



2021 SSPD

Edinburgh, UK and Online

2021 SENSOR SIGNAL PROCESSING FOR DEFENCE (SSPD)

14th and 15th September 2021

Welcome

Programme

Posters

Keynote

Invited Speakers

Technical Committee

2021 Sensor Signal Processing for Defence Conference (SSPD)

Copyright © 2021 by the Institute of Electrical and Electronics Engineers, Inc. All rights reserved.

Copyright and Reprint Permissions

Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923.

For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Service Center, 445 Hoes Lane, Piscataway, NJ 08854. All rights reserved.

IEEE Catalog Number: CFP21SPD-ART (Article)
ISBN: 978-1-6654-3314-3(Article)

Printed copies of this publication are available from:

Curran Associates, Inc
57 Morehouse Lane
Red Hook, NY 12571 USA

Phone: (845) 758-0400
Fax: (845) 758-2633
E-mail: curran@proceedings.com

IEEE eXpress
**Conference
Publishing**

Produced by IEEE eXpress Conference Publishing
For information on producing a conference proceedings and receiving an estimate, contact
conferencepublishing@ieee.org
<http://www.ieee.org/conferencepublishing>

Table of Contents

SSPD 2021 Welcome Message	v
SSPD 2021 Programme	vi
SSPD 2021 Keynote Speakers	x
SSPD 2021 Invited Speakers	xii
SSPD 2021 Conference Committee	xv
Session 1: Imaging and Underwater Signal Processing	
1.1 Fast Classification and Depth Estimation for Multispectral Single-Photon LiDAR Data	1
<i>Mohamed Amir Alaa Belmekki, Stephen McLaughlin, and Abderrahim Halimi</i>	
1.2 Spaceborne SAR Based Assessment of Nuclear Test Effects: The Case of North Korea	6
<i>Nicomino Fiscante, Filippo Biondi, Pia Addabbo, Carmine Clemente, Gaetano Giunta, and Danilo Orlando</i>	
1.3 The Maximal Eigengap Estimator for Acoustic Vector-Sensor Processing	11
<i>Robert Bassett, Jacob Foster, Kay L. Gemba, Paul Leary, and Kevin B. Smith</i>	
Session 2: Poster Presentations	
2.1 Joint Surface Detection and Depth Estimation from Single-Photon Lidar Data using Ensemble Estimators	16
<i>K. Drummond, S. McLaughlin, Y. Altmann, A. Pawlikowska, and R. Lamb</i>	
2.2 Detecting LFM Parameters in Joint Communications and Radar Frequency Bands	21
<i>Kaiyu Zhang, Fraser K. Coutts, and John Thompson</i>	
2.3 Joint Spatio-Temporal Bias Estimation and Tracking for GNSS-Denied Sensor Networks	26
<i>Sofie Macdonald and James R. Hopgood</i>	
2.4 Detection of Human Target Location under Simulated Randomized Rubble using Global Fresnel's Reflection Coefficient	31
<i>Amit Sarkar and Debalina Ghosh</i>	
2.5 Semi-Supervised Domain Adaptation via Adversarial Training	36
<i>Antonin Couturier and Anton-David Almasan</i>	
2.6 Fast Givens Rotation Approach to Second Order Sequential Best Rotation Algorithms	40
<i>Faizan Khattak, Stephan Weiss, and Ian K. Proudler</i>	
2.7 Target Detection and Recognition of Ground Penetrating Radar using Morphological Image Analysis and Graph Laplacian Regularisation	45
<i>Jun Dong, Vladimir Stankovic, and Nigel Davidson</i>	
2.8 Object Detection in EO/IR and SAR Images using Low-SWAP Hardware	50
<i>Richard O. Lane, Adam J. Wragge, Wendy J. Holmes, Stuart J. Bertram, and Tim Lamont-Smith</i>	

2.9 Exponential Filters for Passive Underwater Acoustic Detections - A Global Processing Gain Perspective	55
<i>Stéphane Blouin</i>	
Session 3: RF Sensing and Communication	
3.1 An Approximate Likelihood Ratio Detector for QTMS Radar and Noise Radar	60
<i>David Luong, Bhashyam Balaji, and Sreeraman Rajan</i>	
3.2 Detection of Weak Transient Signals using a Broadband Subspace Approach	65
<i>Stephan Weiss, Connor Delaosa, James Matthews, Ian K. Proudler, and Ben A. Jackson</i>	
3.3 Rate Splitting Multiple Access for Multi-Antenna Multi-Carrier Joint Communications and Jamming	70
<i>Onur Dizdar and Bruno Clerckx</i>	
Session 4: Distributed Processing and Tracking	
4.1 Adaptive Kernel Kalman Filter	75
<i>Mengwei Sun, Mike E. Davies, Ian Proudler, and James R. Hopgood</i>	
4.2 Detection of Malicious Intent in Non-Cooperative Drone Surveillance	80
<i>Jiaming Liang, Bashar I. Ahmad, Mohammad Jahangir, and Simon Godsill</i>	
4.3 Modelling Bi-Static Uncertainties in Sequential Monte Carlo with the GLMB Model	85
<i>Murat Üney, Alexey Narykov, Jason Ralph, and Simon Maskell</i>	
4.4 Graph Filter Design for Distributed Network Processing: A Comparison between Adaptive Algorithms	90
<i>Atiyeh Alinaghi, Stephan Weiss, Vladimir Stankovic, and Ian Proudler</i>	
Session 6: Machine Learning and Information Processing	
6.1 Approximate Proximal-Gradient Methods	95
<i>Anis Hamadouche, Yun Wu, Andrew M. Wallace, and João F. C. Mota</i>	
6.2 Learning a Secondary Source from Compressive Measurements for Adaptive Projection Design	101
<i>Fraser K. Coutts, John Thompson, and Bernard Mulgrew</i>	
Author Index	106
SSPD 2022 Flyer	108

SSPD Conference 2021 - Welcome

Dear Colleagues,

We warmly welcome you to this year's SSPD Conference, our first hybrid conference. This event is the 10th conference of the Sensor Signal Processing for Defence series and provides a chance to present, listen to and discuss the latest scientific findings in signal processing for defence.

We are privileged to have our two keynote speakers, René Vidal from Johns Hopkins Mathematical Institute for Data Science and Hugh Griffiths from the Science Expert Committee (DSEC) / University College London. The SSPD 2021 conference also welcomes our invited speakers; Mark Briers, The Alan Turing Institute; Tien Pham, (CISD) U.S. DEVCOM ARL; Alan Hunter, University of Bath.

A welcome also extends to our panel speakers from Defence, Industry and Academia and the presenters of scientific papers presenting their novel research through live oral presentations. We look forward to some interesting debate and discussion throughout the conference.

We would like to take this opportunity to thank the speakers, reviewers, session chairs and the technical committee for their contribution to this event.

We hope you enjoy our conference.

Mike Davies

Chair, SSPD 2021

Technical sponsorship is provided by IEEE Signal Processing Society and the IEEE Aerospace and Electronic Systems Society. Proceedings will be submitted to the Xplore Digital Library. The conference is organised by the University Defence Research Collaboration (UDRC) in Signal Processing, sponsored by the Defence Science and Technology Laboratory (Dstl) and the Engineering and Physical Sciences Research Council (EPSRC).



Sensor Signal Processing for Defence Programme

Location: Playfair Library Hall, Old College, University of Edinburgh, South Bridge, Edinburgh EH8 9YL.

Tuesday 14th September 2021

8:30 to 9:00 Refreshments

Session 1 – Imaging and Underwater Signal Processing - Chair Stephen McLaughlin, Heriot-Watt University

9:00 Introduction and Welcome to Day 1/Session 1 – Stephen McLaughlin

9:05 – 9:35 Invited Speaker: Multi-Spectral and Multi-Modal Underwater Acoustic Imaging, Alan Hunter, University of Bath.

9:35 – 10:00 Fast Classification and Depth Estimation for Multispectral Single-Photon LiDAR Data, Mohamed Amir Alaa Belmekki¹, Steve McLaughlin¹, Abderrahim Halimi¹, ¹Heriot-Watt University.

10:00 – 10:25 Spaceborne SAR based assessment of nuclear test effects: the case of North Korea, Nicomino Fiscante¹, Filippo Biondi², Pia Addabbo³, Carmine Clemente⁴, Gaetano Giunta¹, Danilo Orlando⁵ ¹University of Study “Roma TRE”, ²University of Study of “L’Aquila”, ³University “Giustino Fortunato”, ⁴University of Strathclyde, ⁵University “Niccolo Cusano”.

10:25 – 10:50 The Maximal Eigengap Estimator for Acoustic Vector-Sensor Processing, Robert L Bassett¹, Jacob Foster¹, Kay Gemba², Paul Leary¹, Kevin B. Smith¹, ¹Naval Postgraduate School, ²U.S. Naval Research Laboratory.

10:50 – 11:20 Refreshments

Session 2 – Mixed Panel Discussion and Posters - Chair - Jordi Barr, Dstl

11:20 Introduction and Welcome to Session 2 – Jordi Barr

11:20 – 12:20 Mixed Panel Discussion: The defence sector is inherently risk-averse in the face of disruptive R&D. How would you enable more rapid adoption of novel, beneficial technology? Moderator, Jordi Barr, Dstl.

12:20 – 12:50 Lightning Presentations

- **P1.** Joint surface detection and depth estimation from single-photon Lidar data using ensemble estimators, Kristofer Drummond¹, Agata Pawlikowska², Robert Lamb², Steve McLaughlin¹, Yoann Altmann¹, ¹Heriot-Watt University, ²Leonardo.
- **P2.** Detecting LFM Parameters in Joint Communications and Radar Frequency Bands, Kaiyu Zhang¹, Fraser K Coutts¹, John Thompson¹, ¹University of Edinburgh.
- **P3.** Joint Spatio-Temporal Bias Estimation and Tracking for GNSS-Denied Sensor Networks, Sofie J. J. Macdonald^{1,2}, James R Hopgood¹, ¹University of Edinburgh, ²Leonardo.
- **P4.** Detection of Human Target Location Under Simulated Randomized Rubble Using Global Fresnel's Reflection Coefficient, Amit Sarkar¹, Debalina Ghosh¹, ¹Indian Institute of Technology Bhubaneswar.

Sensor Signal Processing for Defence Programme

- **P5.** Semi-supervised domain adaptation via adversarial training, Antonin Couturier¹, Anton-David Almasan¹, ¹Thales.
- **P6.** Fast Givens Rotation Approach to Second Order Sequential Best Rotation Algorithms, Faizan Khattak¹, Stephan Weiss¹, Ian Proudler¹, ¹University of Strathclyde.
- **P7.** Target Detection and Recognition of Ground Penetrating Radar using Morphological Image Analysis and Graph Laplacian Regularisation, Jun Dong¹, Vladimir Stankovic¹, Nigel Davidson², ¹University of Strathclyde, ²Dstl.
- **P8.** Object Detection in EO/IR and SAR Images Using Low-SWAP Hardware, Richard O Lane¹, Adam Wragge¹, Wendy Holmes¹, Stuart Bertram¹, Tim Lamont-Smith¹, ¹QinetiQ.
- **P9.** Exponential Filters for Passive Underwater Acoustic Detections - A Global Processing Gain Perspective, Stephane Blouin, Defence Research and Development Canada.

12:50 – 14:00 Lunch and Poster Presentations – There will be an opportunity to view posters either online or at Edinburgh.

Session 3 RF Sensing and Communications - Chair – Gary Heald, Dstl

14:00 Introduction and Welcome to Session 3 – Chair – Gary Heald

14:00 – 15:00 Defence Keynote Speaker: Why Defence Acquisition is Difficult, Hugh Griffiths, Defence Science Expert Committee (DSEC) / University College London.

15:00 – 15:25 An Approximate Likelihood Ratio Detector for QTMS Radar and Noise Radar, David Luong¹, Bhashyam Balaji², and Sreeraman Rajan², ¹Carleton University, ²Defence Research and Development Canada.

15:25 – 15:45 Refreshments

15:45 – 16:10 Detection of Weak Transient Signals Using a Broadband Subspace Approach, Stephan Weiss¹, Connor Delaosa¹, James Matthews², Ian Proudler¹, Ben Jackson³, ¹University of Strathclyde, ²PA Consulting, ³Dstl.

16:10 – 16:35 Rate Splitting Multiple Access for Multi-Antenna Multi-Carrier Joint Communications and Jamming, Onur Dizdar¹, Bruno Clerckx¹, ¹Imperial College London.

16:35 – 16:45 Closing remarks

19:30 Conference Reception Drinks

20:00 Conference Dinner

Sensor Signal Processing for Defence Programme

Wednesday 15th September 2021

8:30 to 9:00 Refreshments

Session 4 Distributed Processing and Tracking – James Hopgood, University of Edinburgh

9:00 Introduction and Welcome to Day 2/Session 4– James Hopgood

9:05 – 9:35 Invited Speaker: Optimising and Understanding the Impact of the NHS COVID-19 app Using Data Science Mark Briers, The Alan Turing Institute.

9:35 – 10:00 Adaptive Kernel Kalman Filter, Mengwei Sun¹, Mike E Davies¹, Ian Proudler², James R Hopgood², ¹University of Edinburgh, ²University of Strathclyde.

10:00 – 10:25 Detection of Malicious Intent in Non-cooperative Drone Surveillance, Jiaming Liang¹, Bashar I. Ahmad², Mohammed Jahangir³, Simon Godsill¹, ¹University of Cambridge, ²Aveillant Thales Land and Air Systems, ³University of Birmingham.

10:25 – 10:50 Modelling bi-static uncertainties in sequential Monte Carlo with the GLMB model, Murat Uney¹, Alexey Narykov¹, Jason F. Ralph¹, Simon Maskell¹, ¹University of Liverpool.

10:50 – 11:15 Graph Filter Design for Distributed Network Processing: A Comparison between Adaptive Algorithms, Atiyeh Alinaghi¹, Stephan Weiss¹, Vladimir Stankovic¹, Ian Proudler¹, ¹University of Strathclyde.

11:15 – 11:45 Refreshments

Session 5 – Mixed Panel Discussion – Chair - TBC - Dstl

11:45 Introduction and Welcome to Session 5 – TBC, Dstl

11:45 – 12:45 Mixed Panel Discussion: The TRL model impedes algorithmic development. How does the defence community need to change to reduce barriers and increase opportunities?, Moderator, TBC, Dstl.

12:45 – 13:45 Lunch

Session 6 – Machine Learning and Information Processing – Chair Mike Davies, University of Edinburgh

13:45 Introduction and Welcome to Session 6 – Chair – Mike Davies

13:45– 14:45 Academic Keynote Speaker: Semantic Information Pursuit, René Vidal, Johns Hopkins Mathematical Institute for Data Science

14:45 – 15:15 Invited Speaker: AI-enabled Multi-Domain Processing and Analytics for Decision Making, Tien Pham, (CISD) U.S. DEVCOM ARL.

15:15 – 15:45 Refreshments

15:45 – 16:10 Approximate Proximal-Gradient Methods, Anis Hamadouche¹, Yun Wu¹, Andrew M Wallace¹, Joao Mota¹, ¹Heriot-Watt University.

16:10 – 16:35 Learning a Secondary Source From Compressive Measurements for Adaptive Projection Design, Fraser K Coutts¹, John Thompson¹, Bernard Mulgrew¹, ¹University of Edinburgh.

Sensor Signal Processing for Defence Programme

16:35– 16:45 Closing remarks

Keynote Speakers

Professor René Vidal, Johns Hopkins Mathematical Institute for Data Science

Rene Vidal is the Herschel Seder Professor of Biomedical Engineering and the Inaugural Director of the Mathematical Institute for Data Science at The Johns Hopkins University. He has secondary appointments in Computer Science, Electrical and Computer Engineering, and Mechanical Engineering. He is also a faculty member in the Center for Imaging Science (CIS), the Institute for Computational Medicine (ICM) and the Laboratory for Computational Sensing and Robotics (LCSR). Vidal's research focuses on the development of theory and algorithms for the analysis of complex high-dimensional datasets such as images, videos, time-series and biomedical data. His current major research focus is understanding the mathematical foundations of deep learning and its applications in computer vision and biomedical data science. His lab has pioneered the



development of methods for dimensionality reduction and clustering, such as Generalized Principal Component Analysis and Sparse Subspace Clustering, and their applications to face recognition, object recognition, motion segmentation and action recognition. His lab creates new technologies for a variety of biomedical applications, including detection, classification and tracking of blood cells in holographic images, classification of embryonic cardio-myocytes in optical images, and assessment of surgical skill in surgical videos.

Abstract: Semantic Information Pursuit

In 1948, Shannon published a famous paper, which laid the foundations of information theory and led to a revolution in communication technologies. Critical to Shannon's ideas was the notion that a signal can be represented in terms of "bits" and that the information content of the signal can be measured by the minimum expected number of bits. However, while such a notion of information is well suited for tasks such as signal compression and reconstruction, it is not directly applicable to audio-visual scene interpretation tasks because bits do not depend on the "semantic content" of the signal, such as words in a document, or objects in an image. In this talk, I will present a new measure of semantic information content called "semantic entropy", which is defined as the minimum expected number of semantic queries about the data whose answers are sufficient for solving a given task (e.g., classification). I will also present an information-theoretic framework called "information pursuit" for deciding which queries to ask and in which order, which requires a probabilistic generative model relating data and questions to the task. Experiments on handwritten digit classification show, for example, that the translated MNIST dataset is harder to classify than the MNIST dataset. Joint work with Aditya Chattopadhyay, Benjamin Haeffele and Donald Geman.

Professor Hugh Griffiths, Defence Science Expert Committee (DSEC) / University College London

Hugh Griffiths holds the THALES/Royal Academy Chair of RF Sensors at University College London and is Chair of the Defence Science Expert Committee (DSEC) in the UK Ministry of Defence. From 2006–2008 he served as Principal of the Defence Academy College of Management and Technology at Shrivenham. He received the MA degree in Physics from Oxford University in 1975, then spent three years working in industry, before joining University College London, where he received the PhD degree in 1986 and the DSc(Eng) degree in 2000 and served as Head of Department from 2001–2006.

His research interests include radar systems and signal processing (particularly bistatic radar and synthetic aperture radar), and antenna measurement techniques. He serves as Editor-in-Chief of the IET Radar, Sonar and Navigation journal. He has published over five hundred papers and technical articles in the fields of radar, antennas and sonar. He has received several awards and prizes, including the IEEE Picard Medal (2017), IET Achievement Medal (2017), the IEEE AES Mimno Award (2015), the IET A.F. Harvey Prize (2012) and the IEEE AES Nathanson Award (1996). He is a Fellow of the IET and a Fellow of the IEEE. He was appointed OBE in the 2019 New Year's Honours List, and in 2021, he was elected Fellow of the Royal Society.



Abstract: Why Defence Acquisition is Difficult

The acquisition of defence equipment – whether tanks, aircraft carriers or fast jets – is notoriously difficult. On the occasions when it goes wrong we hear of cancelled projects or costly spending overruns, amounting to billions of pounds of taxpayers' money. This is not just a UK problem – it happens in other countries too.

This presentation looks at the process of defence acquisition, and particularly the research and development that underpins it. Examples are given of both successes and failures. In particular, it examines the nature of innovation and of low-TRL research and considers the appropriate balance between risk and ambition.

Invited Speakers

Professor Mark Briers, Alan Turing Institute

Mark Briers is Programme Director for The Alan Turing Institute's Defence and Security partnership. Prior to joining Turing, Mark worked in the defence and security sector for over 16 years, directing research programmes in the area of statistical data analysis, and leading large teams to drive impactful research outputs. He completed his PhD in 2007 at Cambridge University where he developed Sequential Monte Carlo based techniques for state-space filtering and smoothing. In recent months, Mark has been providing independent advice to scope the technical development and help to oversee modelling and analytics of the NHS Test and Trace app.

He is an Honorary Senior Lecturer at Imperial College London, where he teaches methodological techniques for use in a Big Data environment and conducts research into statistical methods for cyber security, and he is a Council Member at the Royal Statistical Society. He is an Industrial Fellow alumnus of the Royal Commission for the Exhibition of 1851. Mark is a co-investigator on the EPSRC funded programme grant Computational Statistical Inference for Engineering and Security.

Mark is also Co-Chair of the Institute's Research and Innovation Advisory Committee, which steers scientific direction at the Turing.

Abstract: Optimising and Understanding the Impact of the NHS COVID-19 app Using Data Science

The NHS COVID-19 app has been downloaded over 21 million times. It was the first app in the world to use the new Google / Apple API, allowing novel Bayesian statistical methods to infer distance. Our latest research demonstrates that the app has reduced the number of cases by approximately 600,000 in 2020. The app data are providing valuable insights into the evolution of the pandemic across England and Wales. This presentation will provide an overview of the NHS COVID-19 app, the impact results, and our potential future directions.



Dr Tien Pham, CISD) U.S. DEVCOM ARL

Dr Tien Pham is the Chief Scientist for the Computational & Information Sciences Directorate (CISD) at U.S. DEVCOM ARL. He oversees basic and applied research associated within the ARL Competencies in Military Information Sciences and Network, Cyber & Computational Sciences and the cross-cutting research efforts in Artificial Intelligence and Machine Learning. Has 30+ years of R&D and research program management experience and has published results in wide-ranging research areas from information sciences to networked sensing, multi-modal sensor fusion, and acoustics. He actively participates in a number of technical forums including IEEE, SPIE Defense + Commercial Sensing, Military Sensing Symposium, ISIF Fusion, AAAI Fall Symposium Series, Naval Applications of Machine Learning, Acoustical Society of America and NATO & TTCP. Dr Pham received his B.S. (Magna Cum Laude), M.S. and PhD degrees in Electrical Engineering at the University of Maryland, College Park, USA, in 1988, 1991 and 2006 respectively.

Abstract: AI-enabled Multi-Domain Processing and Analytics for Decision Making

The U.S. Army is evolving its warfighting doctrine to militarily compete, penetrate, disintegrate, and exploit peer adversaries in a Multi-Domain Operations (MDO) Joint Force. To achieve this vision, the Army requires Artificial Intelligence (AI) and Machine Learning (ML) capabilities to enable analytics from the full spectrum of multi-domain data for decision making by highly-dispersed teams of humans and robot agents. This highly diverse learning and reasoning span data types (e.g. video, open-source multi-media, electromagnetic (EM) signals, sensors), warfighting domains (e.g. land, air, cyberspace, spectrum), and warfighting functions (e.g. intelligence, command and control (C2), fires, protection). MDO presents unique challenges that must be overcome by any AI/ML application, including forwarding deployment in complex terrain, dynamic, distributed, resource-constrained environments, and highly contested settings. Moreover, these applications must consider the inherent uncertainties (model and environmental) and the operational timeliness requirements (i.e., time available to learn, infer, and act).

The presentation will discuss the U.S. Combat Capabilities Development Command (DEVCOM) Army Research Laboratory (ARL) collaborative, cross-cutting AI research efforts for MDO to enable processing and analytics for decision making toward the tactical edge. The research efforts will address the technical challenges involving (i) learning and reasoning in complex MDO environments and (ii) resource-constrained AI processing at the point of need.

Professor Alan Hunter, University of Bath

Alan was born in New Zealand and studied at the University of Canterbury (NZ), obtaining a BE (Hons I) degree in electrical and electronic engineering in 2001 and a PhD on synthetic aperture sonar (SAS) in 2006. He left New Zealand for Europe in 2007.

For the next three years, Alan was a research associate at the University of Bristol, where he worked with ultrasonic arrays for non-destructively inspecting engineering components. Here, he developed auto-focusing algorithms for imaging the interiors of objects with complex geometries and material properties.

In 2010, he moved to the Netherlands and spent four years as a defence scientist at TNO (Netherlands Organisation for Applied Research) in The Hague. At TNO, he developed underwater technologies for the Royal Dutch Navy, including sub-sediment imaging sonar, diver detection sonar, and autonomous naval mine-hunting systems.

Before returning to England in 2014, Alan spent three months in La Spezia, Italy as a visiting scientist at the NATO Centre for Maritime Research and Experimentation (CMRE). During this time, he developed precision micro-navigation algorithms for SAS imaging of the seafloor over repeated passes.

Abstract: Multi-Spectral and Multi-Modal Underwater Acoustic Imaging

The useable spectrum of underwater acoustic frequencies is very broad, spanning roughly 6 or 7 orders of magnitude from Hz up to MHz. Traditionally, acoustic imaging of the seafloor has been carried out at high carrier frequencies (in the range of 100 kHz to 1 MHz) using relatively narrow-band signals (quality factors much greater than 1). This has enabled the production of high-resolution acoustic images that often resemble optical photographs. However, there are opportunities to exploit much more of the spectrum and thereby access a richer source of acoustic information. An effective means of achieving this is to operate in multiple separate spectral bands. Bands in the higher frequencies retain the advantages of fine resolution, differences between bands can be used to measure frequency-dependent scattering characteristics, and lower frequency bands (below 100 kHz) increasingly enable penetration beneath the sediment and inside objects. Moreover, low-frequency acoustic waves can couple with elastic wave modes in solids and this gives the potential for sensing material composition and structure for enhanced characterisation. This talk will explore the opportunities and challenges of multi-spectral and multi-modal (i.e., acoustic and elastic) imaging with regard to signal processing, image visualisation, and interpretation.



SSPD2021 Conference Committee

General Chairs

- Mike Davies - University of Edinburgh
- Stephen McLaughlin - Heriot-Watt University
- Jordi Barr - Dstl
- Gary Heald - Dstl

Publicity and Local Arrangements Chair

- Janet Forbes - University of Edinburgh

Technical Programme Committee

- Abderrahim Halimi, Heriot-Watt University
- Alasdair Hunter, Dstl
- Alessio Balleri, Cranfield University
- Andreas Ahrens, Hochschule Wismar
- Andreas Spanias, Arizona State University
- Andrew Wallace, Heriot-Watt University
- Andy Stove, Stove Specialties
- Antonio De Maio, University of Naples "Federico II"
- Athanasios Gkelias, Imperial College London
- Augusto Aubry, Università degli studi di Napoli
- Bernard Mulgrew, University of Edinburgh
- Brian Barber, Dstl
- Bruno Clerckx, Imperial College London
- Carmine Clemente, University of Strathclyde
- Chris Baker, University of Birmingham
- Christoph Wasserzrier, Fraunhofer Institute for High Frequency Physics and Radar Techniques FHR
- Christos Ilioudis, University of Strathclyde
- Cristian Rusu, Dublin University
- David Blacknell, Dstl
- David Cormack, Leonardo
- David Garren, Naval Postgraduate School
- David Greig, Leonardo
- David Harvey, Thales UK
- David Nethercott, Dstl
- Domenico Gaglione, CMRE NATO
- Duncan Williams, Dstl
- Fan Liu, University College London
- Geert Leus, Delft University of Technology

- Harvey Alison, Leonardo
- Henry Gouk, University of Edinburgh
- Hugh Griffiths, UCL
- Ian Proudler, University of Strathclyde
- Ivo Bukovsky, Czech Technical University in Prague
- James Hopgood, University of Edinburgh
- Jason Ralph, University of Liverpool
- Joao Mota, Heriot-Watt University
- John Buck, University of Massachusetts
- John Thompson, University of Edinburgh
- Julian Deeks, Dstl
- Keith Brown, Heriot-Watt University
- Ken McEwan, Dstl
- Krishnaprasad Nambur Ramamohan, Delft University of Technology
- Laura Anitori, TNO
- Mahesh Banavar, Clarkson University
- Maria Greco, University of Pisa
- Mathini Sellathurai, Heriot-Watt University
- Mehrdad Yaghoobi, University of Edinburgh
- Murat Uney, Centre for Maritime Research and Experimentation
- Neil Cade, Leonardo
- Neil Robertson, Queen's University of Belfast
- Nick Goddard, Dstl
- Nikolaos Dionelis, University of Edinburgh
- Oliver Sims, Leonardo
- Paul Thomas, Dstl
- Paul Westoby, Dstl
- Richard Jackson, Dstl
- Rodrigo de Lamare, Pontifical Catholic University of Rio de Janeiro
- Sami Aldalahmeh, Al-Zaytoonah University of Jordan
- Sen Wang, Heriot-Watt University
- Simon Godsill, University of Cambridge
- Simon Maskell, University of Liverpool
- Stephan Weiss, University of Strathclyde
- Stephen Ablett, Dstl
- Suresh Jacob, Dstl
- Vladimir Stankovic, University of Strathclyde
- Wenwu Wang, University of Surrey
- Wolfgang Koch, Fraunhofer FKIE / University of Bonn
- Yoann Altmann, Heriot-Watt University

Fast Classification and Depth Estimation for Multispectral Single-Photon LiDAR Data

Mohamed Amir Alaa Belmekki, *Student Member, IEEE*, Stephen McLaughlin, *Fellow, IEEE*,
Abderrahim Halimi, *Senior Member, IEEE*

School of Engineering and Physical Sciences, Heriot-Watt University, Edinburgh U.K.
e-mail: {mb219, s.mclaughlin, a.halimi}@hw.ac.uk

Abstract—Multispectral 3D LiDAR imaging plays an important role in the remote sensing community as it can provide rich spectral and depth information from targets. This paper proposes a fast pixel-wise classification algorithm for multispectral single-photon LiDAR imaging. The algorithm allows the detection of histograms containing surfaces with specific spectral signatures (i.e., specific materials) and discarding those histograms without reflective surfaces. The proposed Bayesian model is carefully built to allow the marginalization of latent variables leading to a tractable formulation and fast estimation of the parameters of interest, together with their uncertainties. Results on simulated and real single-photon data illustrates the robustness and good performance of this approach.

Index Terms—3D Multispectral imaging, Single-photon LiDAR, Bayesian estimation, Poisson statistics, Multispectral classification.

I. INTRODUCTION

Light detection and ranging (LiDAR) used with a time-correlated single-photon detector has received increased attention in the scientific community over the last few decades in numerous applications such as defence, automotive [1], and environmental sciences [2]. It has successfully proven its efficiency for long-range depth imaging [3]–[5], depth imaging in turbid media [6]–[8] and multispectral imaging [9], [10]. A multispectral LiDAR system operates by illuminating the scene using laser pulses of different wavelengths and recording the arrival times of the reflected photons using a time-correlated single-photon counting (TCSPC) module. A histogram of photon counts with respect to time-of-flight is constructed for each pixel and wavelength. The resulting histograms contain information regarding the object depth profile and its spectral signature (i.e., reflectivity at each wavelength). When the number of spectral bands considered is large enough, it becomes possible to identify and quantify the main materials in the scene, in addition to estimating classical LiDAR-based range profiles.

In the last decade, many works were published promoting the use of multiple wavelengths in LiDAR imaging as it allows robust restoration of the range and spectral profile of the target [10]–[12], spectral classification [13] and spectral unmixing [9], [14]. In [13], joint depth estimation and spectral classification using multispectral 3D LiDAR have been proposed. The algorithm combined a Bayesian model with a Markov chain Monte-Carlo (MCMC) sampling strategy to obtain estimates and their uncertainties. However, the algorithm assumed the absence of background counts due to ambient light, and has a significant

computational cost due to the considered MCMC sampling method, which is incompatible with real-time requirements. A faster algorithm was proposed in [9] by considering an optimization algorithm to estimate a depth image and spatially unmix the contribution of different materials. However, the method did not quantify the estimates uncertainty, as often required for decision making. Recently, two fast algorithms were introduced in [15], [16] to perform target detection task using 3D Lidar imaging. These two algorithms use a Bayesian formulation and show state-of-the-art performance even under the presence of background illumination. Nevertheless, these two algorithms are not able to discern between different materials having distinctive spectral responses.

This paper proposes a new fast algorithm for per-pixel multispectral classification and range estimation, while also providing uncertainty measures about the estimates. Assuming a known spectral library of different materials, the proposed Bayesian strategy allows the detection of pixels with reflections from a target, and to identify the target material based on the known signatures. The proposed algorithm operates on histograms of counts and assumes the presence of at most one surface per-pixel. The model accounts for the Poisson statistics of the data to define the likelihood, and considers suitable prior distributions for the parameters of interest. The resulting marginalized posterior distribution can be efficiently exploited allowing parameters estimation together with uncertainty quantification. The proposed algorithm can be used to compress the high-volume multispectral data by withdrawing uninformative pixels without surfaces/objects. It can also be used as a target detection algorithm to spatially locate materials with specific spectral signatures, e.g., the detection of man-made objects in turbid environments such as metallic objects underwater. The proposed algorithm is tested both on real and simulated data showing promising results even in the sparse photon regime or in presence of a high background level of counts.

The paper is structured as follows. Section II introduces the multispectral LiDAR observation model. The classification problem formulation using Bayesian framework and the estimation strategy are presented in Section III and Section IV, respectively. Results on simulated and real data are presented in Section V. Conclusions and future work are finally reported in Section VI.

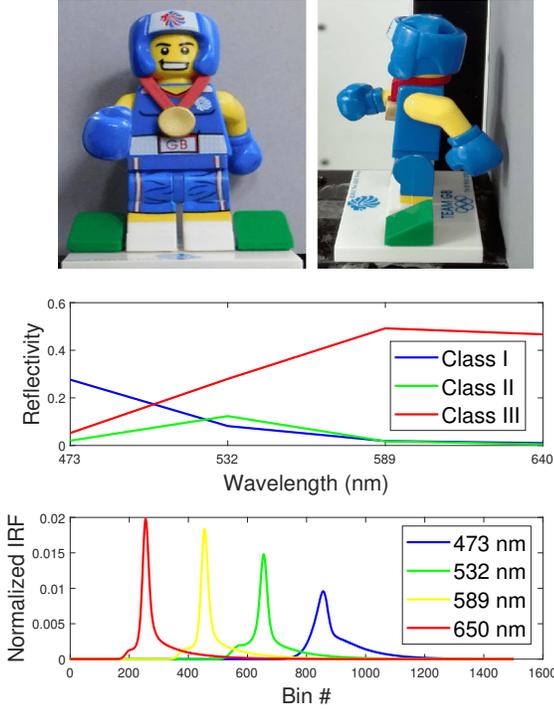


Fig. 1. (Up) The considered scene used in section V, (middle) spectral signatures of the spectral classes associated with Lego scene, (down) the normalized IRFs associated with the wavelengths 473, 532, 589 and 640 nm [19] used in Section V

II. OBSERVATION MODEL

We consider a 3-dimensional cube of histograms \mathbf{Y} of LiDAR photon counts of dimension $N \times L \times T$, where N is the number of scanned spatial positions (i.e., pixels), L is the number of spectral wavelengths and T is the number of time bins. Let $\mathbf{Y}_n = [\mathbf{y}_{n,1}, \mathbf{y}_{n,2}, \dots, \mathbf{y}_{n,L}]^T$ be an $L \times T$ matrix where $\mathbf{y}_{n,l} = [y_{n,l,1}, y_{n,l,2}, \dots, y_{n,l,T}]^T$. According to [17], [18], each photon count $y_{n,l,t}$, where $n \in \{1, \dots, N\}$, $l \in \{1, \dots, L\}$ and $t \in \{1, \dots, T\}$, is assumed to follow a Poisson distribution as follows

$$y_{n,l,t} | r_{n,l}, d_n, b_{n,l} \sim \mathcal{P}[r_{n,l} g_l(t - d_n) + b_{n,l}], \quad (1)$$

where $\mathcal{P}(\cdot)$ denotes a Poisson distribution, $r_{n,l} \geq 0$ is the spectral signature observed at the l_{th} wavelength, $d_n \in \{1, 2, \dots, T\}$ represents the position of an object surface at a given range from the sensor, $b_{n,l} \geq 0$ is the constant background level associated with dark counts and ambient illumination and $g_l(\cdot)$ is the system impulse response function (IRF), whose shape can differ between wavelength channels (see Fig.1), assumed to be known from a calibration step and normalized $\sum_{t=1}^T g_l(t) = 1$. An equivalent model can be considered as in [15] using the signal-to-background ratio (SBR), which is defined as the ratio of the useful detected photons $r_{n,l}$ and the total number of background photons in the histogram $b_{n,l}T$, i.e.: $w_{n,l} = \frac{r_{n,l}}{b_{n,l}T}$ with $w_{n,l} \geq 0$. Thus, (1) can be written in the following form

$$y_{n,l,t} | \omega_{n,l}, d_n, b_{n,l} \sim \mathcal{P}\{b_{n,l} [w_{n,l} T g_l(t - d_n) + 1]\}. \quad (2)$$

This new formulation is interesting as it allows an easy marginalization of the posterior distribution with respect to

(w.r.t.) the background noise parameter as indicated in Section III. Under the assumption that all the observed pixels, wavelengths and bins are mutually independent, the joint likelihood can be written in the following form:

$$p(\mathbf{Y} | \Omega, \mathbf{d}, \mathbf{B}) = \prod_{l=1}^L \prod_{t=1}^T p(y_{n,l,t} | \omega_{n,l}, d_n, b_{n,l}) \quad (3)$$

where $\mathbf{d} = (d_1, \dots, d_N)$ and Ω, \mathbf{B} are two matrices gathering $\omega_{n,l}, \forall n, l$, and $b_{n,l}, \forall n, l$, respectively. Our goal is to perform pixel classification based on an available database of spectral signatures. As a result, the classification task will require an elaborated strategy to mitigate the unknown $b_{n,l}, \omega_{n,l}$ and d_n parameters.

III. HIERARCHICAL BAYESIAN MODEL FOR CLASSIFICATION

The classification problem is an ill-posed problem that is tackled by considering a Bayesian approach. The latter assigns prior distributions for the unknown parameters to regularize the inverse-problem.

A. Prior distributions

A Lidar histogram can either result from background counts (in absence of a target photons due to $r_n = \omega_n = 0$) or a mixture of target and background counts (when $r_n \geq 0$ or $\omega_n \geq 0$). Assuming the presence of K spectral signatures, the classification problem aims to associate a pixel with a target to one of the K spectral classes. The reflectivity prior accounts for this effect by considering a mixture of $K + 1$ distributions as follows

$$P(r_{n,l} | u_n, \alpha_{k,l}^r, \beta_{k,l}^r, K) = \delta(u_n) \delta(r_{n,l}) + \sum_{k=1}^K \delta(u_n - k) \mathcal{G}(r_{n,l}; \alpha_{k,l}^r, \beta_{k,l}^r) \quad (4)$$

where $u_n \in \{0, 1, \dots, K\}$ is a latent variable that indicates the absence of target if $u_n = 0$, otherwise, it indicates the label of the class, $\delta(\cdot)$ is the Dirac delta distribution centred in 0, $\mathcal{G}(r_{n,l}; \alpha_{k,l}^r, \beta_{k,l}^r)$ represents a gamma density whose shape and scale hyperparameters $(\alpha_{k,l}^r, \beta_{k,l}^r)$ are fixed based on the K known spectral signatures. This prior is inspired from the spike-and-slab prior used in [15]. It accounts for $K + 1$ cases, the first represents the absence of a target in the n_{th} pixel and is obtained for $u_n = r_{n,l} = 0$, hence the use of a Dirac distribution (the spike part). The slab part accounts for the presence of one of the K signatures by using a gamma distribution. Thanks to the use of many wavelengths, this prior extends the object detection problem in [15] to a class detection problem using the spectral signature of each class.

Considering the non-negativity of $b_{n,l}, \forall n, l$ and its continuous nature, the background level will be modelled with a gamma distribution as in [20]:

$$P(b_{n,l} | \alpha_l^b, \beta_l^b) = \mathcal{G}(b_l, \alpha_l^b, \beta_l^b) \quad (5)$$

where α_l^b and β_l^b are background hyper-parameters. Considering that only dozens of distinctive wavelengths will be used, we will

assume uncorrelated channels to keep the estimation strategy tractable. Since we are interested in using the model described in (2) instead of (1), assuming that the reflectivity and the background noise are independent and by applying a random variable change, the resulting joint prior distribution will yield:

$$\begin{aligned} p(\boldsymbol{\omega}_n, \mathbf{b}_n | u_n, \boldsymbol{\phi}, K) &= \prod_{l=1}^L p(\omega_{n,l}, b_{n,l} | u_n, \phi_l) \\ &= \prod_{l=1}^L \left[\delta(u_n) \delta(\omega_{n,l}) \mathcal{G}(b_{n,l}, \alpha_l^b, \beta_l^b) \right. \\ &\quad \left. + \sum_{k=1}^K \delta(u_n - k) C_{k,l}(\omega_{n,l}) \mathcal{G}(b_{n,l}, \alpha_{l,k}^\dagger, \beta_{l,k}^\dagger(\omega_{n,l})) \right] \end{aligned} \quad (6)$$

with

$$\begin{aligned} C_{k,l}(\omega_{n,l}) &= \frac{(\beta_l^b)^{\alpha_l^b} (\beta_{k,l}^r)^{\alpha_{k,l}^r} T^{\alpha_{k,l}^r}}{B(\alpha_{k,l}^r, \alpha_l^b)} \frac{\omega_{n,l}^{\alpha_{k,l}^r - 1}}{\beta_{l,k}^\dagger(\omega_{n,l}) \alpha_{l,k}^\dagger} \\ \alpha_{l,k}^\dagger &= \alpha_l^b + \alpha_{k,l}^r \\ \beta_{l,k}^\dagger(\omega_{n,l}) &= \beta_l^b + \beta_{k,l}^r T \omega_{n,l} \end{aligned}$$

where $B(\cdot)$ is the beta function and $\boldsymbol{\phi} = (\phi_1, \phi_2, \dots, \phi_L)$ with $\phi_l = (\alpha_{k,l}^r, \beta_{k,l}^r, \alpha_l^b, \beta_l^b)$, $k \in \{0, \dots, K\}$.

As we suppose that we have no prior knowledge about a pixel's class, the parameter u_n is assumed to be drawn from a uniform distribution, i.e.: $p(u_n = k) = \frac{1}{K+1}$, where $k \in \{0, \dots, K\}$. However, This non-informative class prior can be changed in presence of additional information regarding the classes. The depth parameter d_n is assigned a non-informative uniform prior as follows:

$$p(d_n = t) = \frac{1}{T}, \quad \forall t \in \{1, \dots, T\}. \quad (7)$$

Nonetheless, this can be modified in case of additional information regarding the target position.

B. Joint Posterior distribution

From the joint likelihood in equation (3) derived in Section II and the prior distributions specified in Section III-A, we can obtain the joint posterior distribution for $\boldsymbol{\omega}_n, \mathbf{b}_n, d_n$ and u_n given the 3D histograms \mathbf{Y}_n and the hyperparameters $\boldsymbol{\phi}$ and K . Using Bayes rule and assuming that d_n and u_n are independent from $\boldsymbol{\omega}_n$ and \mathbf{b}_n , the joint posterior distribution of the proposed Bayesian model can be formulated in the following form:

$$p(\boldsymbol{\Theta}_n | \mathbf{Y}_n, \boldsymbol{\phi}, K) \propto p(\mathbf{Y}_n | \boldsymbol{\Theta}_n) p(\boldsymbol{\Theta}_n | \boldsymbol{\phi}, K) \quad (8)$$

where

$$\begin{aligned} \boldsymbol{\Theta}_n &= (\boldsymbol{\omega}_n, \mathbf{b}_n, d_n, u_n) \\ p(\boldsymbol{\Theta}_n | \boldsymbol{\phi}, K) &= p(\boldsymbol{\omega}_n, \mathbf{b}_n, d_n, u_n | \boldsymbol{\phi}, K) \\ &= p(\boldsymbol{\omega}_n, \mathbf{b}_n | \boldsymbol{\phi}, u_n, K) p(d_n) p(u_n | K). \end{aligned} \quad (9)$$

IV. ESTIMATION STRATEGY

The posterior distribution in (8) reflects our knowledge of the unknown parameters to be estimated given the photon data and the available prior knowledge. The Bayesian estimator to be considered, both for the depth and class parameter, is the maximum a posteriori (MAP) estimator as in [9], [14]. From equation (8), we marginalize the background noise and signal-to-background parameters to get the joint depth and class marginal probability as follows:

$$p(u_n, d_n | \mathbf{Y}_n) = \int \int p(\boldsymbol{\omega}_n, \mathbf{b}_n, d_n, u_n | \mathbf{Y}_n) d\mathbf{b}_n d\boldsymbol{\omega}_n. \quad (10)$$

A. Class estimation

The following decision rule is adopted to determine the pixel label

$$H_n = \max_{k=1:K} p(u_n = k | \mathbf{Y}_n) \quad (11)$$

with

$$p(u_n | \mathbf{Y}_n) = \sum_{d_n=1}^T \int \int p(\boldsymbol{\omega}_n, \mathbf{b}_n, d_n, u_n | \mathbf{Y}_n) d\mathbf{b}_n d\boldsymbol{\omega}_n \quad (12)$$

where H_n represents the class of the n^{th} pixel. Note that for $K = 1$ and $L = 1$, we end up with a target detection decision rule as in [15]. We demonstrate that the marginal probability $p(u_n | y_n)$ is :

$$\begin{aligned} p(u_n = k | \mathbf{Y}_n) &= \sum_{d_n=1}^T \int_0^\infty \prod_{l=1}^L [p(u_n = k) p(d_n) D_{n,l,k} \gamma_l^{-1} \\ &\quad F_{n,l,k}(\omega_{n,l}, d_n) d\omega_{n,l}] \end{aligned} \quad (13)$$

with

$$\begin{aligned} \gamma_l &= \frac{\Gamma(\alpha_l^b)}{(\beta_l^b)^{\alpha_l^b}} \prod_{t=1}^T y_{n,l,t}! \\ D_{n,l,k} &= \frac{\Gamma(\bar{y}_{n,l} + \alpha_l^b + \alpha_{k,l}^r) (T \beta_{k,l}^r)^{\alpha_{k,l}^r}}{\Gamma(\alpha_{k,l}^r)} \\ F_{n,l,k}(\omega_{n,l}, d_n) &= \frac{\exp\{\sum_{t=1}^T y_{n,l,t} \ln[\omega_{n,l} T g_l(t - d_n) + 1]\}}{\omega_{n,l}^{1 - \alpha_{k,l}^r} \{\beta_l^b + [T(1 + \omega_{n,l}(1 + \beta_{k,l}^r))]\}^{\alpha_{l,k}^\dagger + \bar{y}_{n,l}}} \end{aligned} \quad (14)$$

where $\bar{y}_{n,l} = \sum_{t=1}^T y_{n,l,t}$. In the event of no target, we can see that the integral is available in its analytical form thanks to the conjugacy between the model (2) and the priors (6). The marginal distribution in (13), however, is intractable. One way to simplify it is to consider that the depth captured is different across all the spectral wavelengths. This simplification improves the tractability of the marginal class probability and will transform (13) into (15) as follows:

$$\begin{aligned} p(u_n = k | \mathbf{Y}_n) &= \prod_{l=1}^L \sum_{d_n=1}^T \left[p(u_n = k) p(d_n) D_{n,l,k} \gamma_l^{-1} \right. \\ &\quad \left. \int_0^\infty F_{n,l,k}(\omega_{n,l}, d_n) d\omega_{n,l} \right]. \end{aligned} \quad (15)$$

The resulting integral with respect to $\omega_{n,l}$ in (15) can be numerically approximated with a quadrature method. The matched filter in (14) can be computed with $\mathcal{O}(T \log T)$ using the fast Fourier transform (FFT) leading to an overall complexity of the integral per-pixel in (15) given by $\mathcal{O}(KLJT \log T)$, where K is the number of classes considered, L is the number of wavelengths, J is the computational cost of the evaluated integrand and T is the number of the temporal bins.

B. Depth estimation

The depth estimate can be obtain as follows

$$\hat{d}_n = \max_{d=1:T} p(d_n | \mathbf{Y}_n) \quad (16)$$

where

$$p(d_n | \mathbf{Y}_n) = \sum_{k=1}^K \prod_{l=1}^L \left[p(u_n = k) p(d_n) D_{n,l,k} \gamma_l^{-1} \int_0^\infty F_{n,l,k}(\omega_{n,l}, d_n) d\omega_{n,l} \right]. \quad (17)$$

Although ω can be marginalized numerically from equation (17), this might lead to a high computational cost. In this paper, we choose to estimate the depth given the easily computed marginal map estimate $\omega_{n,l}^{\text{map}}$ using the simplified model introduced in Section IV-A, leading to

$$\hat{d}_n = \max_{d=1:T} p(d_n | \mathbf{Y}_n, \omega_n^{\text{map}}). \quad (18)$$

The proposed approach allows the evaluation of the full marginal depth posterior. In addition to depth point estimate, this distribution will allow uncertainty quantification (i.e., quantify our confidence regarding the estimates). In this paper, we evaluate the depth uncertainty by considering the depth negative log-cumulative marginal posterior around the MAP estimate, i.e., $\text{NCD} = -\log \left[\sum_{\hat{d}_n - \epsilon}^{\hat{d}_n + \epsilon} p(d_n | \mathbf{Y}_n, \omega_n^{\text{map}}) \right]$, where ϵ is a user fixed constant. Note that a small NCD indicates a high confidence about the estimate, while a large one would be an indication of low confidence.

V. EXPERIMENTAL RESULTS

In this section, we will evaluate the performance of the classification algorithm using the real multispectral single-photon Lego data used in [19] (see Fig. 1). This target has three classes of interest ($K = 3$) whose spectral signatures (related to α^r and β^r) are extracted from pixels acquired considering a negligible background contribution and after maximum acquisition time per-pixel (see signatures in Fig.1). α_l^b and β_l^b are relatively non-informative such that $(\alpha_l^b, \beta_l^b) = (1, \frac{T}{r_l^M})$ with r_l^M being the average number of signal photons per pixel for the l_{th} spectral wavelength. Two experiments have been performed. First, we consider a spatially sub-sampled data to analyse the behavior of the algorithm w.r.t. SBR and photons levels. The subsampled data has $N = 40 \times 40$ pixels, $L = 4$ wavelengths and $T = 1500$ time bins (bin width of 2ps), and is corrupted so that the SBR varies from 0.01 to 100. The object, of size 42 mm tall and 30 mm wide, was scanned with a 40ms acquisition time per pixel and per wavelength (total of 160ms per pixel) at a standoff distance of 1.8m with the IRF $g_l(\cdot)$

depicted in Fig. 1 (the reader is advised to consult [19] for more details). Considering the foregoing parameters and for maximum acquisition time, the average computational time per-pixel of the proposed algorithm is ≈ 55 ms using Matlab 2020a on a Intel Core i7-8700@3.2GHz, 16 GB RAM. Fig. 3 top and bottom represent, respectively, the root mean square error (RMSE) in meters defined by $\text{RMSE} = \sqrt{\frac{1}{N} \|\mathbf{d}^{\text{ref}} - \hat{\mathbf{d}}\|^2}$, where \mathbf{d}^{ref} is obtained from sampling the whole scene with the maximum acquisition time and under a negligible background illumination, and the accuracy ($\text{ACC} = \frac{TP+TN}{TP+TN+FP+FN}$) w.r.t SBR and the average signal photons, where TP , TN , FP and FN represent respectively: true positive, true negative, false positive and false negative. These two figures provide the user with the required number of useful photons (which is proportional to the scanning time) needed to have a given depth precision and accuracy for different SBR levels. The second experiment shows the obtained results when considering the full 200×200 pixels data. Fig. 2 represents (from left to right) the depth estimate, the class estimate and depth uncertainty measure w.r.t acquisition time per-pixel and per-wavelength and for Lego data in absence of background illumination (i.e., $\text{SBR} = \omega = 66$) and high illumination background where the SBR is as low as $\omega = 1.3$. These figures exhibit qualitatively the robustness of the proposed algorithm. Fig. 2(c) represents the NCD for $\epsilon = 1.5$ mm where higher values indicate low-confidence regions. It is observed that higher depth uncertainty is observed for lower acquisition time data, and for low photon regions (e.g., black mouth and eyes, curved surfaces, etc).

VI. CONCLUSIONS

This paper has presented a new fast approach that performs depth estimation and class detection for multispectral LiDAR data which shows robustness even under challenging background illumination. In addition to the class and depth estimates, the proposed algorithm also provides the uncertainty measure associated with each parameter of interest. Besides being able to classify a scene w.r.t to defined spectral signatures, this algorithm can also complete a target detection task where objects of interest are materials with a specific spectral signature. This enables detecting objects which have a particular spectral signature under turbid media, e.g., metallic object detection underwater. Despite assuming one surface per-pixel, this algorithm can be used as a pre-processing step for multiple-surface-per-pixel to discard irrelevant pixels. Future work will investigate the use of this algorithm within an adaptive sampling strategy to improve Lidar data acquisition.

ACKNOWLEDGEMENTS

This work was supported by the UK Royal Academy of Engineering Research Fellowship Scheme (RF/201718/17128), and EPSRC Grants EP/T00097X/1, EP/S000631/1, EP/S026428/1, the Dasa project DSTLX1000147844 and the MOD University Defence Research Collaboration (UDRC) in Signal Processing. We thank the single-photon group led by Prof. G. S. Buller for providing the real single-photon data used in this work.

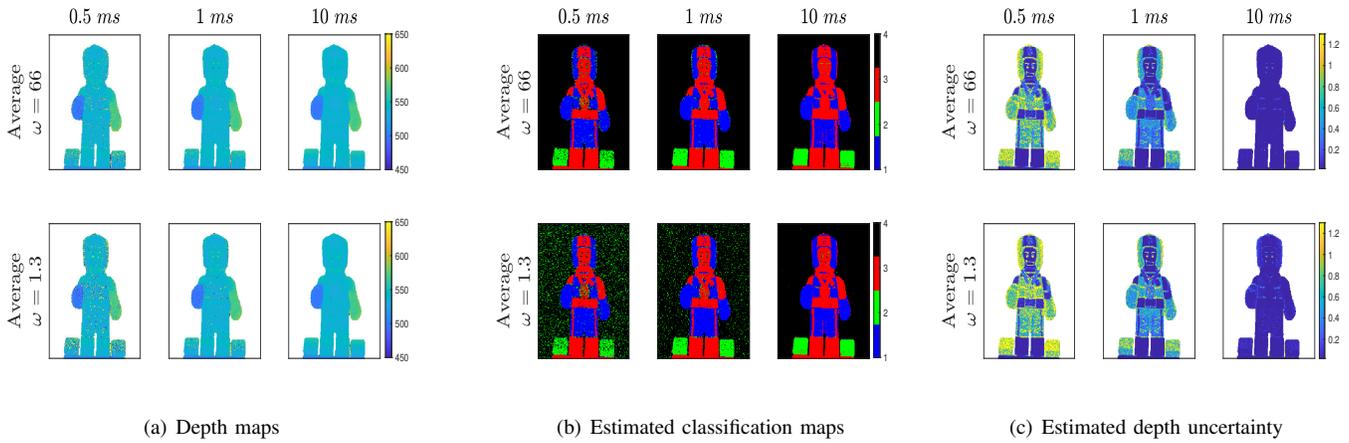


Fig. 2. Results of the proposed algorithm for different acquisition times and two SBR levels. (a) estimated depth maps (white pixels indicate no depth), (b) estimated classes, (c) the negative-log cumulative density (NCD) for $\epsilon = 1.5\text{mm}$ (large values indicate high uncertainty).

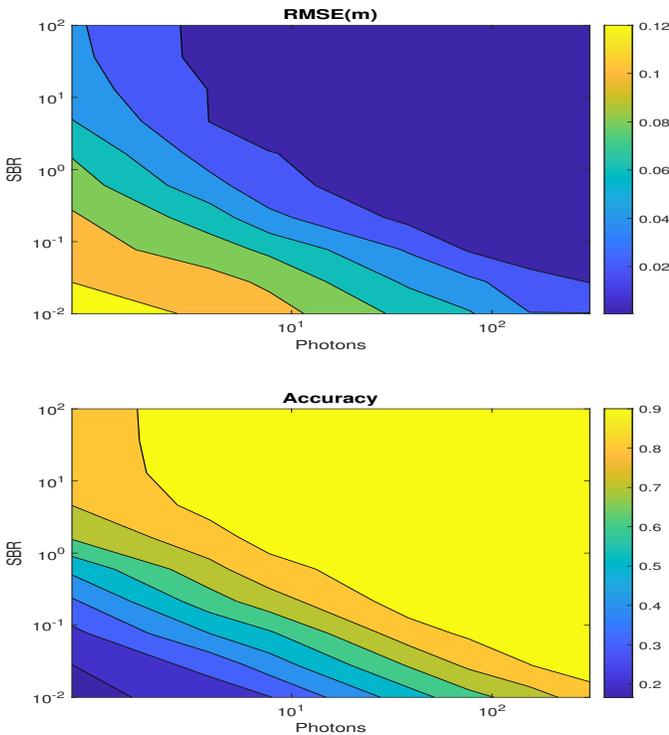


Fig. 3. Depth RMSEs (up) and Classification accuracy (down) of the sub-sampled lego data ($N = 40 \times 40$, $T = 1500$, $L = 4$, $K = 3$) w.r.t. signal-to-background ratio (SBR) and the average signal photons per pixel and per wavelength.

REFERENCES

[1] A. M. Wallace, A. Halimi, and G. S. Buller, "Full waveform lidar for adverse weather conditions," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 7, pp. 7064–7077, 2020.

[2] A. M. Wallace, A. McCarthy, C. J. Nichol, X. Ren, S. Morak, D. Martinez-Ramirez, I. H. Woodhouse, and G. S. Buller, "Design and evaluation of multispectral lidar for the recovery of arboreal parameters," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 8, pp. 4942–4954, 2013.

[3] A. McCarthy, R. J. Collins, N. J. Krichel, V. Fernández, A. M. Wallace, and G. S. Buller, "Long-range time-of-flight scanning sensor based on high-speed time-correlated single-photon counting," *Applied optics*, vol. 48, no. 32, pp. 6241–6251, 2009.

[4] A. M. Pawlikowska, A. Halimi, R. A. Lamb, and G. S. Buller, "Single-

photon three-dimensional imaging at up to 10 kilometers range," *Optics express*, vol. 25, no. 10, pp. 11919–11931, 2017.

[5] Z. Li, E. Wu, C. Pang, B. Du, Y. Tao, H. Peng, H. Zeng, and G. Wu, "Multi-beam single-photon-counting three-dimensional imaging lidar," *Optics express*, vol. 25, no. 9, pp. 10 189–10 195, 2017.

[6] A. Maccarone, A. McCarthy, X. Ren, R. E. Warburton, A. M. Wallace, J. Moffat, Y. Petillot, and G. S. Buller, "Underwater depth imaging using time-correlated single-photon counting," *Optics express*, vol. 23, no. 26, pp. 33 911–33 926, 2015.

[7] A. Halimi, A. Maccarone, A. McCarthy, S. McLaughlin, and G. S. Buller, "Object depth profile and reflectivity restoration from sparse single-photon data acquired in underwater environments," *IEEE Transactions on Computational Imaging*, vol. 3, no. 3, pp. 472–484, 2017.

[8] G. Satat, M. Tancik, and R. Raskar, "Towards photography through realistic fog," in *2018 IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2018, pp. 1–10.

[9] Y. Altmann, A. Maccarone, A. Halimi, A. McCarthy, G. Buller, and S. McLaughlin, "Efficient range estimation and material quantification from multispectral lidar waveforms," in *2016 Sensor Signal Processing for Defence (SSPD)*. IEEE, 2016, pp. 1–5.

[10] A. Halimi, R. Tobin, A. McCarthy, J. Bioucas-Dias, S. McLaughlin, and G. S. Buller, "Robust restoration of sparse multidimensional single-photon lidar images," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 138–152, 2019.

[11] J. Tachella, Y. Altmann, M. Márquez, H. Arguello-Fuentes, J.-Y. Tourneret, and S. McLaughlin, "Bayesian 3d reconstruction of subsampled multispectral single-photon lidar signals," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 208–220, 2019.

[12] A. Halimi, A. Maccarone, R. Lamb, G. S. Buller, and S. McLaughlin, "Robust and guided bayesian reconstruction of single-photon 3d lidar data: Application to multispectral and underwater imaging," *arXiv preprint arXiv:2103.10122*, 2021.

[13] Y. Altmann, A. Maccarone, A. McCarthy, G. Buller, and S. McLaughlin, "Joint range estimation and spectral classification for 3d scene reconstruction using multispectral lidar waveforms," in *2016 IEEE Statistical Signal Processing Workshop (SSP)*. IEEE, 2016, pp. 1–5.

[14] Y. Altmann, A. Maccarone, A. McCarthy, G. Newstadt, G. S. Buller, S. McLaughlin, and A. Hero, "Robust spectral unmixing of sparse multispectral lidar waveforms using gamma markov random fields," *IEEE Transactions on Computational Imaging*, vol. 3, no. 4, pp. 658–670, 2017.

[15] J. Tachella, Y. Altmann, S. McLaughlin, and J.-Y. Tourneret, "On fast object detection using single-photon lidar data," in *Wavelets and Sparsity XVIII*, vol. 11138. International Society for Optics and Photonics, 2019, p. 111380T.

[16] A. Halimi, A. Wallace, G. S. Buller, and S. McLaughlin, "Fast surface detection using single-photon detection events," in *2020 Sensor Signal Processing for Defence Conference (SSPD)*. IEEE, 2020, pp. 1–5.

[17] S. Hernandez-Marin, A. M. Wallace, and G. J. Gibson, "Bayesian analysis of lidar signals with multiple returns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 12, pp. 2170–2180, 2007.

[18] J. Rapp and V. K. Goyal, "A few photons among many: Unmixing signal and noise for photon-efficient active imaging," *IEEE Trans. Comput. Imaging*, vol. 3, no. 3, pp. 445–459, Sept. 2017.

[19] R. Tobin, Y. Altmann, X. Ren, A. McCarthy, R. A. Lamb, S. McLaughlin, and G. S. Buller, "Comparative study of sampling strategies for sparse photon multispectral lidar imaging: towards mosaic filter arrays," *Journal of Optics*, vol. 19, no. 9, p. 094006, 2017.

[20] Y. Altmann, X. Ren, A. McCarthy, G. S. Buller, and S. McLaughlin, "Robust bayesian target detection algorithm for depth imaging from sparse single-photon data," *IEEE Transactions on Computational Imaging*, vol. 2, no. 4, pp. 456–467, 2016.

Spaceborne SAR based assessment of nuclear test effects: the case of North Korea

Nicomino Fiscante
Department of Engineering
University of Study “Roma TRE”
 Rome, Italy
 nicomino.fiscante@uniroma3.it

Filippo Biondi
Department of Engineering
University of Study of “L’Aquila”
 Rome, Italy
 filippo.biondi@marina.difesa.it

Pia Addabbo
Department of Engineering
University “Giustino Fortunato”
 Benevento, Italy
 p.addabbo@unifortunato.eu

Carmine Clemente
Department of Electronic and Electrical Engineering
University of Strathclyde
 Glasgow, Scotland
 carmine.clemente@strath.ac.uk

Gaetano Giunta
Department of Engineering
University of Study of “Roma TRE”
 Rome, Italy
 gaetano.giunta@uniroma3.it

Danilo Orlando
Department of Engineering
University “Niccolò Cusano”
 Rome, Italy
 danilo.orlando@unicusano.it

Abstract—This study aims at estimating the Earth surface deformations due to the nuclear tests carried out by Democratic Peoples Republic of Korea from the 3rd of September 2017 by processing a time series of synthetic aperture radar images acquired by the COSMO-SkyMed satellite constellation. For active satellite sensors working in the X-band, phase information can be unreliable if scenarios with dense vegetation are observed. This uncertainty makes difficult to correctly estimate both the interferometric fringes and the phase delay generated by the variation in the space-time domain of the atmospheric parameters. To this end, in our research we apply the Sub-Pixel Tracking technique, so that the displacement information is extrapolated during the coregistration process. Our results demonstrate the capability of spaceborne remote sensing to help characterize large underground nuclear tests in revealing an accurate estimate of the spatial displacement due to the explosions. The work also reveals the presence of a possible underground tunnel network.

Index Terms—Displacement-field, Nuclear tests, Persistent Scatterers Interferometry, SAR Interferometry, Synthetic Aperture Radar, Sub-pixel correlation, Sub-Pixel Offset Tracking.

I. INTRODUCTION

On the 3rd of September 2017, the Democratic People’s Republic of Korea (DPRK), or North Korea, announced the successful test of a thermonuclear device. Different seismological agencies reported body wave magnitudes of well above 6.0, consequently estimating the explosive yields a quantity of the order of hundreds of kilotons [1]. Earthquakes induced by underground nuclear explosions, given the enormous amount of energy emitted, can produce considerable ground displacements that could be estimated using satellite synthetic aperture radar (SAR) data.

The most-widely used technique for displacement estimation is based on the persistent scatterers interferometry SAR (PS-InSAR) [2]. Unfortunately, this method has some limitations, such as the need of many interferometric images to obtain a reliable atmospheric phase screen and it cannot be applied to areas completely covered by vegetation.

A different approach exploits Differential InSAR (DInSAR) inversion strategy assisted by Small Baseline Subset (SBAS) [3]. Results demonstrate the effectiveness of this technique to retrieve the full 4-D displacement field associated, but in [4] it has been shown that SBAS method has also a number of limitations such as: geometric and temporal decorrelation, scale constraints, a limit on the spatial displacement gradient, geometric distortions and assumptions of linearity in the displacement process.

Sub-Pixel Offset Tracking (SPOT) is a promising alternative method for Earth displacement estimation. This technique has previously been applied to different scenarios overcoming technical defects and limitations of conventional DInSAR techniques [5]. In this paper, we apply the SPOT technique for estimating the Earth displacement caused by the underground nuclear explosions occurred from the 3rd of September 2017 in North Korea. A temporal series of interferometric SAR images, acquired from the COSMO-SkyMed (CSK) satellite constellation, are processed over a large area and highlighting a significant displacement. The results demonstrates the effectiveness of the method in estimating the location where the explosions began.

The outline of this paper is the following. In Section II a description of the nuclear test site is provided. In Section III the applied methodology to estimate the surface displacement is explained while in Section IV the experimental results are illustrated and discussed.

II. THE PUNGGYE-RI NUCLEAR TEST SITE

North Korea established its only known underground nuclear test site, shown in Figure 1, 17 kilometers north of the village of Punggye-ri at the foot of Mt. Mantap (2,209 m) in North Hamgyong Province. Between October 2006 and September 2017, six declared underground tunnel-emplaced nuclear explosive tests have been conducted from this site. All of these events generated seismic waves that were recorded at

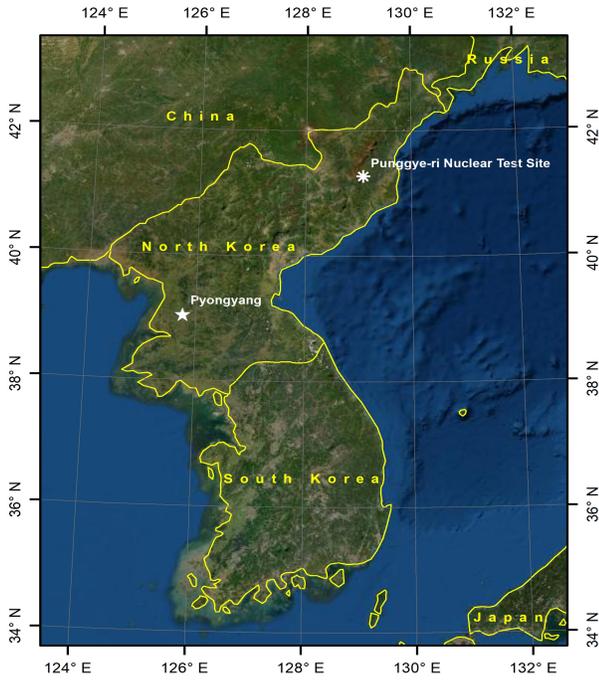


Fig. 1. Geolocation of the Punggye-ri nuclear test site.

regional and teleseismic distances and all took place within a few km of each other at the Punggye-ri nuclear facility, which is visible on the Google Earth platform at geographical coordinate: 41.279°N, 129.087°E.

Specifically, the best source of information on North Korea’s nuclear capabilities are the distant rumblings of nuclear tests detected by seismic monitoring stations belonging to the Comprehensive Test Ban Treaty Organization (CTBTO)’s International Monitoring System (IMS), the United States Geological Survey (USGS), and the Norwegian Seismic Array (NORSAR) organization with a mandate to monitor for nuclear explosions. Seismologists measure the magnitude of the body wave of these rumblings and, by employing several methods that use the seismological findings to facilitate the distinction between explosion and tectonic events, are able to understand if a seismic phenomenon is due to an explosion [6]. For scientific communities concerned with seismic monitoring with respect to nuclear surveillance, such as the CTBTO, the most pertinent issues that must be addressed are location determination, source discrimination, and yield estimate. This information is vital in understanding the DPRK’s progress towards nuclear armament, and it is the obligation of these non governmental organisations to address the issues outlined above for global nuclear security [1, 7]. The above mentioned organizations were able to retrieve information about six explosive events, located at the North Korean site reported in Table I.

On the 24th of May 2018 the DPRK invited a small delegation of international journalists at the Punggye-ri Nuclear Test Site to document the destruction of its underground tunnels and buildings. When visiting journalists were briefed on the dismantlement operations [8], officials from North Korea’s

Nuclear Weapons Institute spoke on the nuclear tests site and it was displayed a large scale topographic map. Figure 2 shows the areas of the Punggye-ri nuclear facility that were about to be demolished. Specifically, the map provides information on the tunnel entrance locations, the trajectory of tunnel excavations, along with what was claimed to be the layout of the two tunnels at the East and North Portals used to conduct North Korea’s six acknowledged underground nuclear tests. It also showed the two other tunnels prepared for, but never used, nuclear testing at the South and West Portals. The map appears to be a “to scale” schematic and the locations appear to be very precise and correlate well with most remotely derived geolocations for all tests [1, 6, 7], identified in sequence numbered one to six. However, the map does not appear to be a detailed engineering diagram in terms of layout of the tunnels as there is no hint to the specific “zig zag” layout and “fish hook” termination as expected configuration by a multitude of recent news reports [9]. In general, the map corroborates closely with seismic data and satellite imagery analysis of the test geolocations and tunnel directions [10].

III. METHODOLOGY

Sub-Pixel Offset Tracking is a largely used technique to measure large scale ground displacements in both range and azimuth directions and it is complementary to DInSAR and PS-InSAR methodologies in the case of radar phase information instability [11]. The total displacement of the sub-pixel normalized cross-correlation is described by the chirp-Doppler complex quantity $D^{(i,j)}$ [11, 12], for the (i, j) th master and slave interferometric pair. More precisely, considering a temporal series of N interferometric pairs, the total offset can be expressed as:

$$D^{(i,j)} = D_{\text{disp}}^{(i,j)} + D_{\text{topo}}^{(i,j)} + D_{\text{orb}}^{(i,j)} + D_{\text{cont}}^{(i,j)} + D_{\text{atm}}^{(i,j)} + D_{\text{noise}}^{(i,j)} \quad \text{for } (i, j) = 1, \dots, N, \quad (1)$$

where:



Fig. 2. Dismantlement operations for the Punggye-ri Nuclear Test Site. The map layout of the nuclear facility with the declared epicenters of the six underground tests.

TABLE I
MAIN INFORMATION AND DATA COLLECTED ABOUT THE NORTH KOREA'S NUCLEAR TESTS SERIES.

Event	Date	GPS coordinate ¹	Depth (meters) ¹	Quake (magnitude)			Yields (kilotons) ¹		
				USGS	CTBTO	NORSAR	USGS	CTBTO	NORSAR
DPRK 1	09/10/2006	41.294° N, 129.094° E	438	4.3	4.0	4.0	1.00	0.40	0.40
DPRK 2	25/05/2009	41.303° N, 129.037° E	556	4.7	4.5	4.5	3.41	1.85	1.85
DPRK 3	12/02/2013	41.299° N, 129.004° E	576	5.1	4.9	5.0	11.66	6.31	8.58
DPRK 4	06/01/2016	41.300° N, 129.047° E	772	5.1	4.9	4.8	11.66	6.31	4.64
DPRK 5	09/09/2016	41.287° N, 129.078° E	714	5.3	5.1	5.1	21.54	11.66	11.66
DPRK 6	03/09/2017	41.343° N, 129.036° E	798	6.3	6.1	6.1	464.16	251.19	251.19

¹ Estimated on seismic signal.

- $\mathbf{D}_{\text{disp}}^{(i,j)}$ is the quantity of interest to be estimated and indicates the possible spatial displacement;
- $\mathbf{D}_{\text{topo}}^{(i,j)}$ represents the residual topography induced phase due to a non perfect knowledge of the actual height profile;
- $\mathbf{D}_{\text{orb}}^{(i,j)}$ accounts residual errors in satellite orbital data;
- $\mathbf{D}_{\text{cont}}^{(i,j)}$ accounts attitude and control errors of the flying satellite trajectory;
- $\mathbf{D}_{\text{atm}}^{(i,j)}$ denotes the phase components due to the change in the atmospheric and ionospheric dielectric constant between the two master/slave acquisitions;
- and $\mathbf{D}_{\text{noise}}^{(i,j)}$ accounts for decorrelation phenomena such as: spatial, temporal, thermal, etc

Therefore, solving (1) with respect to $\mathbf{D}_{\text{disp}}^{(i,j)}$, would lead to a compensation of all the other offset components as described in [13, 14]. Writing $\mathbf{D}^{(i,j)} = |\mathbf{D}^{(i,j)}| \exp(j\theta^{(i,j)})$, where the magnitude is the distance between the master and slave is estimated by two-dimensional cross-correlation and the phase, considered respect to the horizontal axis, is an harmonic parameter indicating the direction of displacement [15].

Figure 3 shows the applied processing workflow, which is made by several consecutive steps. Specifically, starting from the image databases where the master and slave SAR images are stored, the first step performs the coregistration of a single interferometric pair. The next steps perform the displacement estimation and the geocoding process, respectively. The workflow is repeated for each $(i, j) = 1, \dots, N$ interferometric pair and then averaged.

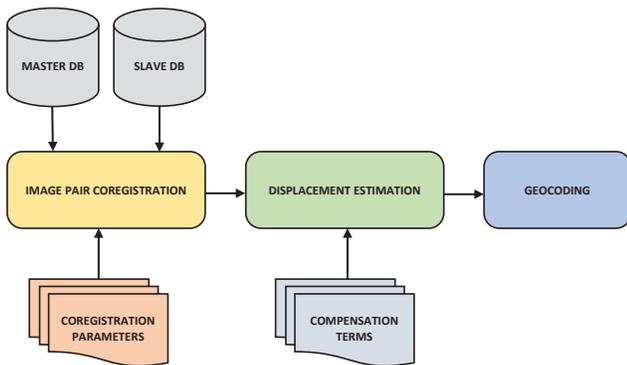


Fig. 3. Applied processing chain workflow.

IV. EXPERIMENTAL RESULTS

The experimental results are derived by processing the interferometric time series of SAR Spotlight data acquired by CSK satellite constellation. Specifically, we consider 27 images, of which 13 are pre-events and 14 are post-events. In this study we are not interested in the estimation of a displacement time series because a nuclear explosion occurs in a very small interval of time and we assume that after this event no other geological force caused significant movement of the Earth's surface. The main objective is to analyse the displacement generated by the enormous forces due to nuclear blasting. Through the geocoded data it is possible to estimate the position where the explosion occurred. In addition, we try to analyse a possible network of underground tunnels characterizing the whole test site.

Figure 4 shows two subsidence maps. More precisely, Figure 4 (top) is the output obtained by processing only the SAR images observed before the DPRK 1 nuclear test, and it confirms that no significant subsidence phenomena are present. Moreover, Figure 4 (bottom) shows the output obtained processing the SAR images acquired after the DPRK-6 event. In this result a very interesting subsidence phenomenon, that covers an area of about 3.5 kmq, can be observed. The displacement generated by the explosions can be seen as it is possible to observe that the north side of Mt. Mantap is in subsidence, characterized by a physical movement of the Earth's surface downwards of about -20 cm. Conversely, the south side of the mountain has a positive subsidence of about $+50$ cm, which means a rise of the Earth's surface.

Figure 5 illustrates in detail the subsidence effect caused by nuclear explosions on Mt. Mantap; in fact, it is possible to identify the chambers used in the experiment of the 3rd of September 2017 and in the previous ones. Specifically, Figure 5 (top) shows the possible positions of the six nuclear tests, in accordance with Table I, as well as the respective two access tunnels from the North Portal and the East Portal respectively. Figure 5 (bottom) shows in detail the DPRK 6 displacement location. This analysis is particularly interesting because it is possible to observe, considering the maximum value of the estimated displacement, the position where the detonation of the nuclear weapon took place. Furthermore, analyzing in detail the chamber explosion of the DPRK 6 event it is possible to note that the displacement has a main lobe with an asymmetric profile. This asymmetry could be

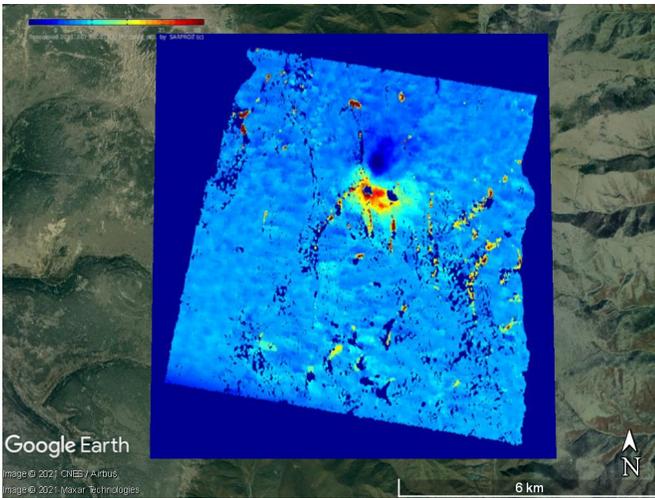
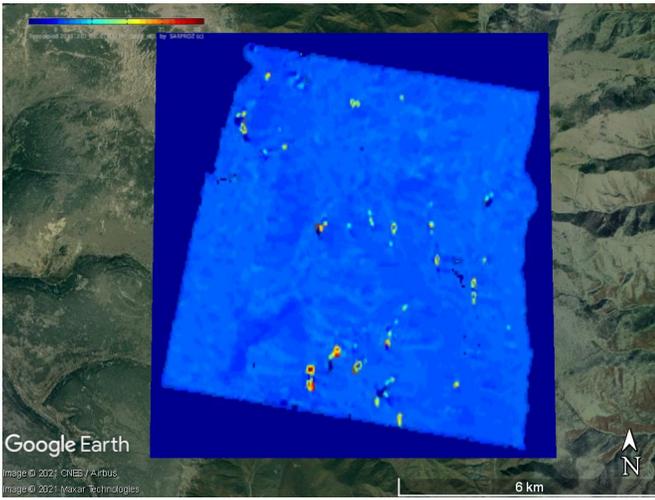


Fig. 4. Global view of the main explosion area (top), and a detailed view of the DPRK-6 site (bottom). Amplitude values are in centimetres.

due to different factors such as: a not perfect uniformity of the explosion or the internal material compositions of Mt. Mantap. Figure 6 illustrates the displacement-field indicating both direction and intensity for DPRK 5 and DPRK 6 events. This representation is useful to understand the origin of the force field that generated the displacement, i.e., the position of the nuclear device, and to give interesting information about the tunnel networks.

V. CONCLUSIONS

This study has the objective of estimating the Earth deformations due to the underground nuclear test carried out by North Korea after the 3rd of September 2017 by processing time series of multitemporal interferometric SAR images observed by the CSK satellite constellation. The area of interest is the Punggye-ri nuclear facility that has housed all six of North Korea’s nuclear tests. In this work we apply the SPOT technique to estimate the displacement map to overcome the

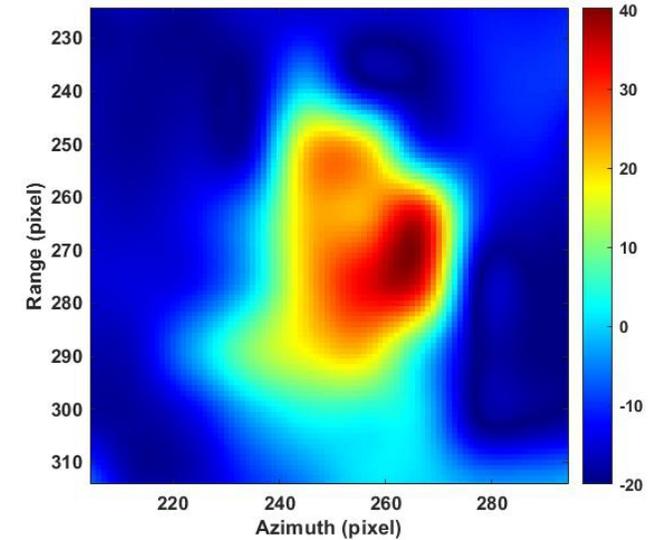
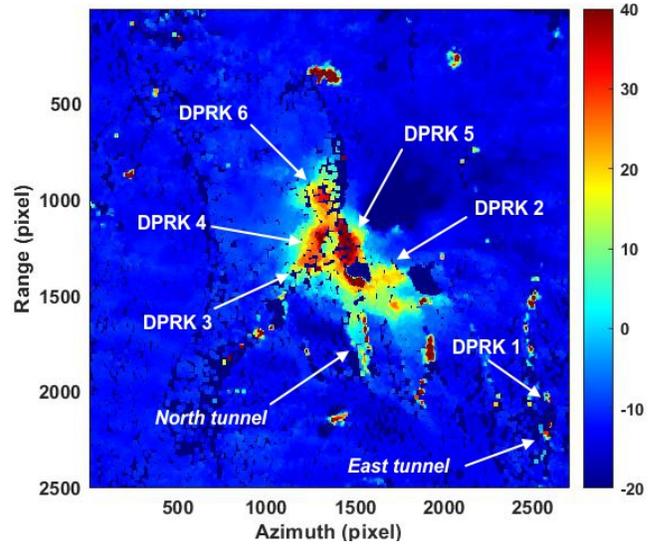


Fig. 5. Global view of the main explosion area (top), and a detailed view of the DPRK 6 site (bottom). Amplitude values are in centimetres.

main limitations encountered by the widely-used techniques that process data acquired by SAR sensors in the X band [2–4]. The surface deformation map shows significant and complex ground displacement located mainly on Mt. Mantap, caused by explosions, collapse, compaction and landslides. The results reveal the nuclear explosion locations, the effect on the surface and the underground network of tunnels which stems from the generated displacement field.

REFERENCES

- [1] Incorporated Research Institutions for Seismology, “Special Event: 2017 north korean nuclear test,” Sept. 2017. [Online]. Available: <https://ds.iris.edu/ds/nodes/dmc/specialevents/2017/09>
- [2] A. Ferretti, C. Prati, and F. Rocca, “Permanent scatterers in SAR interferometry,” *IEEE Transactions on*

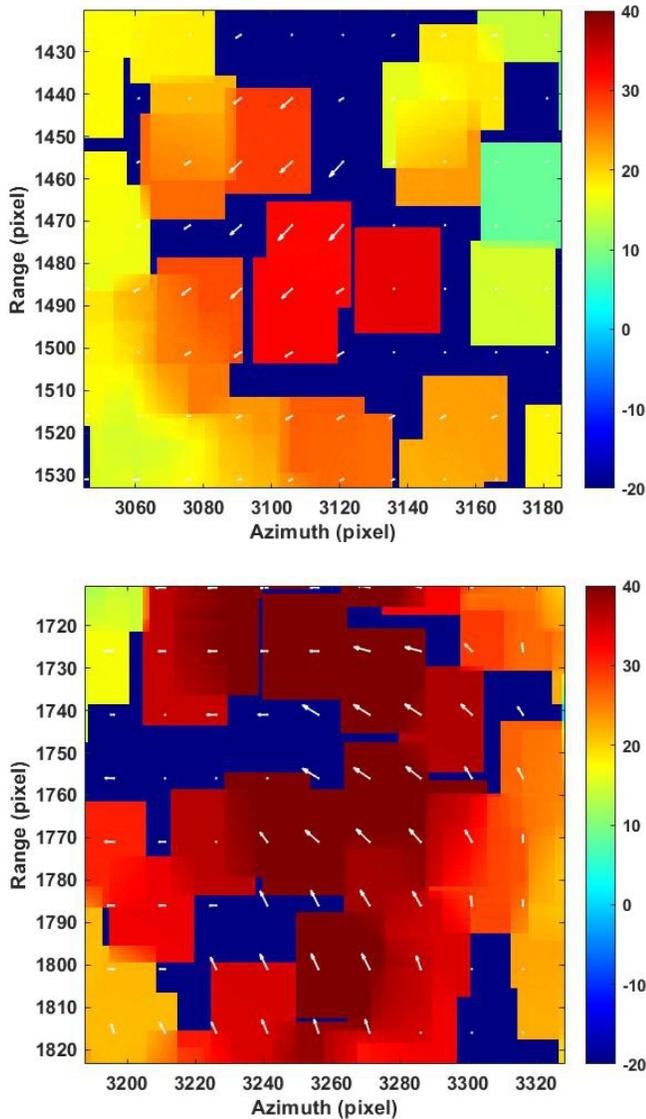


Fig. 6. Displacement-field over the DPRK 5 event location (top) and over the DPRK 6 event location (bottom). Amplitude values are in centimetres.

- Geoscience and Remote Sensing*, vol. 39, no. 1, pp. 8–20, 2001.
- [3] P. Berardino, G. Fornaro, R. Lanari, and E. Sansosti, “A new algorithm for surface deformation monitoring based on small baseline differential SAR interferograms,” *IEEE Transactions on geoscience and remote sensing*, vol. 40, no. 11, pp. 2375–2383, 2002.
 - [4] A. Singleton, Z. Li, T. Hoey, and J.-P. Muller, “Evaluating Sub-Pixel Offset Techniques as an alternative to D-InSAR for monitoring episodic landslide movements in vegetated terrain,” *Remote Sensing of Environment*, vol. 147, pp. 133–144, 2014.
 - [5] R. Michel, J.-P. Avouac, and J. Taboury, “Measuring ground displacements from sar amplitude images: Application to the landers earthquake,” *Geophysical Research Letters*, vol. 26, no. 7, pp. 875–878, 1999.
 - [6] S. Gibbons, F. Pabian, S. Näsholm, T. Kvaerna, and S. Mykkeltveit, “Accurate relative location estimates for the North Korean nuclear tests using empirical slowness corrections,” *Geophysical Journal International*, vol. 208, no. 1, pp. 101–117, 10 2016. [Online]. Available: <https://doi.org/10.1093/gji/ggw379>
 - [7] T. Wang, Q. Shi, M. Nikkhoo, S. Wei, S. Barbot, D. Dreger, R. Bürgmann, M. Motagh, and Q.-F. Chen, “The rise, collapse, and compaction of Mt. Mantap from the 3 september 2017 North Korean nuclear test,” *Science*, vol. 361, no. 6398, pp. 166–170, 2018. [Online]. Available: <https://science.sciencemag.org/content/361/6398/166>
 - [8] M. Greenfield, “We’re back at the hotel after two 12 hour train journeys into the mountains. we spent 9 hours on Thurs at Punggye-ri test site, watching 3 tunnel entrances demolished by explosion. Our full report will be on Sky News at 6am UK. A few pics:,” Twitter user @ SkyGreenfield, May 5, 2018. [Online]. Available: twitter.com/skygreenfield/status/999803019804925958
 - [9] 38 North, “North Korea’s Punggye-ri Nuclear Test Site: analysis reveals its potential for additional testing with significantly higher yields.” [Online]. Available: <https://www.38north.org/2017/03/punggye031017>
 - [10] J. Palmer and N. Selby, “Recreating and exploring the dprk nuclear test site in 3D to calculate possible overburdens for detonations,” 2019.
 - [11] F. Biondi, P. Addabbo, D. Orlando, and C. Clemente, “Micro-motion estimation of maritime targets using pixel tracking in cosmo-skymed synthetic aperture radar data—an operative assessment,” *Remote Sensing*, vol. 11, no. 14, p. 1637, 2019.
 - [12] F. Biondi, “COSMO-SkyMed staring spotlight SAR data for micro-motion and inclination angle estimation of ships by pixel tracking and convex optimization,” *Remote Sensing*, vol. 11, no. 7, p. 766, 2019.
 - [13] S. L. Ullo, P. Addabbo, D. Di Martire, S. Sica, N. Fiscante, L. Cicala, and C. V. Angelino, “Application of DInSAR Technique to High Coherence Sentinel-1 Images for Dam Monitoring and Result Validation Through In Situ Measurements,” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 3, pp. 875–890, March 2019.
 - [14] F. Biondi, C. Clemente, and D. Orlando, “An atmospheric phase screen estimation strategy based on multichromatic analysis for differential interferometric synthetic aperture radar,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 7269–7280, 2019.
 - [15] E. Sansosti, P. Berardino, M. Manunta, F. Serafino, and G. Fornaro, “Geometrical SAR image registration,” *IEEE Transactions on Geoscience and Remote Sensing*, vol. 44, no. 10, pp. 2861–2870, 2006.

The Maximal Eigengap Estimator for Acoustic Vector-Sensor Processing

Robert Bassett

*Operations Research Department
Naval Postgraduate School
Monterey, CA 93943, USA
robert.bassett@nps.edu*

Jacob Foster

*Operations Research Department
Naval Postgraduate School
Monterey, CA 93943, USA
jacob.foster@nps.edu*

Kay L. Gemba

*Acoustics Division, Code 7160
U.S. Naval Research Laboratory
Washington, D.C. 20375, USA
kay.gemba@nrl.navy.mil*

Paul Leary

*Physics Department
Naval Postgraduate School
Monterey, CA 93943, USA
pleary@nps.edu*

Kevin B. Smith

*Physics Department
Naval Postgraduate School
Monterey, CA 93943, USA
kbsmith@nps.edu*

Abstract—This paper introduces the maximal eigengap estimator for finding the direction of arrival of a wideband acoustic signal using a single vector-sensor. We show that in this setting narrowband cross-spectral density matrices can be combined in an optimal weighting that approximately maximizes signal-to-noise ratio across a wide frequency band. The signal subspace resulting from this optimal combination of narrowband power matrices defines the maximal eigengap estimator. We discuss the advantages of the maximal eigengap estimator over competing methods, and demonstrate its utility in a real-data application using signals collected in 2019 from an acoustic vector-sensor deployed in the Monterey Bay.

Index Terms—direction of arrival, acoustic vector-sensor, signal subspace, eigengap

I. INTRODUCTION

Direction of Arrival (DOA) estimation of acoustic signals is a problem which spans multiple application areas. Examples include biology, where tracking marine mammals can provide detailed information on their habits, and defense, where monitoring and locating vessels has utility in many naval applications. One advantage of an acoustic vector-sensor over a conventional hydrophone array is that acoustic vector-sensors have a smaller footprint, while still providing signal direction. The specialized nature of data collected by an acoustic vector-sensor motivates signal processing techniques customized for its analysis.

In this paper, we consider DOA estimation of a single wideband source using a single acoustic vector-sensor. We introduce the *maximal eigengap estimator*, which combines cross-spectral density (CSD) matrices across a wide frequency range to maximize signal-to-noise ratio (SNR) of the resulting signal subspace. In the setting of a single acoustic vector sensor and a single source, we provide a tractable formulation

All authors acknowledge support from the Office of Naval Research, Grants No. N0001420WX01523, N0001421WX01634, and N0001421WX00410. We are also grateful to the Monterey Bay Aquarium Research Institute (MBARI) for their role in collecting the data used in this paper.

of the maximal eigengap estimator, thus providing a new DOA estimation method customized for this setting.

Previous work on DOA estimation of wideband signals aggregates narrowband information across a wide frequency band in a variety of ways. In [1], CSD matrices for each frequency are combined via linear combination, where the ideal weighting of each matrix is given by the SNR in that frequency. Because this SNR is unknown, the authors take an equal weighting of each cross-spectral matrix. The authors of [2] instead aggregate narrowband frequency information by forming an estimating equation for the signal subspace which is a linear combination of generalized eigenvector equations for each frequency. It is suggested that the weighting should be related to SNR in each frequency (this suggestion is confirmed in an analysis of asymptotic variance [3]), but numerical examples suggest that a uniform weighting works equally well. Other popular methods combine signal or noise subspace information uniformly [4], [5]. Our contribution, the maximal eigengap estimator, continues this pattern of aggregating narrowband information, and is similar in theme to [1], with the important difference that we capitalize on the acoustic vector-sensor setting to *optimize* SNR over the weights in a linear combination of cross-spectral power matrices.

This paper is organized as follows. In the next section, we formally introduce the maximal eigengap estimator and our main theoretical contribution, Theorem 1, which provides a tractable reformulation of the estimator. Section III applies the maximal eigengap estimator to signals collected in 2019 by an acoustic vector-sensor deployed in the Monterey Bay, where ground-truth information on vessel locations provide a realistic test case. We conclude by summarizing our results.

Before proceeding we establish some notation. We denote vectors and matrices by bold text, and scalars by plain text. For a vector or matrix \mathbf{a} , we denote its transpose by \mathbf{a}^T and its conjugate transpose by \mathbf{a}^H . Similarly, we denote the complex conjugate of a scalar and pointwise conjugate of vector/matrix

with superscript $*$. Minimal and maximal eigenvalues of a Hermitian matrix \mathbf{a} (which are real by the spectral theorem [6]) are denoted $\lambda_{\min}(\mathbf{a})$ and $\lambda_{\max}(\mathbf{a})$, with corresponding eigenvectors $\mathbf{v}_{\min}(\mathbf{a})$ and $\mathbf{v}_{\max}(\mathbf{a})$. The expectation operator is written $\mathbb{E}[\cdot]$. The Frobenius norm of a matrix \mathbf{a} is denoted $\|\mathbf{a}\|$, and $\|\cdot\|$ denotes an arbitrary norm.

II. MAXIMAL EIGENGAP ESTIMATOR

A. Signal Model

Consider a single acoustic vector-sensor with a single signal source. The sensor's output at time t , $\mathbf{z}(t)$, has four channels, consisting of an omnidirectional hydrophone and three particle velocity measurements. We assume a plane wave signal, where it can be shown (see [7]) that the scaled sensor output is

$$\mathbf{z}(t) = \mathbf{w} s(t) + \mathbf{n}(t), \quad (1)$$

where $s(t)$ is the acoustic pressure at the sensor at time t , $\mathbf{w} = (1, k_x, k_y, k_z)^T$ is a vector such that $\mathbf{k} = (k_x, k_y, k_z)^T$ is a unit vector pointing towards the stationary source, and $\mathbf{n}(t) \in \mathbb{R}^4$ is a noise term. The vector \mathbf{k} can be written $(\cos \theta \cos \phi, \sin \theta \cos \phi, \sin \phi)^T$, where θ and ϕ give the azimuth and elevation angles, respectively, of \mathbf{k} . Throughout, we focus our attention on estimating the azimuth angle θ using the x and y velocity channels, setting $\phi = 0$. This reflects that azimuth is the primary quantity of interest in many DOA estimation problems.

Denote by $\mathbf{x}(t)$, $\mathbf{e}(t)$, and \mathbf{u} , the restriction of $\mathbf{z}(t)$, $\mathbf{n}(t)$, and \mathbf{w} , respectively, to their second (x) and third (y) channels. Because we assume $\phi = 0$, \mathbf{u} is a unit vector. Equation (1) restricted to x and y velocity channels is then

$$\mathbf{x}(t) = \mathbf{u} s(t) + \mathbf{e}(t). \quad (2)$$

A frequency domain representation of (2) is

$$\mathbf{X}(\omega) = \mathbf{u} S(\omega) + \mathbf{E}(\omega). \quad (3)$$

Denote $\mathbb{E}[S(\omega)^2]$ by $P_S(\omega)$, and $\mathbb{E}[\mathbf{E}(\omega)\mathbf{E}(\omega)^H]$ by $\Sigma(\omega)$. Assume that

- (A1) $\mathbb{E}[S(\omega)\mathbf{E}(\omega)^*] = \mathbb{E}[S(\omega)^*\mathbf{E}(\omega)] = 0$ for each ω .
- (A2) The condition number of $\Sigma(\omega)$ is bounded by some constant C , uniformly in ω .

$$\frac{\lambda_{\max}(\Sigma(\omega))}{\lambda_{\min}(\Sigma(\omega))} \leq C$$

Assumption (A1) allows us to form the CSD matrix of $\mathbf{X}(\omega)$, denoted $\mathbf{P}_{\mathbf{X}}(\omega)$, as

$$\begin{aligned} \mathbf{P}_{\mathbf{X}}(\omega) &= \mathbb{E}[\mathbf{X}(\omega)\mathbf{X}(\omega)^H] \\ &= P_S(\omega)\mathbf{u}\mathbf{u}^T + \Sigma(\omega). \end{aligned} \quad (4)$$

When $P_S(\omega)$ dominates $\Sigma(\omega)$, it can be shown that the maximal eigenvector $\mathbf{v}_{\max}(\mathbf{P}_{\mathbf{X}}(\omega))$ is close to \mathbf{u} . Moreover, if the spatial covariance $\Sigma(\omega)$ is a scalar multiple of the identity matrix this recovery is exact. A precise statement is given by the following proposition, which follows from a direct application of [8, Theorem 8.5].

Proposition 1. *Assume $P_S(\omega) > 0$ and $\|\Sigma(\omega)\| < P_S(\omega)/2$. Let \mathbf{u}^\perp be a unit vector perpendicular to \mathbf{u} , $\mathbf{U} = [\mathbf{u}, \mathbf{u}^\perp]$ a matrix with columns \mathbf{u} and \mathbf{u}^\perp , and $\tilde{p}(\omega)$ the off-diagonal term in $\mathbf{U}^T \Sigma(\omega) \mathbf{U}$. Then*

$$\|\mathbf{v}_{\max}(\mathbf{P}_{\mathbf{X}}(\omega)) - \mathbf{u}\|_2 \leq \frac{2|\tilde{p}(\omega)|}{P_S(\omega) - 2\|\Sigma(\omega)\|}. \quad (5)$$

By the bound given in (5), we have tighter control on the difference of $\mathbf{v}_{\max}(\mathbf{P}_{\mathbf{X}}(\omega))$ and the signal's DOA \mathbf{u} when either

- (i) The maximum eigenvector of $\Sigma(\omega)$ is closely aligned with \mathbf{u} , in the sense that $|\mathbf{v}_{\max}(\Sigma(\omega))^T \mathbf{u}^\perp|$ (which directly controls $|\tilde{p}(\omega)|$) is small.
- (ii) The signal dominates the noise, in the sense that $P_S(\omega)$ is large and $\Sigma(\omega)$ is small (in Frobenius norm). Recalling that $\|\Sigma(\omega)\| = \sqrt{\lambda_{\max}(\Sigma(\omega))^2 + \lambda_{\min}(\Sigma(\omega))^2}$, we can alternatively insist that these eigenvalues are small.

Using assumption (A2), we can upper bound $\|\Sigma(\omega)\|$ as

$$\|\Sigma(\omega)\| \leq \lambda_{\min}(\Sigma(\omega)) \sqrt{1 + C^2},$$

and the denominator in (5)

$$P_S(\omega) - 2\|\Sigma(\omega)\| \geq P_S(\omega) - 2\lambda_{\min}(\Sigma(\omega)) \sqrt{1 + C^2}. \quad (6)$$

From (5) and (6), we see that large $P_S(\omega)$ and small $\lambda_{\min}(\Sigma(\omega))$ make $\mathbf{v}_{\max}(\mathbf{P}_{\mathbf{X}}(\omega))$ a better estimate of \mathbf{u} . However, note that both $P_S(\omega)$ and $\lambda_{\min}(\Sigma(\omega))$ are properties of unobserved variables. From (4), we note that $\lambda_{\max}(\mathbf{P}_{\mathbf{X}}(\omega))$ approximates $P_S(\omega)$ and $\lambda_{\min}(\mathbf{P}_{\mathbf{X}}(\omega))$ approximates $\lambda_{\min}(\Sigma(\omega))$. Hence our desire for large $P_S(\omega)$ and small $\lambda_{\min}(\Sigma(\omega))$ can be stated as the following.

Goal: for better estimation of the signal DOA, we should have $\lambda_{\max}(\mathbf{P}_{\mathbf{X}}(\omega))$ large and $\lambda_{\min}(\mathbf{P}_{\mathbf{X}}(\omega))$ small.

B. Combining Narrowband Information

For wideband signals, we seek to aggregate the narrowband information described in the previous section over a range of ω values, only some of which contain the signal of interest. For frequency bins $F = \{\omega_1, \dots, \omega_{|F|}\}$, our goal is to decide, for each $\omega \in F$, whether the signal of interest is contained in the narrowband CSD matrix $\mathbf{P}_{\mathbf{X}}(\omega)$, and if so to what degree it agrees with the signal DOA in other frequency bins. To accomplish this goal, we extend (4) by searching for a weighted combination of the power matrices that gives a tighter bound in (5). Namely, we propose solving, for some to-be-specified norm $\|\cdot\|$, the following problem, where $\mathbf{a} \in \mathbb{R}^{|F|}$ is a weight vector such that a_ω denotes the entry of \mathbf{a} corresponding to frequency bin $\omega \in F$.

$$\begin{aligned} \max_{\mathbf{a} \in \mathbb{R}^{|F|}} \lambda_{\max} \left(\sum_{\omega \in F} a_\omega \mathbf{P}_{\mathbf{X}}(\omega) \right) - \lambda_{\min} \left(\sum_{\omega \in F} a_\omega \mathbf{P}_{\mathbf{X}}(\omega) \right) \\ \text{s.t. } \|\mathbf{a}\| \leq 1, \quad \mathbf{a} \geq 0 \end{aligned} \quad (7)$$

In (7), the $\|\cdot\| \leq 1$ constraint defines the maximum size of the weight vector. Because the λ_{\max} term measures the power

in the signal subspace, and λ_{\min} term the power in the noise subspace, this objective maximizes the estimated SNR¹.

One concern with the formulation (7) is that some frequencies (in practice, often the lower ones) have more power in both signal and noise subspaces, so these terms have the potential to dominate the objective function. To remedy this, we propose three standardization schemes for the CSD matrices $\mathbf{P}_{\mathbf{X}}(\omega)$. We denote a standardized $\mathbf{P}_{\mathbf{X}}(\omega)$ by $\mathbf{Q}_{\mathbf{X}}(\omega)$. Three intuitive options for standardization are:

- (a) Take each matrix to have unit trace, so that

$$\mathbf{Q}_{\mathbf{X}}(\omega) = \frac{\mathbf{P}_{\mathbf{X}}(\omega)}{\text{trace}(\mathbf{P}_{\mathbf{X}}(\omega))},$$

and the total power (aggregated across both signal and noise subspaces) in each frequency bin is 1.

- (b) Take each matrix to have minimal eigenvalue 1, so that

$$\mathbf{Q}_{\mathbf{X}}(\omega) = \frac{\mathbf{P}_{\mathbf{X}}(\omega)}{\lambda_{\min}(\mathbf{P}_{\mathbf{X}}(\omega))},$$

and the maximal eigenvalue of the matrix gives its SNR.

- (c) No standardization, so that

$$\mathbf{Q}_{\mathbf{X}}(\omega) = \mathbf{P}_{\mathbf{X}}(\omega).$$

We compare the merits of each standardization method in section III.

Lastly, we replace $\mathbf{P}_{\mathbf{X}}(\omega)$ in (7) with an estimate $\hat{\mathbf{P}}_{\mathbf{X}}(\omega)$ obtained from observations. We extend the notation for standardized matrices as one would expect, so that $\hat{\mathbf{Q}}_{\mathbf{X}}(\omega)$ is the standardized $\hat{\mathbf{P}}_{\mathbf{X}}(\omega)$. Then (7) becomes

$$\begin{aligned} \max_{\mathbf{a} \in \mathbb{R}^{|F|}} \lambda_{\max} \left(\sum_{\omega \in F} a_{\omega} \hat{\mathbf{Q}}_{\mathbf{X}}(\omega) \right) - \lambda_{\min} \left(\sum_{\omega \in F} a_{\omega} \hat{\mathbf{Q}}_{\mathbf{X}}(\omega) \right) \quad (8) \\ \text{s.t. } \|\mathbf{a}\| \leq 1, \quad \mathbf{a} \geq 0. \end{aligned}$$

We define the *maximal eigengap estimator* as the real part of a maximum eigenvector, $\text{Re} \left\{ \mathbf{v}_{\max} \left(\sum_{\omega \in F} \hat{a}_{\omega} \hat{\mathbf{Q}}_{\mathbf{X}}(\omega) \right) \right\}$, where \hat{a} denotes the maximizer in (8). The following Theorem is instrumental in computing the maximal eigengap estimator.

Theorem 1. *The objective in (8) is the square root of a convex quadratic function in \mathbf{a} .*

Proof. Denote the entries of $\hat{\mathbf{Q}}_{\mathbf{X}}(\omega)$ by

$$\hat{\mathbf{Q}}_{\mathbf{X}}(\omega) = \begin{pmatrix} q_{\omega} & r_{\omega} \\ r_{\omega}^* & s_{\omega} \end{pmatrix},$$

where we have used the conjugate symmetry of $\hat{\mathbf{Q}}_{\mathbf{X}}(\omega)$ in the off-diagonal terms. The eigenvalues

$$\lambda_{\max} \left(\sum_{\omega \in F} a_{\omega} \hat{\mathbf{Q}}_{\mathbf{X}}(\omega) \right) \quad \text{and} \quad \lambda_{\min} \left(\sum_{\omega \in F} a_{\omega} \hat{\mathbf{Q}}_{\mathbf{X}}(\omega) \right)$$

¹In detail, maximizing difference in (7) is equivalent to maximizing $\log(\lambda_{\max}/\lambda_{\min})$ which is equivalent to maximizing the SNR $\lambda_{\max}/\lambda_{\min}$.

are given by the zeros of the characteristic polynomial for $\sum_{\omega \in F} a_{\omega} \hat{\mathbf{Q}}_{\mathbf{X}}(\omega)$,

$$\det \left(\begin{array}{cc} \sum_{\omega \in F} a_{\omega} q_{\omega} - \lambda & \sum_{\omega \in F} a_{\omega} r_{\omega} \\ \sum_{\omega \in F} a_{\omega} r_{\omega}^* & \sum_{\omega \in F} a_{\omega} s_{\omega} - \lambda \end{array} \right) = 0. \quad (9)$$

These roots can be computed explicitly using the quadratic formula, and their difference is the discriminant of the quadratic in (9), which is nonnegative because the eigenvalues of Hermitian matrices are real.

$$\sqrt{\left(\sum_{\omega \in F} a_{\omega} (q_{\omega} + s_{\omega}) \right)^2 - 4 \left(\left(\sum_{\omega \in F} a_{\omega} q_{\omega} \right) \left(\sum_{\omega \in F} a_{\omega} s_{\omega} \right) - \left| \sum_{\omega \in F} a_{\omega} r_{\omega} \right|^2 \right)} \quad (10)$$

Because we maximize this expression in (8), and it is always nonnegative, we can instead maximize the square of (10)

$$\left(\sum_{\omega \in F} a_{\omega} (q_{\omega} + s_{\omega}) \right)^2 - 4 \left(\left(\sum_{\omega \in F} a_{\omega} q_{\omega} \right) \left(\sum_{\omega \in F} a_{\omega} s_{\omega} \right) - \left| \sum_{\omega \in F} a_{\omega} r_{\omega} \right|^2 \right). \quad (11)$$

Expanding (11), we have

$$\begin{aligned} \sum_{\omega_i \in F} \sum_{\omega_j \in F} \left(a_{\omega_i} (q_{\omega_i} + s_{\omega_i}) (q_{\omega_j} + s_{\omega_j}) a_{\omega_j} \right. \\ \left. - 4 \left(a_{\omega_i} q_{\omega_i} s_{\omega_j} a_{\omega_j} - a_{\omega_i} r_{\omega_i} r_{\omega_j}^* a_{\omega_j} \right) \right) \quad (12) \end{aligned}$$

Define a matrix $\tilde{\mathbf{R}} \in \mathbb{C}^{|F| \times |F|}$ with

$$\tilde{\mathbf{R}}_{\omega_i, \omega_j} = (q_{\omega_i} + s_{\omega_i}) (q_{\omega_j} + s_{\omega_j}) - 4 \left(q_{\omega_i} s_{\omega_j} - r_{\omega_i} r_{\omega_j}^* \right), \quad (13)$$

so that the quadratic form $\mathbf{a}^T \tilde{\mathbf{R}} \mathbf{a}$ gives (12).

Next we apply a common technique to produce a real symmetric quadratic form \mathbf{R} which is equal to the quadratic form given by $\tilde{\mathbf{R}}$. First, note that expression (11) is *real* and *scalar*, because a_{ω} , and q_{ω} , and s_{ω} are real (recall that $\hat{\mathbf{Q}}_{\mathbf{X}}(\omega)$ is Hermitian). Hence (11) is equal to its conjugate, its transpose, and its conjugate transpose. Because (12) and (11) are equal, it follows that

$$\left(\mathbf{a}^T \tilde{\mathbf{R}} \mathbf{a} \right)^T = \left(\mathbf{a}^T \tilde{\mathbf{R}} \mathbf{a} \right)^H = \mathbf{a}^T \tilde{\mathbf{R}}^* \mathbf{a} = \mathbf{a}^T \tilde{\mathbf{R}} \mathbf{a}. \quad (14)$$

Define

$$\mathbf{R} = \frac{\frac{1}{2} \left(\tilde{\mathbf{R}} + \tilde{\mathbf{R}}^H \right) + \frac{1}{2} \left(\tilde{\mathbf{R}} + \tilde{\mathbf{R}}^H \right)^T}{2}, \quad (15)$$

which is real, symmetric, and from (14) yields the same quadratic form as $\tilde{\mathbf{R}}$.

Finally, we remark that the quadratic form induced by \mathbf{R} is convex because $\mathbf{a}^T \mathbf{R} \mathbf{a}$ is nonnegative for any choice of \mathbf{a} . This follows from the derivation of (10), in which we note that the discriminant is nonnegative because the eigenvalues of the conjugate symmetric matrix $\sum_{\omega \in F} a_{\omega} \hat{\mathbf{Q}}_{\mathbf{X}}(\omega)$ must be real. \square

Theorem 1 permits an equivalent definition of the weights \hat{a} in the maximal eigengap estimator of (8) as the maximizer of the following expression, where \mathbf{R} is defined in (15).

$$\begin{aligned} \max_{\mathbf{a} \in \mathbb{R}^{|F|}} \mathbf{a}^T \mathbf{R} \mathbf{a} \quad (16) \\ \text{s.t. } \|\mathbf{a}\| \leq 1 \quad \mathbf{a} \geq 0. \end{aligned}$$

The primary benefit of this reformulation is the simple form of problem (16), which facilitates computation. We consider using both the 2-norm $\|\cdot\|_2$ and the 1-norm $\|\cdot\|_1$ to define the weight constraint. Note that, regardless of the norm used, (16) maximizes a convex function over a convex set, so a maximizer exists and is contained within the extreme points of the feasible set [9, Corollary 32.3.4].

III. APPLICATIONS & EXPERIMENTS

In this section we apply the maximal eigengap estimator to a set of signals collected in 2019 by an acoustic vector-sensor deployed in the Monterey Bay. We combine these signals with Automatic Identification System (AIS) data providing the GPS locations of vessels in the Bay throughout 2019. By comparing actual vessel DOAs with estimates provided by the maximal eigengap estimator, we assess the performance of the maximal eigengap estimator in a realistic setting.

A. Data Details

The acoustic signals we consider were collected by a Geospectrum M20-105 vector-sensor deployed at a depth of 891 meters on the Monterey Accelerated Research System cabled observatory, which is operated by the Monterey Bay Aquarium Research Institute. The time window of the data spans from February 1st to December 31st, 2019, and except for a handful of minor, maintenance-related disruptions these data are a continuous representation of the acoustic soundscape in Monterey Bay over the time frame considered.

The locations of vessels are given by GPS data provided by the US Coast Guard, with a time resolution of five minutes. We consider all observations where a vessel is due West, within 15 km, and has bearing between 190° and 350° of the sensor. We also require that there are no additional vessels identified within 20 km of the sensor. We pair each of these

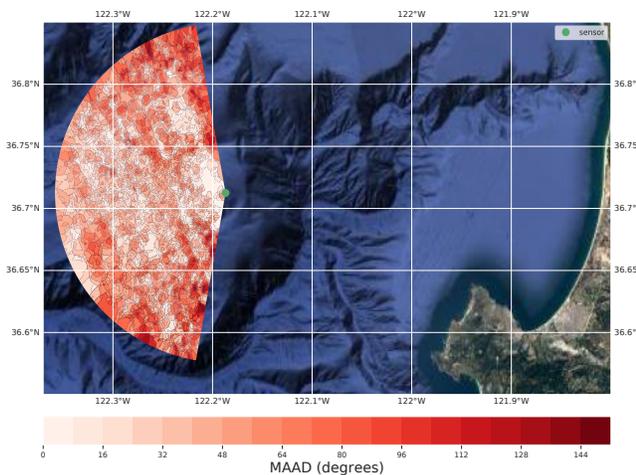


Fig. 1. The acoustic vector-sensor within the Monterey Bay overlaid with an interpolation of the mean absolute angular deviation for the maximal eigengap estimator (with 2-norm and no scaling, see Table I) constructed using K-nearest neighbors regression over all observed vessels.

Reference	Abbreviation	Norm	Scaling
This paper	1-Trace	1-norm	Trace
This paper	2-MinEig	2-norm	Minimum Eigenvalue
This paper	2-None	2-norm	None
[7]	Covar	NA	NA

TABLE I
THE ESTIMATORS CONSIDERED. THE MAXIMAL EIGENGAP (16) USES VARIOUS NORMS AND CSD MATRIX SCALINGS. THE COVARIANCE ESTIMATOR OF [7] IS INCLUDED AS A COMPETING METHOD.

vessel locations with the corresponding five minutes of signal recorded by the acoustic vector-sensor, resulting in 3674 observations labeled by signal DOA. By restricting our attention to vessels West of the sensor, we focus on detecting an eigenvector defining the signal subspace, and not considering vessels within 10° of North or South avoids complications arising from the ambiguity of its sign. We note that once a signal subspace has been detected, additional processing can be used to resolve this ambiguity using the acoustic vector-sensor’s omnidirectional channel [7].

Figure 1 gives the location of the vector-sensor within the Monterey Bay, with an interpolated error function for the estimator superimposed. Bathymetrically, the sensor is located on a shelf within the Monterey Canyon. To the sensor’s East, the ocean is shallower and contains mostly fishing and recreational vessels, whereas West of the sensor a pair of nearby shipping channels yields traffic that is primarily commercial. The data contain an unknown proportion of errors caused by noisy GPS reports and interfering signals such as small recreational vessels or aquatic mammals. The multi-season nature of the data also produces dynamic propagation conditions which impact the strength of both source and interfering signals at the sensor [10].

B. Numerical Performance

To assess the maximal eigengap estimator’s performance, we apply it to the acoustic signals with known DOAs described in III-A. We take frequency bins F ranging from 75 Hz to 300 Hz with 2 Hz resolution, and use averaged periodograms to estimate the CSD matrices $\hat{\mathbf{P}}_{\mathbf{X}}(\omega)$ [11]. In II, we introduced two opportunities for variation in the implementation of the maximal eigengap estimator (16), depending on the norm used to constrain the weight vector and the scaling of the CSD matrix included in the data preprocessing (a)-(c). We consider three of these variations in the following experiments: 1-norm with trace scaling, 2-norm with minimal eigenvalue scaling, and 2-norm with no scaling. As a comparison of how the maximal eigengap estimator performs relative to existing methods, we also implement and apply the covariance-based DOA estimator from [7]. Table I gives the different estimators considered.

We comment briefly on the implementation details of the maximal eigengap estimator using the 1 and 2-norms. For the 1-norm constraint, the weight vector in the maximal eigengap estimator can be computed in closed form. Indeed, when the 1-norm is used in (16), the extreme points of the constraint set are $\{\mathbf{0}, \mathbf{e}_1, \dots, \mathbf{e}_{|F|}\}$, where $\mathbf{0}$ denotes the zero vector and \mathbf{e}_i the i th standard basis vector. Then it is clear that the maximizer

in (16) occurs at the standard basis vector corresponding to the largest diagonal term in \mathbf{R} .

In the 2-norm constrained case, (16) resembles a maximal eigenvalue problem, but with an additional nonnegativity constraint. When the entries of \mathbf{R} are nonnegative, the Perron-Frobenius theorem guarantees that $\mathbf{v}_{\max}(\mathbf{R})$ satisfies the nonnegativity constraint, but from its construction \mathbf{R} may have negative entries. Instead, we approximate a solution to (16) by projecting $\mathbf{v}_{\max}(\mathbf{R})$ onto the nonnegative orthant. In practice, we find that the $\mathbf{v}_{\max}(\mathbf{R})$ is primarily composed of positive entries, and that occasional negative entries are close to zero, which suggests that this approximation is reasonable.

The results of these experiments demonstrate that the maximal eigengap estimator is a more accurate method for DOA estimation of a wideband signal using an acoustic vector-sensor than those existing in the literature. Figure 2 gives the mean absolute angular deviation (MAAD) of the various methods as a function of the upper bound on the distance between the sensor and vessels. Though variations of the maximal eigengap estimator perform similarly, the 2-None variation performs strictly worse than the 1-Trace and 2-MinEig variations when greater than 5 km from the sensor. Most importantly, all variations of the maximal eigengap estimator outperform its competitor, the covariance method introduced in [7]. This performance difference is especially large for vessels close to the sensor, where the mean absolute angular deviation is approximately 30 degrees less for the maximal eigengap estimator. We conjecture that this difference is due to the flexibility of the maximal eigengap estimator, which selects from among a set of frequency bins those that present a similar signal subspace. The optimal frequency bins change depending on the vessel's distance from the sensor, and the maximal eigengap estimator has the ability to adapt to this change using its optimal weighting of CSD matrices.

Further comparisons of the estimators' performance demonstrate the utility of the maximal eigengap estimator. Figure 3 presents histograms of the absolute angular error for the

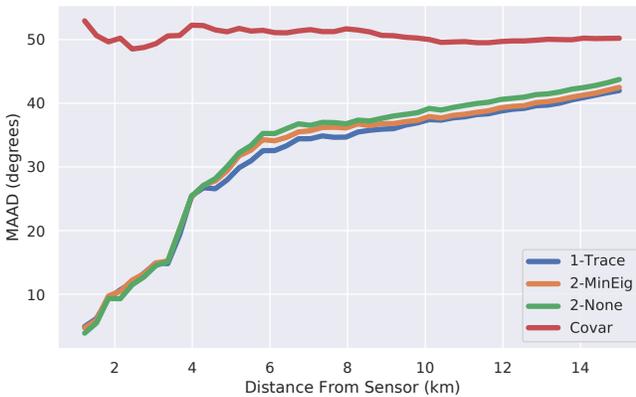


Fig. 2. Mean absolute angular deviation for each DOA estimator in Table I, as a function of the distance between a vessel and the sensor. The three variants of the maximal eigengap estimator perform much better than the covariance method of [7] when the vessels are closer to the sensor, and this advantage decreases as the distance increases.

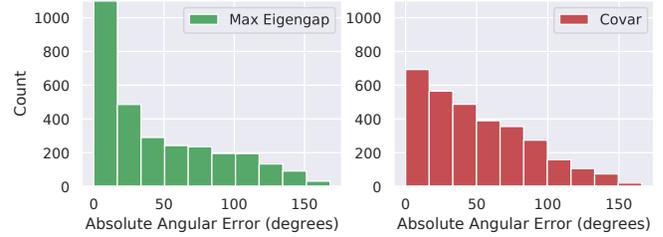


Fig. 3. Histograms of absolute angular error for the maximal eigengap estimator (with 2-norm and no scaling) and the covariance method of [7], applied to all vessels within 15 km of the sensor. The maximal eigengap estimator has error near zero more often than its competitor.

2-None variant of the maximal eigengap estimator and the competing covariance method. In this figure, the maximal eigengap has 50% more observations in the smallest error bin than the covariance method. Figure 1 presents the absolute angular deviation of the maximal eigengap estimator as a function of vessel location, constructed by interpolating error over all vessels in the data set. Certain vessel locations present more difficulty for DOA estimation than others. We suspect that this difficulty can primarily be attributed to propagation conditions arising from bathymetric features of those locations.

IV. CONCLUSION

In this paper we introduce the maximal eigengap estimator for DOA estimation of a wideband signal collected with a single acoustic vector-sensor. The maximal eigengap estimator's utility is in its formulation, which optimally and tractably combines signal subspace information across a frequency range. We demonstrate that the maximal eigengap estimator outperforms existing techniques for DOA estimation of maritime vessels on a set of labeled data collected by an acoustic vector-sensor.

REFERENCES

- [1] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 33, no. 4, pp. 823–831, 1985.
- [2] J. A. Cadzow, "Multiple source location-the signal subspace approach," *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. 38, no. 7, pp. 1110–1125, 1990.
- [3] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Proc.*, vol. 39, no. 5, pp. 1110–1121, 1991.
- [4] Y.-S. Yoon, L. M. Kaplan, and J. H. McClellan, "Tops: New doa estimator for wideband signals," *IEEE Trans. Signal Proc.*, vol. 54, no. 6, pp. 1977–1989, 2006.
- [5] M. M. Hyder and K. Mahata, "Direction-of-arrival estimation using a mixed $\ell_{2,0}$ norm approximation," *IEEE Trans. Signal Proc.*, vol. 58, no. 9, pp. 4646–4655, 2010.
- [6] S. Axler, *Linear algebra done right*. Springer Science & Business Media, 1997.
- [7] A. Nehorai and E. Paldi, "Acoustic vector-sensor array processing," *IEEE Trans. Signal Proc.*, vol. 42, no. 9, pp. 2481–2491, 1994.
- [8] M. J. Wainwright, *High-dimensional statistics: A non-asymptotic viewpoint*. Cambridge University Press, 2019, vol. 48.
- [9] R. T. Rockafellar, *Convex analysis*. Princeton university press, 2015.
- [10] K. L. Gemba, J. Sarkar, B. Cornuelle, W. S. Hodgkiss, and W. Kuperman, "Estimating relative channel impulse responses from ships of opportunity in a shallow water environment," *J. Acoustic. Soc. Am.*, vol. 144, no. 3, pp. 1231–1244, 2018.
- [11] S. M. Kay, *Modern spectral estimation: theory and application*. Prentice Hall, 1988.

Joint surface detection and depth estimation from single-photon Lidar data using ensemble estimators

K. Drummond

*School of Engineering and Physical Sciences
Heriot-Watt University and Leonardo UK
Edinburgh, United Kingdom
kd122@hw.ac.uk*

S. McLaughlin, Y. Altmann

*School of Engineering and Physical Sciences
Heriot-Watt University
Edinburgh, United Kingdom*

A. Pawlikowska, R. Lamb

*Leonardo UK
Edinburgh, United Kingdom*

Abstract—This paper addresses the problem of joint surface detection and depth estimation from single-photon Lidar (SPL) data. Traditional 3D ranging methods for SPL usually perform surface detection and range estimation sequentially to alleviate the computational burden of joint detection and estimation. Adopting a Bayesian formalism, the joint detection/estimation problem is formulated as a single inference problem. To avoid the intractable integrals usually involved with variable marginalization, we consider discrete variables and the resulting problem is recast as a model selection/averaging problem. We illustrate our method for a case where the expected signal-to-background (e.g., the target reflectivity and ambient illumination level) is unknown but the proposed framework can be adapted to more complex problems where the target depth can be obtained by combining several estimators. We demonstrate the additional benefits of the proposed method in also providing a conservative approach to uncertainty quantification of the calculated depth estimates, which can be used for real time analysis. The benefits of the proposed methods are illustrated using synthetic and real SPL data for targets at up to 8.6 km.

Index Terms—Single-photon Lidar, Bayesian estimation, Detection, Ensemble estimation.

I. INTRODUCTION

Single-Photon Lidar (SPL) is a reliable Lidar technology for readily providing 3D scene reconstruction. SPL systems are based on time-correlated single-photon counting (TCSPC) [1]–[3], whereby the laser source emits a short pulse towards the scene, part of which is reflected by the target. Once an individual photon of the reflected pulse is detected, which we refer to as a “desirable” detection event, the time interval between the pulse emission and the photon detection is recorded. This time interval is the photon’s time-of-flight (ToF) and it is logged in the corresponding time-bin of a histogram. ToF histograms can be adversely affected by “undesirable” detection events, which arise due to ambient illumination and dark counts. To provide more accurate and reliable depth estimates of the target, many pulses are repeatedly emitted towards the scene to build denser histograms and improve the signal quality.

SPL has proven to have great benefits for use in a range of different fields, including autonomous vehicles [4], agriculture [5] and defence [6]. The high sensitivity of single photon detectors allows for the use of low-power, eye-safe laser sources [6]. Furthermore, the picosecond timing resolution enables greater surface-to-surface resolution at ranges up to 200 km [7]. Thanks to advances in single-photon avalanche diode (SPAD) array technology, acquisition of data can now be achieved at video rates or higher [8], [9]. Consequently,

greater interest is being focused on faster data processing to reconstruct 3D scenes as fast and as reliably as possible. Whilst great strides have been achieved along those lines, e.g., [10], important challenges such as surface detection and reliable uncertainty quantification still need to be addressed.

In this paper, we propose a novel, pixel-wise, joint detection and depth estimation method, which detects objects/surfaces in the field of view, estimates their distance and rapidly provides uncertainty measures that can then be used in more sophisticated object recognition algorithms or subsequent decision-making processes. We use a similar observation model as described previously [11]–[14]. These methods treat the unknown model parameters (e.g., the target reflectivity) as continuous variables that are classically determined using sequential or iterative processes [15]–[18]. This can be computationally intensive, especially if the target depth and reflectivity are jointly estimated. We overcome this problem by treating the reflectivity as a discrete parameter, allowing us to perform joint detection and range estimation at a fixed (and low) computational cost. We also extend the observation model from previous works by allowing non-uniform background distributions without the method being significantly more computationally intensive. This enables the analysis of data corrupted by pile-up in SPAD detectors [19].

As multiple sources of error arise when reconstructing 3D surfaces, it is ever more important to quantify the uncertainty in depth estimation. Recent works have shown it is possible to use uncertainty quantification methods for joint depth estimation and detection, e.g., [20]. However, that method is far too slow for reconstruction at real time speeds, which motivates our work.

The remainder of this paper is organized as follows. Section II recalls the statistical observation model used for SPL and describes the proposed method for joint surface detection and depth estimation. Results of simulations conducted with synthetic single-pixel histogram data and real SPL data are presented and discussed in Section III. Conclusions are finally reported in Section IV.

II. BACKGROUND THEORY

A. Observation model

In this paper, we consider a set of K photon time of arrival (ToA) values $\mathbf{y} = \{y_k\}_{k=1}^K$, such that $y_k \in (0, T)$, where it is implicitly assumed that T is the repetition period of the

laser source [21]. Indices representing pixel dependency are omitted to simplify notation. The probability density function for a photon ToA, y_k , for a given pixel is given by

$$f(y_k|d, w) = w h_0 \left(y_k - \frac{2d}{c} \right) + (1 - w) \mathcal{V}_{(0,T)}, \quad (1)$$

where d is the range of the target surface within the admissible range gate and c is the speed of light, such that $2d/c$ is the characteristic ToF associated with the illuminated target. The function $h_0(\cdot)$ in Eq. (1) is the normalised Impulse Response Function (IRF) of the Lidar system which is not required to be Gaussian nor symmetric. It is generally measured during system calibration [12], [21]. The second term $\mathcal{V}_{(0,T)}$, assumed known, represents the distribution of undesirable background photon detection events mentioned in Section I. This need not necessarily be a uniform distribution, especially in situations of high ambient illumination conditions where pile-up effects due for instance to scattering media are more prominent. The shape of $\mathcal{V}_{(0,T)}$ arises generally due to the dominating properties of the ambient illumination over the undesirable detection events caused by dark counts, which are usually constant and relatively low compared with the ambient illumination undesirable detection events. The variable w in Eq. (1) is the probability of a detection event to be a desirable detection event. This probability is related to the signal to background ratio (SBR) by $SBR = w/(1 - w)$.

When K photons are detected, and the dead-times of the SPAD detector can be neglected, the photon ToA's are mutually independent (given d and w) and the joint likelihood can be expressed as [12], [21]

$$f(\mathbf{y}|d, w) = \prod_{k=1}^K f(y_k|d, w). \quad (2)$$

Our goal is to estimate d from Eq. (1), given that w is also unknown. Moreover, we are also interested in estimating w , as it provides information about the presence ($w > 0$) or absence ($w = 0$) of a surface, as well as its reflectivity.

B. Proposed estimation strategy

We assume that d can take a finite number N_D of values, i.e., $d \in \{d_1, \dots, d_{N_D}\}$. This set is constructed from a subset of the N_T number of original non-overlapping time bins spanning $(0, T)$ and N_D is a number of non-overlapping time bins within this subset, spanning $(\tau, T - \tau)$, such that $N_D \leq N_T$ and $\tau \geq 0$, where the width of the time bins is arbitrary (usually given by the timing resolution of the SPAD used). The parameter τ is chosen to ensure that the support of $h_0(\cdot)$ is always included in $(0, T)$ (for any admissible value of d) and thus the value of τ depends on the width of the IRF peak.

Let's first assume that w is known. The probability distribution $f(d|\mathbf{y}, w)$ can be obtained from the joint likelihood Eq. (2), using Bayes theorem

$$f(d|\mathbf{y}, w) = \frac{f(\mathbf{y}|d, w)f(d)}{f(\mathbf{y}|w)}, \quad (3)$$

where $f(d)$ is a user-defined depth prior distribution, $f(\mathbf{y}|w)$ is the marginal term and $f(\mathbf{y}|w) = \sum_{j=1}^{N_D} f(\mathbf{y}|d = d_j, w)f(d =$

$d_j)$ is a tractable normalizing constant (thanks to d being discrete). This posterior can then be used easily to compute Bayesian estimators (e.g., maximum a posteriori (MAP) or minimum mean squared error (MMSE), denoted $\mu(w)$) for the depth, as well as the posterior variance, denoted $\sigma^2(w)$. Unfortunately, depth inference using Eq. (3) is challenging since w is unknown in practice and setting its value poorly can greatly impact the depth estimation.

To alleviate this issue, a classical approach consists (assuming that w is continuous) of assigning w a prior distribution and of computing the following marginal posterior distribution

$$f(d|\mathbf{y}) = \int f(d|\mathbf{y}, w)f(w|\mathbf{y}) dw, \quad (4)$$

where $f(w|\mathbf{y})$ is the marginal posterior distribution of w . Unfortunately, manipulating $f(d|\mathbf{y})$ in Eq. (4) is challenging due to the integral w.r.t. w which needs to be approximated numerically for any value of d . To overcome this difficulty, we consider the parameter w as discrete with $w \in \{w_1, w_2, \dots, w_M\}$, where we allow $w_1 = 0$ to be in the admissible set of w and M is a user determined value. Using this discretization of w , Eq. (4) becomes

$$f(d|\mathbf{y}) = \sum_{m=1}^M f(d|\mathbf{y}, w_m)f(w_m|\mathbf{y}), \quad (5)$$

which becomes tractable, provided that $f(w_m|\mathbf{y})$ can be computed easily. The marginal posterior distribution $f(d|\mathbf{y})$ can be seen as a mixture of M distributions, whose weights are given by $f(w_m|\mathbf{y})$. Thus, its mean and variance can be easily derived from the mean and variance of each of its components. More precisely, by simplifying the notation $\mu_m = \mu(w_m)$ and $\sigma_m^2 = \sigma^2(w_m)$ for the mean and variance of $f(d|\mathbf{y}, w_m)$, the mean and variance of $f(d|\mathbf{y})$ in (5) are given by

$$\bar{\mu} = \sum_{m=1}^M f(w = w_m|\mathbf{y})\mu_m, \quad (6)$$

$$\bar{\sigma}^2 = \left(\sum_{m=1}^M f(w = w_m|\mathbf{y})(\sigma_m^2 + \mu_m^2) \right) - \bar{\mu}^2. \quad (7)$$

Since w can only take a finite number of values, the marginal posterior $f(w|\mathbf{y})$ can be computed exactly using

$$f(w|\mathbf{y}) = \frac{f(\mathbf{y}|w)f(w)}{\sum_{m=1}^M f(\mathbf{y}|w = w_m)f(w = w_m)}, \quad (8)$$

and $f(\mathbf{y}|w)$ defined below Eq. (3). While in Eq. (6), we consider the marginal mean of d a posteriori (having marginalised over w), it is of course also possible to condition the estimation of d based on the marginal MAP (MMAP) estimate of w . In that case, we can first compute

$$\hat{w} = \underset{w}{\operatorname{argmax}} f(w|\mathbf{y}), \quad (9)$$

and derive the mean and variance of $f(d|\mathbf{y}, \hat{w})$, which, as will be illustrated in Section III, often has a mean close to that of $f(d|\mathbf{y})$ but a smaller variance (since it does not account for the uncertainty associated with w). It is important to notice that the computation of the summary statistics in Eqs. (6) and (7)

requires the computation of MN_D likelihoods ($f(\mathbf{y}|d_n, w_m)$). The associated cost can thus be bounded by controlling the grids of d and w . Note also that Eq. (8), as a by-product of the depth estimation procedure, allows us to perform surface detection. For instance, we can decide that a surface is present based on $f(w > \alpha|\mathbf{y})$ where α is user-defined. More complex decision criteria can also be used as the proposed method is not based on single decision criterion. Finally, this method differs from the surface detection method detailed in [11] in which the depth estimation is performed after the detection step, whereas our proposed method performs both simultaneously.

C. Depth estimation using ensemble estimators

Eq. (5) is a mixture distribution and the resulting depth estimation method can be seen as a model averaging method, whose weights are given by $f(w_m|\mathbf{y})$. These M models are simply characterized by a different value of w_m . However, more general models can also be used. Consider a set of M models denoted $\{\mathcal{M}_m\}_m$, for which we can compute $f(d|\mathbf{y}, \mathcal{M}_m)$ and its mean and variance. Robust depth estimation via model averaging can then be achieved using

$$f(d|\mathbf{y}) = \sum_{m=1}^M \pi_m f(d|\mathbf{y}, \mathcal{M}_m), \quad (10)$$

where $\pi_m \geq 0, \forall m$ and $\sum_{m=1}^M \pi_m = 1$. These weights can be either arbitrarily defined (e.g., using $1/M$ when $f(\mathcal{M}_m|\mathbf{y})$ cannot be computed, or using $\pi_m = f(\mathcal{M}_m|\mathbf{y})$ otherwise). Due to space constraints, we only illustrate our method for models parametrized by w , but the proposed approach could also be used, for instance if several IRFs or background distributions were considered simultaneously. The former would arise for instance in long-range imaging applications when the peak broadens due to reflection on surfaces which are oblique to the beam direction. The latter can arise from varying levels of back-scattered light in scattering media. Models can also differ by the priors assigned to d (and the other model parameters such as w).

III. RESULTS

We first evaluate the performance of the proposed algorithm using synthetic single-pixel data and then two real SPL datasets, provided by Leonardo UK [6]. In all results presented, the background distribution $\mathcal{V}_{(0,T)}$ is assumed to be known, regardless of whether it is a uniform distribution or not (and estimated from the data as a pre-processing step).

A. Single-Pixel analysis

First, we generate two synthetic histograms of length $T = 1500$ bins, with $K = 100$ and $K = 1000$, respectively. The real IRF data obtained from [22] is used, with histogram resolution at 2ps per bin and $FWHM = 30$ bins = 60ps, and the maximum of the peak is set to bin 746. We set the ground truth signal photon probability to $w = 0.2$. The final pixel data generated are shown in Fig. 1 (bottom). For this initial investigation, we set $M = 20$ and the values of admissible values of w are uniformly spread in $[0, 1]$. Fig. 1 (top) shows that $f(w|\mathbf{y})$ is more concentrated as K increases since more detected photons help the discrimination between signal and

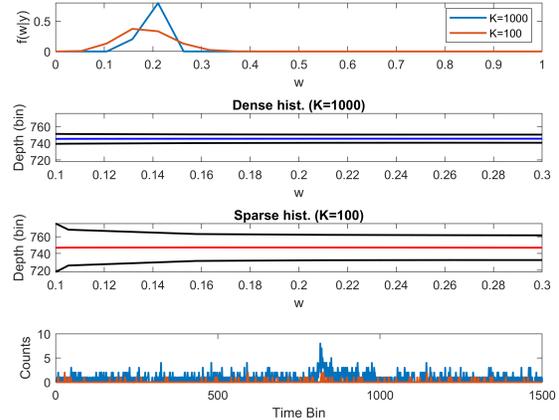


Fig. 1: Graphic results of photon dense (blue) and sparse (red) single pixel histogram data. Top: Plot of probability a posteriori $f(w|\mathbf{y})$ vs w . Middle: Plot of mean μ_m values for photon dense (Second) and sparse (Third) data, with error range plots (shown in black). Bottom: Single Pixel Histogram data plot.

	Dense hist. ($K = 1000$)			Sparse hist. ($K = 100$)		
	Mean of d	Var. of d	Est. w	Mean of d	Var. of d	Est. w
$f(d \mathbf{y}, \hat{w})$ using (9)	745.44	2.68	0.21	747.13	28.55	0.16
$f(d \mathbf{y})$ and $f(w \mathbf{y})$	745.43	2.74	0.20	747.02	62.50	0.19

TABLE I: Comparison of the different estimates of d (conditioned on w or not) and w (marginal MAP or marginal MMSE), for $K = 1000$ and $K = 100$. The actual value of (d, w) is $(746, 0.2)$. Note that $f(d|\mathbf{y})$ and $f(w|\mathbf{y})$ are computed using (5) and (8), respectively.

background photons. Similarly, the second and third row of Fig. 1, which depict $\mu(w = w_m)$ and $\sigma^2(w = w_m)$, illustrate how the estimated depth mean and variance using $f(d|\mathbf{y}, w)$ depend on K and on the unknown value of w (plots restricted to $w > 0.1$ below which the means degrade drastically).

Table I summarizes the different estimates of d and w for the two histograms with $K = 100$ and $K = 1000$. First, we can note that the marginal MMSE estimator of w (i.e., the mean of $f(w|\mathbf{y})$ used in the bottom row of Table I) is usually more reliable than the MMAP estimator in (9) which is more sensitive to the resolution of the w -grid (and M), especially for small values of K . Second, the MMSE depth estimates, conditioned on the MMAP estimator of w (top row) or computed from the marginal posterior $f(d|\mathbf{y})$ (bottom row) are similar. However, the estimated variance is larger in the latter case, since it incorporates the uncertainty about this unknown parameter w . This estimator is thus more conservative in terms of uncertainty quantification.

B. Real SPL data analysis

Here, we use two real SPL datasets acquired by Leonardo to illustrate the potential benefits of the proposed method. The

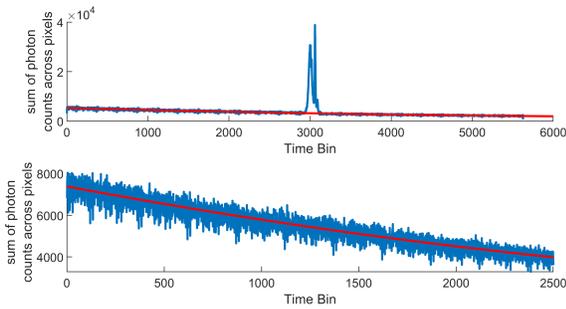


Fig. 2: Sum of photon counts (blue) across all Lidar pixels for college tower data (top) and pylon data (bottom). The red curves represent the estimated background contributions.

first SPL cube consists of 100×50 pixels and focuses on the tower of an Edinburgh college tower, taken at $\approx 3\text{km}$ range, already considered in [6]. The second dataset consists of 80×40 pixels representing a pylon, taken at $\approx 8.6\text{km}$ range, and is an example of a challenging, line-based object. These datasets primarily differ by their average SBR, much higher in the first dataset ($SBR \approx 0.22$) than in the second ($SBR \approx 0.01$). It should be noted the SBR described is not the peak SBR, where the signal to background ratio is calculated from the time bin where the signal/peak reaches its maximum value, but rather the ratio of desirable to undesirable detection events from all histogram temporal bins. The difference in the SBR of both datasets can be seen in Fig. 2, which depicts two histograms obtained by accumulating the histograms of all the pixels in the set. These accumulated histograms can be used to estimate an average background distribution which is not constant in these two cases. Here, we used second-order polynomials to fit $\mathcal{V}_{(0,T)}$ and compared the results obtained assuming constant background instead. The background for the SBR is equivalent to the integral of the $\mathcal{V}_{(0,T)}$ distribution curve over all the temporal bins.

For both datasets, our detection method is compared to the detection method proposed by Tachella et al. [11], assuming a constant background distribution. It is worth recalling that estimates of w can be used to estimate to the target intensity (number of signal photons $I = wK$) and the number of background photons $B = (1 - w)K$ and we used \bar{w} , the mean of $f(w|\mathbf{y})$ in these expressions, leading to \bar{I} and \bar{B} . The proposed method has been applied with larger M values than for the single pixel analysis for a more precise estimation of w . A target is assumed to be present in each pixel if and only if $f(w > w_0) > 0.5$, where w_0 is user-defined and scene dependent. Due to space constraints, the final depth variance results are not presented but their scale is of the order the timing resolution of the SPAD used.

1) *College Tower Data*: For this dataset, target detection is a relative simple task and the admissible grid of w is set using $M = 100$ equality spaced $\{w_m\}_m$ and we used $w_0 = 0.02$. The estimated probability of target presence maps are displayed in Fig. 3, which shows that for this scene, the proposed method leads to results similar to that using Tachella's method [11], irrespective of the background model

adopted. This can be explained by the fact that the peaks in the histograms can be easily identified even when assuming a constant background.

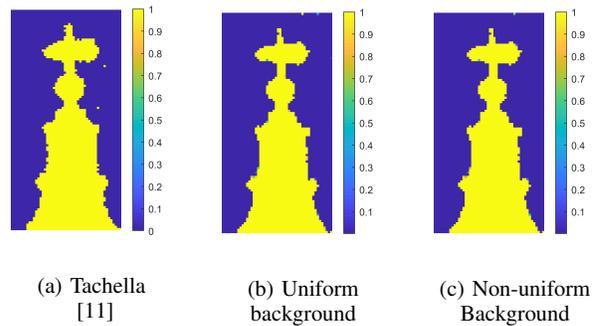


Fig. 3: College tower data comparison of probability of detection results for the method by Tachella et al. (a), and for the proposed method using uniform (b) and non uniform (c) background distributions, where $w_0 = 0.02$.

The final mean depth estimates $\bar{\mu}$, reflectivity estimates \bar{I} and background estimates \bar{B} are presented in Fig. 4.

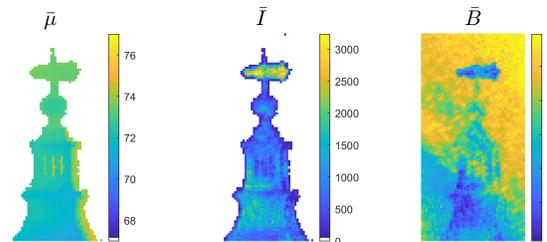


Fig. 4: Final mean depth (left), reflectivity (middle) and background (right) estimates for the tower data using the proposed method, for $w_0 = 0.02$.

2) *Pylon Data*: For this dataset for which we know the SBR is low, the admissible grid of w is set using $M = 200$ logarithmically spaced $\{w_m\}_m$, with $w_2 = 10^{-5}$, $w_M = 10^{-1}$ and $w_0 = 0.008$. The estimated target presence maps are displayed in Fig. 5, which shows that for this scene, the proposed method leads to results noisier than those obtained by Tachella et al. [11] (taking the conjugate gamma density shape parameter $\alpha_b = 100$) when we assume a uniform background distribution, and leads to improved results for detecting the pylon structure when we assume a non-uniform background distribution. In this instance, the peaks in the histograms cannot be easily identified when assuming a constant background, and so a non-uniform distribution assumption is required to obtain better results from the cross-correlation calculations.

The final mean depth, reflectivity and background estimates obtained using our proposed method are shown in Fig. 6, under the condition $w_0 = 0.008$ and $\bar{I} > 0.25$.

IV. CONCLUSION

In this paper, we proposed a novel method for joint surface detection and depth estimation from SPL data using discrete

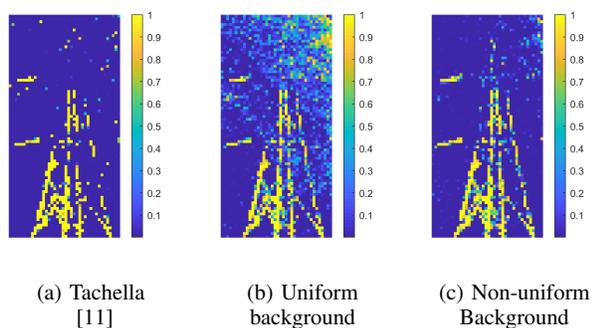


Fig. 5: Pylon data comparison of probability of detection results for Tachella et al. method (a), where $\alpha_b = 100$, and proposed method using uniform (b) and non uniform (c) background distributions, where $w_0 = 0.008$.

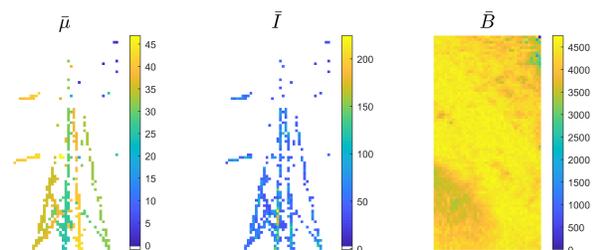


Fig. 6: Final mean depth (left), reflectivity (middle) and background (right) estimates for the pylon data using the proposed method, for $w_0 = 0.008$ and $\bar{I} > 0.25$.

variables to avoid intractable marginalizations and producing satisfactory results in the final estimates using model selection/averaging. While we illustrated the method for a case where only the expected signal-to-background is unknown, in the future we aim to adapt the framework to more complex problems where the target depth can be obtained combining several arbitrary estimators. Furthermore we plan to propose a GPU implementation to enable reliable depth estimation and uncertainty quantification at real-time speeds.

ACKNOWLEDGEMENTS

This work was supported by CENSIS and Leonardo UK. This work was supported by the Royal Academy of Engineering under the Research Fellowship scheme RF201617/16/31, the Engineering and Physical Sciences Research Council of the UK (EPSRC) Grant number EP/S000631/1 and the UK MOD University Defence Research Collaboration (UDRC) in Signal Processing.

REFERENCES

- [1] G. Buller and R. Collins, "Single-photon generation and detection," *Measurement Science and Technology*, vol. 21, no. 1, 2010.
- [2] A. McCarthy, X. Ren, A. D. Frera, N. R. Gemmill, N. J. Krichel, C. Scarcella, A. Ruggeri, A. Tosi, and G. S. Buller, "Kilometer-range depth imaging at 1550 nm wavelength using an ingaas/inp single-photon avalanche diode detector," *Opt. Express*, vol. 21, no. 19, pp. 22098–22113, Sep 2013. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-21-19-22098>
- [3] W. Becker, *Advanced Time-Correlated Single-Photon Counting Techniques*, ser. Springer Series in Chemical Physics. Springer, 2005.
- [4] J. Rapp, J. Tachella, Y. Altmann, S. McLaughlin, and V. K. Goyal, "Advances in single-photon Lidar for autonomous vehicles," *IEEE Signal Processing Magazine*, vol. 37, no. 4, 2020.
- [5] P. Yimyam and A. F. Clark, "3D reconstruction and feature extraction for agricultural produce grading," *8th International Conference on Knowledge and Smart Technology (KST)*, Chiangmai, no. 136-141, 2016.
- [6] A. M. Pawlikowska, A. Halimi, R. A. Lamb, and G. S. Buller, "Single-photon three-dimensional imaging at up to 10 kilometers range," *Opt. Express*, vol. 25, no. 11919–11931, 2017.
- [7] Z.-P. Li, J.-T. Ye, X. Huang, P.-Y. Jiang, Y. Cao, Y. Hong, C. Yu, J. Zhang, Q. Zhang, C.-Z. Peng, F. Xu, and J.-W. Pan, "Single-photon imaging over 200km," *Optica*, vol. 8, no. 3, pp. 344–349, Mar 2021. [Online]. Available: <http://www.osapublishing.org/optica/abstract.cfm?URI=optica-8-3-344>
- [8] M. Entwistle, M. A. Itzler, J. Chen, M. Owens, K. Patel, X. Jiang, K. Slomkowski, and S. Rangwala, "Geiger-mode APD camera system for single-photon 3D LADAR imaging," in *Advanced Photon Counting Techniques VI*, M. A. Itzler, Ed., vol. 8375, June 2012, p. 83750D.
- [9] R. Henderson, N. Johnston, H. Chen, D. Li, G. Hungerford, R. Hirsch, P. Yip, D. McLoskey, and D. Birch, "A 192 x 128 time correlated single photon counting imager in 40nm CMOS technology," in *44th European Solid-State Circuits Conference (ESSCIRC) 2018*. IEEE, Sept. 2018.
- [10] J. Tachella, Y. Altmann, N. Mellado, A. McCarthy, R. Tobin, G. S. Buller, J.-Y. Tourneret, and S. McLaughlin, "Real-time 3D reconstruction from single-photon Lidar data using plug-and-play point cloud denoisers," *Nature Communications*, vol. 10, no. 1, Nov 2019. [Online]. Available: <http://dx.doi.org/10.1038/s41467-019-12943-7>
- [11] J. Tachella, Y. Altmann, S. McLaughlin, and J. Y. Tourneret, "Fast surface detection in single-photon Lidar waveforms," in *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5.
- [12] Q. Legros, J. Tachella, R. Tobin, A. McCarthy, S. Meignen, G. S. Buller, Y. Altmann, S. McLaughlin, and M. E. Davies, "Robust 3D reconstruction of dynamic scenes from single-photon Lidar using beta-divergences," *IEEE Transactions on Image Processing*, vol. 30, p. 1716–1727, 2021. [Online]. Available: <http://dx.doi.org/10.1109/TIP.2020.3046882>
- [13] Y. Altmann, S. McLaughlin, and M. E. Davies, "Fast online 3D reconstruction of dynamic scenes from individual single-photon detection events," *IEEE Transactions on Image Processing*, vol. 29, pp. 2666–2675, 2020.
- [14] J. Rapp and V. K. Goyal, "A few photons among many: Unmixing signal and noise for photon-efficient active imaging," *IEEE Transactions on Computational Imaging*, vol. 3, no. 3, pp. 445–459, 2017.
- [15] D. Shin, A. Kirmani, V. K. Goyal, and J. H. Shapiro, "Photon-efficient computational 3D and reflectivity imaging with single-photon detectors," *IEEE Trans. Comput. Imaging*, vol. 1, no. 2, pp. 112–125, Jun 2015.
- [16] Y. Altmann and S. McLaughlin, "Range estimation from single-photon Lidar data using a stochastic EM approach," in *2018 26th European Signal Processing Conference (EUSIPCO)*, ser. European Signal Processing Conference (EUSIPCO). United States: IEEE, Dec. 2018, pp. 1112–1116.
- [17] Y. Altmann, X. Ren, A. McCarthy, G. Buller, and S. McLaughlin, "Lidar waveform-based analysis of depth images constructed using sparse single-photon data," *IEEE Transactions on Image Processing*, vol. 25, no. 5, pp. 1935–1946, May 2016.
- [18] X. Ren, P. W. R. Connolly, A. Halimi, Y. Altmann, S. McLaughlin, I. Gyongy, R. K. Henderson, and G. S. Buller, "High-resolution depth profiling using a range-gated CMOS SPAD quanta image sensor," *Opt. Express*, vol. 26, no. 5, pp. 5541–5557, Mar 2018. [Online]. Available: <http://www.opticsexpress.org/abstract.cfm?URI=oe-26-5-5541>
- [19] A. Gupta, A. Ingle, A. Velten, and M. Gupta, "Photon-flooded single-photon 3D cameras," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [20] Y. Altmann, A. Maccarone, A. Halimi, A. McCarthy, G. Buller, and S. McLaughlin, "Efficient range estimation and material quantification from multispectral Lidar waveforms," in *2016 Sensor Signal Processing for Defence (SSPD)*, 2016, pp. 1–5.
- [21] Q. Legros, S. McLaughlin, Y. Altmann, S. Meignen, and M. E. Davies, "Robust depth imaging in adverse scenarios using single-photon Lidar and beta-divergences," in *2020 Sensor Signal Processing for Defence Conference (SSPD)*, 2020, pp. 1–5.
- [22] Y. Altmann, X. Ren, A. McCarthy, G. Buller, and S. McLaughlin, "Robust bayesian target detection algorithm for depth imaging from sparse single-photon data," *IEEE Transactions on Computational Imaging*, vol. 2, no. 4, pp. 456–467, Dec. 2016, arXiv admin note: text overlap with arXiv:1507.02511.

Detecting LFM Parameters in Joint Communications and Radar Frequency Bands

Kaiyu Zhang, Fraser K. Coutts, and John Thompson

Institute for Digital Communications, University of Edinburgh, Edinburgh, EH9 3FG, UK

Email: Kaiyu.Zhang@ed.ac.uk

Abstract—As a significant radar waveform, linear frequency modulation (LFM) is widely used in both military and civilian applications. Recently, there has been increasing interest in using the same radio spectrum to enable the radar and communication signals to coexist. This paper explores methods capable of estimating LFM parameters, particularly in the presence of co-channel orthogonal frequency division multiplexing (OFDM) signals. In particular, this paper applies the discrete chirp Fourier transform (DCFT) to this specific scenario and also compares this approach to the idea of reusing an OFDM receiver to estimate the LFM parameters. Through simulations, we demonstrate the use of the Hough transform to confirm that these can be identified to a high degree of accuracy. In addition, we discuss how this “reuse” approach opens up new applications for storing waveforms of interest for later data analysis.

I. INTRODUCTION

Nowadays, as the number of mobile terminal users increases and diversified sensors are applied in various scenarios, a great amount of information is usually disposed of in traditional signal processing to reduce complexity and improve efficiency in radio frequency (RF) receivers. However, such information could contain latent features of interest and be worth storing or recovering for further processing; for example, for detecting abnormal waveforms or other signal processing tasks. In the meantime, in these various scenarios, spectrum resources grow scarce and different waveforms or different sensors are more likely to interfere with each other. Thus, robust signal detection in the presence of co-channel interference becomes an important and imminent problem to be solved.

The two most common radio frequency receivers in defence are communication and radar. Many researchers are now focussed on communication and radar systems working simultaneously to improve spectral efficiency. For example, [1] discusses two scenarios under the topic of communication and radar spectrum sharing (CRSS): radar-communication coexistence (RCC) and dual-functional radar-communication (DFRC). With respect to DFRC, [2] and [3] present orthogonal frequency division multiplexing (OFDM) as a promising candidate that is used in multiple scenarios, notably for autonomous vehicles. Also, [4] introduces the use of the cyclic prefix (CP) in OFDM system to combat channel effects by circular convolution; this inherent ability to overcome channel effects makes OFDM a common modulation scheme widely used in communication systems.

Reference [5] introduces the linear frequency modulation (LFM) waveform as one kind of continuous waveform (CW)

that is frequently used in defence and civilian applications. In addition, by utilising a combined communication and radar waveform, [6] proposes chirp modulation to achieve radar and communication functions simultaneously and [7] discusses modulating a LFM signal by adjusting the data with orthogonal sequences in an integrated radar and communication scenario.

Thus, CRSS is the application scenario selected for study in this paper, where OFDM and LFM waveforms represent the communication waveform and the radar waveform, respectively. In the CRSS scenario with LFM and OFDM waveforms, one waveform could be seen as the interference to the other. Much of the research to date focuses on the application and the performance of CRSS. However, in this paper the question to be studied is how to detect LFM waveforms in the presence of co-channel OFDM signals. The receiver may then choose to store such waveforms for later processing. With respect to CRSS scenarios, other techniques, such as compressive sensing (CS), have already been applied in [8] to improve the communication symbol-error-rate performance when uncoordinated radar waveforms are also present. At a receiver, the received waveforms may contain information which is useful for signal processing and may be worth storing. Based on this point, CS [9] is also a promising technique for analysing or storing large amounts of data and can be applied on detection, classification, estimation, and filtering problems. Recently, it has been shown that CS methods are also able to compress data and store results for further processing [10]. Specifically, [10] utilised an information-theoretic projection design approach to control the input-output mutual information terms for two independent sources, which is extremely suitable for CRSS scenarios. Thus CS is a significant technique which will be investigated and applied to CRSS in future research.

To detect the LFM parameters, there are several methods proposed in the literature. Reference [11] proposes joint estimation of the phase, frequency, and frequency rate based on the application of least squares to the unwrapped phase of the signal. In [12], an estimation algorithm based on a simple iterative approach is proposed whose main characteristics include accuracy, reduction in error propagation effects, and operation over a wide range of phase parameter values. In [13] and [14], methods are proposed to estimate chirp parameters or achieve chirp detection based on fast Fourier transform (FFT) techniques. In addition, a novel algorithm for the parameter estimation of multicomponent chirp signals in complicated noise environments is proposed in [15]. Research in [16]

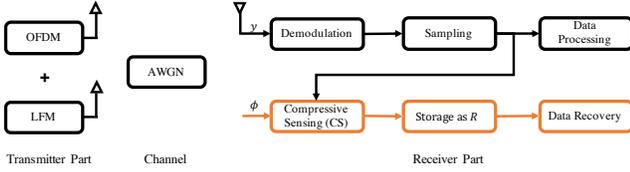


Fig. 1. Joint OFDM and LFM System.

studies the application of the discrete chirp Fourier transform (DCFT) for LFM waveform parameter estimation and [17] discusses the maximum chirplet transform. The latter is a simplification of the chirplet transform, where an iterative detection followed by window subtraction is employed to avoid re-computation of the spectrum.

This paper will focus on one specific application for LFM estimation, where a combined co-channel OFDM and LFM signal model is considered. Specifically, most OFDM receivers contain an efficient FFT implementation. The main contribution of our paper is to show that we are able to estimate the LFM parameters by reusing an existing OFDM receiver directly in order to reduce hardware complexity. We also compare this approach with the more standard DCFT technique from [16]. This work is the starting point for an exploration into the use of general CS techniques for the analysis of two independent signal source, combined communications and radar signals.

The layout of this paper is as follows: Section II introduces the system diagram and the system model of OFDM and LFM transmitting simultaneously. Section III explains how the DCFT can be used to estimate the LFM parameters in the system model. Section IV proposes an alternative approach to detect LFM parameters by reusing an OFDM receiver in combination with the Hough transform [18]. Section V discusses the two methods in Section III and Section IV and also introduces future research topics.

II. SYSTEM MODEL

In this section, we provide an overview of the system model utilised throughout the paper. The model is constructed from three main components — a transmitter subsystem, a channel, and a receiver subsystem — and is illustrated in Fig. 1. Here, the black blocks are standard signal processing elements, and the orange blocks are the objective of future research.

Assuming there are M_s subcarriers in the OFDM waveform, N_c symbols for each symbol block, and the length of CP is N_{cp} symbols, the complex-valued time domain OFDM signal $x(t)$ considered in this paper is defined as in [4]

$$x(t) = \sum_{k=0}^{M_s-1} X_k e^{j2\pi kt/T}, \quad 0 < t < T + T_g, \quad (1)$$

where X_k is the data symbol modulated via quadrature phase shift keying (QPSK) modulation, and the frequency interval for each adjacent subcarrier is $\Delta W = 1/T$, while T is symbol time. Equation (1) is usually implemented using the inverse discrete Fourier transform (IDFT).

Furthermore, for the radar waveform, assuming f_i is the linear frequency and f_c is the carrier frequency, we define the instantaneous frequency of an LFM waveform at time t as $f_{\text{LFM}}(t)$, which is expressed as:

$$f_{\text{LFM}}(t) = f_i t + f_c. \quad (2)$$

Thus, the corresponding complex analogue form of LFM is

$$f(t) = e^{j(\beta_0 t^2 + \alpha_0 t)}. \quad (3)$$

The derivative of the phase term in (3) is the instantaneous frequency of LFM as in (2), therefore

$$\beta_0 = \pi f_i, \quad \alpha_0 = 2\pi f_c. \quad (4)$$

Thus, (2) can be rewritten as

$$f_{\text{LFM}}(t) = \frac{\beta_0}{\pi} t + \frac{\alpha_0}{2\pi}. \quad (5)$$

At the receiver part, the received signal is the $(N_c + N_{cp}) \times 1$ vector \mathbf{y} as follows:

$$\mathbf{y} = \mathbf{y}_{\text{ofdm}} + \mathbf{y}_{\text{lfm}} + \mathbf{w}, \quad (6)$$

where \mathbf{y}_{ofdm} is the $(N_c + N_{cp}) \times 1$ OFDM data, \mathbf{y}_{lfm} is the $(N_c + N_{cp}) \times 1$ LFM data, and \mathbf{w} is the $(N_c + N_{cp}) \times 1$ additive white Gaussian noise (AWGN).

The data processing block after sampling and demodulation in Fig. 1 can be a LFM receiver or an OFDM receiver. For the orange blocks, CS is considered as a data analysing process to store data, which is different to the normal receivers. These processing blocks will be discussed further in Section V of the paper.

III. DISCRETE CHIRP FOURIER TRANSFORM

Research in [16] studies using the N -point DCFT to estimate chirp parameters and this paper also presents theoretical analyses on the selection of N . In this section, we will briefly introduce the DCFT method, apply it to OFDM and LFM combined signals, and present the simulation results.

The DCFT is similar to the discrete Fourier transform (DFT), which is efficiently applied in OFDM signal reception via the FFT. The N -point DFT is defined as

$$X[k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n] W_N^{nk}, \quad k = 0, 1, \dots, N-1, \quad (7)$$

where $W_N = \exp(-2\pi j/N)$. Similarly, the expression for the N -point DCFT method is presented as

$$X[l, k] = \frac{1}{\sqrt{N}} \sum_{n=0}^{N-1} x[n] W_N^{ln^2 + kn}, \quad l, k = 0, 1, \dots, N-1. \quad (8)$$

When applying the DCFT to equation (3) to estimate LFM parameters, [16] proposed the following equation when the sampling rate is $t = n/N^{1/3}$;

$$f[n, l, k] = e^{j\left(\beta_0 \frac{n^2}{N^{\frac{2}{3}}} + \alpha_0 \frac{n}{N^{\frac{1}{3}}}\right)} = W_N^{-(ln^2 + kn)}. \quad (9)$$

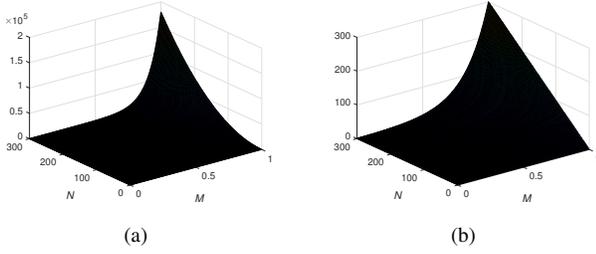


Fig. 2. Frequency estimation ranges: (a) \tilde{f}_l and (b) \tilde{f}_c based on N and M .

Based on (3) and (9), the relationship between l , k and β_0 , α_0 are as follows:

$$l = \frac{\beta_0}{2\pi} N^{\frac{1}{3}} \quad k = \frac{\alpha_0}{2\pi} N^{\frac{2}{3}}. \quad (10)$$

Thus, we are able to determine the estimated l as \tilde{l} and the estimated k as \tilde{k} by (8), then the estimated β_0 as $\tilde{\beta}_0$ and the estimated α_0 as $\tilde{\alpha}_0$ are also able to be deduced from equation (10). Furthermore, [16] also discusses limitations on the selection of N , which should be an odd value to avoid multiple peaks appearing in the DCFT spectrum. Even values of N are typically preferred for efficient implementations of the DFT. Thus, being constrained to odd values of N during computation of the DCFT might inhibit low-complexity solutions. The estimation range of an N -point DCFT under this sampling rate, when $\tilde{l} \leq N$ and $\tilde{k} \leq N$ can be characterised by

$$\tilde{f}_l \leq 2N^{\frac{2}{3}} \quad \tilde{f}_c \leq N^{\frac{1}{3}}. \quad (11)$$

When modifying the sampling rate to $t = n \times 1/N^M$, where M is a value in the interval of $(0, 1]$, equation (3) can be rewritten as

$$f[n, l, k] = e^{j(\beta_0 \frac{n^2}{N^{2M}} - \alpha_0 \frac{n}{N^M})} = W_N^{-(ln^2 + kn)}. \quad (12)$$

Thus, the relationship between \tilde{l} , \tilde{k} and $\tilde{\beta}_0$, $\tilde{\alpha}_0$ under M sampling rate are as follows:

$$\tilde{l} = \frac{\tilde{\beta}_0}{2\pi N^{2M-1}}, \quad \tilde{k} = \frac{\tilde{\alpha}_0}{2\pi N^{M-1}}. \quad (13)$$

In the meanwhile, the corresponding estimation range of \tilde{f}_l and \tilde{f}_c are as follows:

$$\tilde{f}_l \leq 2N^{2M} \quad \tilde{f}_c \leq N^M. \quad (14)$$

Based on (14), Fig. 2(a) and Fig. 2(b) present the estimation range of \tilde{f}_l and \tilde{f}_c depending on different values of N and M , respectively. From Fig. 2, we observe that the estimation ranges of \tilde{f}_l and \tilde{f}_c increase with N and M simultaneously. In both plots, the maximum estimation range occurs when both N and M are maximised. In addition, computational complexity is increasing more expensively with the increasing of the value of N .

TABLE I
OFDM PARAMETERS

Name	Value
Length of CP	16
FFT length	64
Symbol Period	0.08 s
Bandwidth	1k Hz
Modulation	QPSK
Total Subcarriers	52 (Freq Index -26 to $+26$)
DC Subcarrier	Null (0 subcarrier)
Signal to Noise Ratio (SNR)	0 dB

TABLE II
LFM PARAMETERS

Name	Value
Start Frequency	0 Hz
End Frequency	500 Hz
Repetition Period	10 s

Considering the application scenario and referring to the OFDM standard parameters of [19], we set the OFDM parameters and the LFM parameters for this paper as shown in Table I and in Table II, respectively. From the parameters in Table II, the frequency settings for LFM in the joint radar and communication system are as follows: $f_l = 50$ Hz and $f_c = 0$ Hz. Furthermore, in the received signal (6), the SNR for the OFDM signal with respect to the AWGN is 0 dB, while the transmission power of the OFDM signal is equal to the transmission power of the LFM signal.

After obtaining the receiver signal (6), we transform the data using the DCFT with a sampling rate of $t = n \times 1/N^M$, where $N = 25$ and $M = 1$. Then we acquire $f[n, l, k]$ for different l and k as shown in Fig. 3. This figure indicates that the peak of the magnitude of the DCFT is 27.9; the corresponding coordinates $(0, 1)$ are the most suitable for the frequency estimation. By placing $\tilde{l} = 1$ and $\tilde{k} = 0$ into (4) and (13), we estimate the LFM parameters as: $\tilde{f}_l = 50$ Hz and $\tilde{f}_c = 0$ Hz, respectively.

IV. REUSING THE OFDM RECEIVER FOR LFM PARAMETER ESTIMATION

In [13] and [14], FFT-based methods to estimate the chirp parameters and to achieve chirp detection are proposed, respectively, and [20] discusses the communication performance of an OFDM receiver while receiving OFDM and LFM combined signals.

One main feature of the OFDM receiver is that it can easily implement a frequency transform through the use of the FFT [4]. In this section, we propose an alternative estimation approach by reusing the OFDM receiver to estimate the LFM parameters in the OFDM-LFM combined signal, which is defined in (6). Compared with the DCFT method applied in the LFM receiver of Fig. 1, this method directly applies the

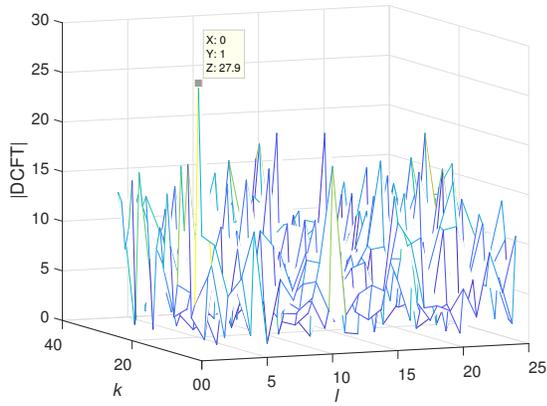


Fig. 3. Magnitude of DCFT spectrum



Fig. 4. Data processing diagram for output of OFDM receiver.

OFDM receiver as the first data processing step in Fig. 1 to estimate the LFM parameters.

Fig. 4 introduces the main data processing steps applied to the combined signal in (6) as follows:

1) *Discard CP*: We dispose of the CP in each symbol period at the receiver part. In our simulations, when removing the CP, 20% of the symbol samples are removed for each symbol period, thus we use zero-padding to extend the remaining samples over 0.064s into one full symbol period 0.08 s.

2) *FFT*: Apply FFT to the data after removing the CP, which is inherently utilised in the OFDM receiver.

3) *Hough Transform*: In this step, our work builds on ideas published in [21], which studied creating images of communications signals for analysis. Furthermore, [22] applies the Hough transform to multi-component LFM signals to estimate the LFM parameters. Therefore, we apply the Hough transform [18], which is defined as follows:

$$\rho = x \cos \theta + y \sin \theta, \quad (15)$$

where x, y is the coordinates under Cartesian coordinate system, ρ represents the distance between the origin and the given line, and θ denotes the angle between the X -axis and the given line.

4) *Parameter Estimation*: Fit the ρ and θ from the Hough transform into equation (2) to estimate the LFM parameters.

Step 1) and Step 2) are two regular signal processing functions in an OFDM receiver. Step 3) and Step 4) are two additional procedures to detect the LFM parameters and can be easily applied after the traditional OFDM receiver.

Based on the OFDM and LFM parameters in Table I and Table II and the data process shown in Fig. 4, we obtain the spectrum of the processed OFDM-LFM signal as shown in Fig. 5. Here, the peak of the spectrum of the combined signal is caused by the LFM waveform. Therefore, in order to estimate

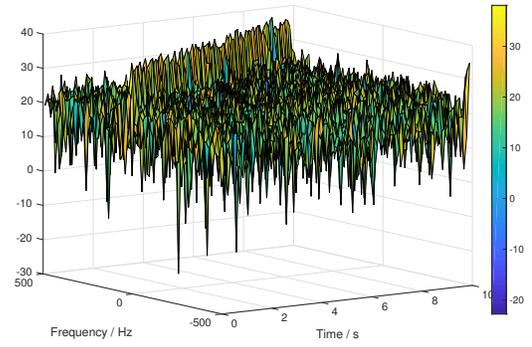


Fig. 5. Combined OFDM and LFM spectrum.

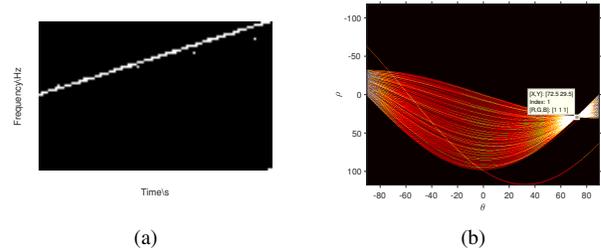


Fig. 6. Hough transform processing. (a) Black and white image obtained after thresholding of Fig. 5 (b) Result of Hough transform applied to black and white image.

the LFM parameters by the Hough transform, we need to perform the following steps. Firstly, we convert Fig. 5 into a 2D black and white image by selecting 30 dB as a threshold as shown in Fig. 6(a). Then we apply the Hough transform to calculate the gradient of the white line as shown in Fig. 6(b) and subsequently estimate the LFM parameters.

In Fig. 6(b), the coordinate of the peak of the Hough transform is $(72.5, 29.5)$. To express the Hough transform result in a more intuitive manner, we convert Fig. 6(b) into a 3D plot as shown in Fig. 7.

From Fig. 7 with the high resolution zoom shown in Fig. 6(b), we are able to identify parameter values of $\theta = 72.5^\circ$ and $\rho = 29.5$. When considering the characteristics of the LFM waveform, which possesses a frequency that increases with time, we fit the data into (15). Thus, the expression of white line in Fig. 6(a) is as follows:

$$y = 0.3153x + 30.9316. \quad (16)$$

In the meantime, regarding the scale of the axes in Fig. 5, we set $y = 0.08f_{\text{LFM}}(t)$ and $x = 12.5t$ in (16). Furthermore, the line has a y-intercept of 30.9 in (16); this is approximately equal to the frequency bin at which ‘‘DC’’ (0 Hz) is situated, such that

$$f_{\text{LFM}}(t) = 49.2656t + 0. \quad (17)$$

Then, regarding (2), we obtain the parameter estimation as follows: $\hat{f}_l = 49.2656$ Hz and $\hat{f}_c = 0$ Hz. When compared with the ground truth parameter value, $f_l = 50$ Hz, the parameter estimation error is 1.47%.

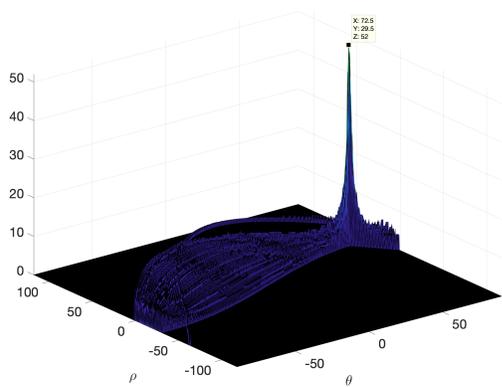


Fig. 7. Peak of Hough Transform of Fig. 6(a)

V. DISCUSSION AND FUTURE WORK

This paper discusses one alternative way to estimate LFM parameters by reusing an OFDM receiver instead of a traditional LFM receiver. Compared with a traditional LFM estimation method, such as the DCFT, the alternative OFDM receiver approach combined with the Hough transform is a new technique for waveform parameter estimation. Moreover, this alternative method is not restricted by the values of N and M selected in the DCFT, where N must be an odd number and M should be in the interval $(0, 1]$. Furthermore, the estimation range of the alternative method mainly depends on the resolution selected in the Hough transform, however the accuracy of the DCFT depends on both the values of N and M . In addition, the alternative method is able to utilise the OFDM receiver to detect the LFM parameters, which provides the possibility of formulating an RCSS implementation.

In future research, we will investigate the implementation of the orange blocks as shown in Fig. 1 for joint communication and radar scenarios. Since CS techniques are able to reduce the data dimensionality, we will explore how to apply general CS techniques to this CRSS application and how to implement them efficiently for use on low power devices. Exploiting CS, radar and communication data is able to be stored in compact form and can be reconstructed without much distortion. In particular, the CS technique in [10] achieves data recovery from compressed data using a Bayesian inference model and the authors were able to use projection design in a CS scenario to control the input-output mutual information terms for two independent sources, which could possibly be radar and communication data, respectively. Employing the method from [10] will further extend our current research into estimating the LFM parameters by reusing OFDM receivers for the purpose of radar signal detection.

ACKNOWLEDGMENT

This work was supported by the Engineering and Physical Sciences Research Council of the UK (EPSRC) Grant number EP/S000631/1 and the UK MOD University Defence

Research Collaboration (UDRC) in Signal Processing. The authors gratefully acknowledge insightful discussions with Prof. Bernie Mulgrew and Dr Mehrdad Yaghoobi.

REFERENCES

- [1] F. L. et. al., "Joint radar and communication design: Applications, state-of-the-art, and the road ahead," *IEEE Trans. Commun.*, vol. 68, no. 6, pp. 3834–3862, 2020.
- [2] D. Garmatyuk, J. Schuerger, and K. Kauffman, "Multifunctional software-defined radar sensor and data communication system," *IEEE Sensors Journal*, vol. 11, no. 1, pp. 99–106, 2011.
- [3] C. Sturm and W. Wiesbeck, "Waveform design and signal processing aspects for fusion of wireless communications and radar sensing," *Proceedings of the IEEE*, vol. 99, no. 7, pp. 1236–1259, 2011.
- [4] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.
- [5] N. Levanon and E. Mozeson, *Radar signals*. John Wiley & Sons, 2004.
- [6] A. Berni and W. Gregg, "On the utility of chirp modulation for digital signaling," *IEEE Trans. Commun.*, vol. 21, no. 6, pp. 748–751, 1973.
- [7] Z. Zhao and D. Jiang, "A novel integrated radar and communication waveform based on LFM signal," in *ICEIEC 2015*, pp. 219–223.
- [8] L. Zheng, M. Lops, and X. Wang, "Adaptive interference removal for uncoordinated radar /communication coexistence," *IEEE Journal of Selected Topics in Signal Processing*, vol. 12, no. 1, pp. 45–60, 2018.
- [9] M. A. Davenport, P. T. Boufounos, M. B. Wakin, and R. G. Baraniuk, "Signal processing with compressive measurements," *IEEE Journal of Selected topics in Signal processing*, vol. 4, no. 2, pp. 445–460, 2010.
- [10] F. K. Coutts, J. Thompson, and B. Mulgrew, "Gradient of mutual information in linear vector Gaussian channels in the presence of input noise," in *28th EUSIPCO*. IEEE, 2020, pp. 2264–2268.
- [11] P. M. Djuric and S. M. Kay, "Parameter estimation of chirp signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 12, pp. 2118–2126, 1990.
- [12] M. Z. Ikram, K. Abed-Meraim, and Yingbo Hua, "Estimating the parameters of chirp signals: an iterative approach," *IEEE Trans. Signal Process.*, vol. 46, no. 12, pp. 3436–3441, 1998.
- [13] J. Speer, K. George, and D. Mutz, "FFT based algorithm to demodulate high frequency chirp signals," in *IEEE Aerospace Conference*, 2016, pp. 1–6.
- [14] C. C. et. al., "Chirp signal detection using FFT peak frequency difference [correspondence]," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 52, no. 3, pp. 1449–1453, 2016.
- [15] J. G. et. al., "Parameter estimation of multicomponent chirp signals via sparse representation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 47, no. 3, pp. 2261–2268, 2011.
- [16] Xiang-Gen Xia, "Discrete chirp-Fourier transform and its application to chirp rate estimation," *IEEE Trans. Signal Process.*, vol. 48, no. 11, pp. 3122–3133, 2000.
- [17] F. Millioz and M. Davies, "Sparse detection in the chirplet transform: Application to FMCW radar signals," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2800–2813, 2012.
- [18] R. O. Duda and P. E. Hart, "Use of the Hough transformation to detect lines and curves in pictures," *Communications of the ACM*, vol. 15, no. 1, pp. 11–15, 1972.
- [19] IEEE, "IEEE standard for information technology–telecommunications and information exchange between systems local and metropolitan area networks–specific requirements part 11: Wireless LAN medium access control (MAC) and physical layer (PHY) specifications."
- [20] R. H. Alsisi and R. K. Rao, "Constant envelope DCT- and FFT-based OFDM systems with continuous phase chirp modulation over fading channels," in *ICITISEE*, 2017, pp. 211–216.
- [21] S. Phonsri, S. S. Mukherjee, and M. Sellathurai, "Computer vision and bi-directional neural network for extraction of communications signal from noisy spectrogram," in *CAMA*, 2015, pp. 1–4.
- [22] F. e. a. Xu, "Parameter estimation of multi-component LFM signals based on STFT+Hough transform and fractional Fourier transform," in *2nd IMCEC*, 2018, pp. 839–842.

Joint Spatio-Temporal Bias Estimation and Tracking for GNSS-Denied Sensor Networks

Sofie Macdonald* and James R. Hopgood†

Institute for Digital Communications, School of Engineering

University of Edinburgh

Email: *s1229110@ed.ac.uk, †james.hopgood@ed.ac.uk

Abstract—Sensor parameter estimation is a key process that must be considered when performing data fusion in a multi-sensor object tracking scenario. For example, significant relative time delays in sensor data arriving at a fusion centre can result in a reduction of track accuracy, false tracks, or early termination of a true object track. The same issues may arise in the presence of some relative angular bias between sensors. This article presents a technique for simultaneous target tracking and estimation of relative time delays and angular biases in data for a multi-sensor system with no access to a global frame of reference. The proposed technique makes use of a hierarchical Bayesian model and couples a grid-based search method with an array of augmented state Kalman filters to accomplish this. Results are provided comparing the root-mean-squared error in a simulated single object tracking scenario. The performance of a single sensor, two sensors with correct registration, two sensors with incorrect registration, and two sensors with registration correction are compared. The results demonstrate a significant improvement in tracking performance when registration errors are corrected with the proposed method, as well as an increase in accuracy over object tracking with only a single sensor.

Index Terms—hierarchical Bayesian model, spatio-temporal alignment, single object tracking, GNSS-denied, bias estimation, Kalman filter, grid-based search

I. INTRODUCTION

A. Problem Overview

Sensor data fusion is a valuable technique in sensor networks. Use of multiple sensors to carry out a task – such as behaviour analysis – has advantages over the same task performed by only a single sensor [1]. Calibration of these sensors is key to their reliable performance. Consider the rapid deployment of radars on ships where there may be no time for calibration ahead of time - impromptu registration becomes a necessary capability. Sensors with unknown relative registration biases are likely to return data which can lead to, in the case of object tracking, false tracks or complete loss of tracks. In a crisis situation, this could be disastrous.

Sensors may not share a common spatial reference frame (spatial bias) and may not be synchronised in time (temporal bias). Therefore, to allow successful data fusion, it is important to remove these biases such that the sensors share a frame of reference (FoR). Additionally, the network may not have any reliable access to an external reference such as the Global Positioning System (GPS). GPS can provide much information about sensor position and pose, however, in a real scenario this will likely be with some error. Furthermore, regions

exist which are entirely GPS-denied. These include indoors, underwater, hostile regions (where signals may experience interference), and in space exploration.

Developed in 1981, the Network Time Protocol (NTP) is a method for the clock synchronisation of computer systems [2]. It requires access to the internet. The sensors considered in this work have no internet access and therefore NTP is not considered a solution to the problem this work addresses. Additionally, NTP deals with round trip delays and overall clock synchronisation, which is not considered here.

B. State-of-the-Art & Contributions

Registration methods for networks of sensors have been proposed previously in the literature. Some fall into the category of spatial alignment [3], [4], while others focus primarily on temporal alignment [5]. A new KF-based algorithm for spatial bias estimation among stationary time synchronised sensors tracking N targets is proposed in [6]. This is founded on the reconstruction of Kalman gains at the fusion centre. Fortunati et al. [7] consider the spatial alignment problem between local and remote sensors and derive a linear LS estimator to align the data. A modified exact maximum likelihood (MEML) registration algorithm, which was shown to outperform the standard exact maximum likelihood (EML) algorithm in a small radar network, has been presented [8]. This was achieved through the determination of an exact likelihood function. A neural EKF (NEKF) has been developed for the alignment of two-sensor systems [9], while in [10] a deep learning based 3D point cloud registration system is proposed. In 2013, [11] presented a Bayesian algorithm based on importance sampling to estimate sensor bias for asynchronous sensors. Recently, [12] formulate a nonlinear LS approach for the three dimensional asynchronous multi-sensor registration problem.

The works mentioned above do not consider joint tracking and spatio-temporal alignment. This paper builds on the work in [13]. An algorithm is proposed for the spatial and temporal alignment of co-located radars with relative registration errors between them. The method allows simultaneous joint bias estimation and object tracking and does not rely on access to a global FoR. Instead, the radars are calibrated relative to each other. A grid-based search method is implemented. A two dimensional grid represents the bias hypothesis state space, where one dimension represents temporal bias hypotheses and the other represents spatial bias hypotheses. In this work

temporal bias is a fixed, integer value of radar sampling interval, whilst spatial bias is a relative angular offset between the radars. The method is implemented within a hierarchical Bayesian model (HBM) [14] – which is a powerful tool for state prediction – and an array of augmented state Kalman filters (ASKFs) utilised for the object tracking. A likelihood function suited for KFs is derived and evaluated during the data filtering stage. Values assigned to bias pairs are then used to update the corresponding weights of points on the grid. This work considers centralised networks – in other words, all measurements collected by all sensors are transmitted to a data fusion centre. A plot fusion architecture is employed. The novel contributions of this work include a joint sensor calibration and object tracking method implemented within an HBM, and the derivation of a likelihood function for the parameter estimation. An analysis of a range of simulations where sensor configuration is varied is provided.

C. Paper Organisation

Section II presents the joint spatio-temporal estimation problem and an overview of the HBM processes; Section III provides model definitions, scenarios, and implementation; results are displayed in Section IV; and Section V provides a conclusion and brief discussion of future work.

II. PARAMETER ESTIMATION AND DATA FUSION

The main problem addressed in this work is that of estimating a relative temporal delay, τ , in sensor data and a relative angular offset, ϕ , between multiple sensors. These are estimated jointly and in a recursive manner alongside the object tracking procedure. Consider the scenario of two co-located sensors tracking the same object. Both sensors collect measurements with some time-varying measurement noise. The FoR of one sensor is rotated by ϕ relative to that of the other sensor and there is relative delay τ in data transmission from this sensor. These biases must be corrected so that accurate sensor fusion can be accomplished. The framework of choice is the HBM. This hierarchical approach has been successfully applied to the solution of problems in a wide range of fields. These include spatio-temporal forecasting in urban traffic modelling [15], unsupervised learning and estimation of crowd emotional states in crowd monitoring [16], and the modelling of the brain cortex for pattern recognition [17], amongst others. In the field of target tracking, HBMs have been used for joint multiple-target tracking (MTT) and registration parameter estimation [18] and for simultaneous localisation and mapping (SLAM) [19]. The HBM utilised in this work has two levels: the high-level process (known as the parent process) estimates the unknown (or ‘hidden’) parameters, i.e. the sensor calibration parameters; and the low-level process (known as the offspring process) estimates the object states, i.e. the tracking function. The two processes are linked by a likelihood function calculated in the offspring process and employed for parameter estimation in the parent process. This function is problem dependent.

A. Offspring Process

The tracking problem considered is linear when the calibration is known and so the offspring process utilises a bank of ASKFs for the tracking procedure. It is important to describe it here as this is where the bias hypotheses are incorporated. An ASKF is a Kalman filter with augmented state vector and extended transition and observation matrices. Augmentation of the state vector at time k involves the concatenation of previous state estimates with the current state estimate in the following manner:

$$\hat{\mathbf{X}}_k = [\hat{\mathbf{x}}_k \quad \cdots \quad \hat{\mathbf{x}}_{k-\hat{\tau}_{max}}]^T \quad (1)$$

Here, $\hat{\tau}_{max}$ denotes the largest temporal bias hypothesis made in units of the sampling interval, δt , and:

$$\hat{\mathbf{x}}_k = [\hat{p}_{x,k} \quad \hat{v}_{x,k} \quad \hat{p}_{y,k} \quad \hat{v}_{y,k}]^T \quad (2)$$

Here, $\hat{p}_{x/y,k}$ is the filter object x and y position estimates, and $\hat{v}_{x/y,k}$ is the filter object x and y velocity estimates. The transition matrix of the system model, \mathbf{F}_k , is extended to become:

$$\mathbf{F}_k = \begin{bmatrix} \mathbf{F}_0 & 0 & 0 & \cdots & 0 & 0 \\ \mathbb{I} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \mathbb{I} & 0 & \cdots & 0 & 0 \\ & & & \vdots & & \\ 0 & 0 & 0 & \cdots & \mathbb{I} & 0 \end{bmatrix} \quad (3)$$

where \mathbb{I} is a 4×4 unit matrix, and:

$$\mathbf{F}_0 = \begin{bmatrix} 1 & \delta t & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & \delta t \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (4)$$

The observation matrix, \mathbf{H}_i , is also extended and arranged to reflect the bias hypothesis of any grid point i , $[\hat{\phi}_s^i \quad \hat{\tau}_s^i]$ for $s \in S = \{1, \dots, n_s\}$, the set of sensor IDs, where n_s is the total number of sensors. $\hat{\phi}_s^i$ represents the spatial bias (angular offset) hypothesis at grid point i for sensor s and $\hat{\tau}_s^i$ represents the temporal bias hypothesis at grid point i for sensor s . A rotation matrix, Θ^i , is then applied to incorporate the angular offset estimate:

$$\mathbf{H}_i = \Theta^i \mathbf{H}^i \quad (5)$$

where:

$$\Theta^i = \begin{bmatrix} \theta_1^i & 0 & 0 & \cdots & 0 \\ 0 & \theta_2^i & 0 & \cdots & 0 \\ 0 & 0 & \theta_3^i & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & \theta_{n_s}^i \end{bmatrix} \quad (6)$$

and:

$$\theta_s^i = \begin{bmatrix} \cos(\hat{\phi}_s^i) & -\sin(\hat{\phi}_s^i) \\ \sin(\hat{\phi}_s^i) & \cos(\hat{\phi}_s^i) \end{bmatrix} \quad (7)$$

\mathbf{H}^i takes the following general form:

$$\mathbf{H}^i = [\mathbf{H}_1^i \quad \mathbf{H}_2^i \quad \cdots \quad \mathbf{H}_{n_s}^i]^T \quad (8)$$

where:

$$\mathbf{H}_s^i = [\mathcal{O}_{2 \times 4\hat{\tau}_s^i} \quad \mathbf{H}_0 \quad \mathcal{O}_{2 \times (4\hat{\tau}_{max} + 4)}] \quad (9)$$

and

$$\mathbf{H}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \quad (10)$$

In the above, $\mathcal{O}_{m \times n}$ is an $m \times n$ zero matrix. The offspring process is responsible for estimating an object state as it evolves through time. Here, this obeys the nearly constant velocity (NCV) motion model. Measurements from both sensors are fused and supplied to the filter, along with a spatio-temporal bias hypothesis. With these inputs, the filter performs the state estimate update and evaluates the likelihood function. This is then passed to the parent process for the weights update.

B. Parent Process

The parent process performs the estimation of the registration parameters, ϕ and τ . With a grid-based approach the hypotheses of the registration parameters at a time-step k for sensor s are represented by a set of points and their corresponding weights $\Psi_{k,s} = \{\psi_s^i, w_{k,s}^i\}_{i=1}^N = \{[\hat{\phi}_s^i \quad \hat{\tau}_s^i]_{s=1}^{n_s}, w_{k,s}^i\}_{i=1}^N$. Each i references a different bias pair (grid point) and N is the total number of pairs. Every pair has some corresponding weight, $w_{k,s}^i$, which reflects the belief that registration parameters ψ_s^i are closest to the true values, ϕ and τ . Because the grid is not time-varying, a grid-based search method ([20], pg. 9) can be used to update the weights at each time-step. Weights are predicted and updated recursively, following [13] equations (2a) and (2b). Equation (2a) can be understood as a convolution of prior weights with a kernel function. The kernel function is selected to be the binomial distribution $r \sim \mathcal{B}(n, p)$ and, following investigation, the parameters $n = N$ and $p = 0.5$ were found to be suitable in this work. These were selected as a compromise between an uninformative distribution and one which searches the state space efficiently. The parent process is initialised with the assumption that, before any data processing, all sensor bias pairs are equally likely. Therefore, the supplied prior distribution is flat. Equation (2b) requires a likelihood function to update weights and here the Kalman filter likelihood function, $\ell_k(\psi_k^i | \mathbf{Z}_k)$, is derived from the integral form of the KF likelihood conditioned on ψ_k^i (\mathbf{Z}_k is the set of all sensor measurements up to time-step k , and \mathbf{z}_k and \mathbf{x}_k are the set of sensor measurements and the object state at time k , respectively):

$$\ell_k(\psi_k^i | \mathbf{Z}_k) = \int p(\mathbf{z}_k | \mathbf{x}_k, \psi_k^i) \times p(\mathbf{x}_k, \psi_k^i | \mathbf{Z}_{k-1}) d\mathbf{x}_k \quad (11)$$

where:

$$p(\mathbf{x}_k, \psi_k^i | \mathbf{Z}_{k-1}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}, \psi_k^i) p(\mathbf{x}_{k-1}, \psi_k^i | \mathbf{Z}_{k-1}) d\mathbf{x}_{k-1} \quad (12)$$

The Gaussian propagation identity [21] is used to evaluate both of these integrals and the resulting function simplifies to:

$$\ell_k(\psi_k^i | \mathbf{Z}_k) = \mathcal{N}(\mathbf{z}_k | \mathbf{H}_i \mathbf{F}_k \hat{\mathbf{x}}_{k-1}, \mathbf{\Lambda}_k) \quad (13)$$

where $\mathbf{\Lambda}_k$ is the covariance of $\ell_k(\psi_k^i | \mathbf{Z}_k)$ with form:

$$\mathbf{\Lambda}_k = \mathbf{R}_k + \mathbf{H}_i (\mathbf{Q}_k + \mathbf{F}_k \mathbf{P}_{k-1} \mathbf{F}_k^T) \mathbf{H}_i^T \quad (14)$$

where \mathbf{R}_k is the measurement noise covariance and defined in the measurement model in Section III-B.2) in Equation (20), and \mathbf{P}_{k-1} is the KF estimate covariance at time-step $k-1$. \mathbf{Q}_k is the filter process noise covariance, defined in Equation (17). The pseudocode for the parent and offspring prediction and update steps of this work are based loosely on that found in [13], where a HBM is used for joint registration and fusion of heterogeneous sensors.

III. MODELING, DATA, AND SCENARIOS

A. Implementation

The parent process of this hierarchical model is represented by an evenly distributed, two-dimensional grid of distinct points. A single grid point represents a joint spatio-temporal bias hypothesis with a calculated weight. The weights are continuously updated based on the value of the likelihood function, $\ell_k(\psi_k^i | \mathbf{Z}_k)$, that is the output of the offspring process.

B. Model Definitions

1) *Object Motion Model*: Multiple sensors track a single object which evolves through time according to the NCV model. This is defined as:

$$\mathbf{x}_k = \mathbf{F}_0 \mathbf{x}_{k-1} + \mathbf{w}_k \quad (15)$$

where \mathbf{x}_k is a four dimensional Cartesian state vector, with the following elements:

$$\mathbf{x}_k = [p_{x,k} \quad v_{x,k} \quad p_{y,k} \quad v_{y,k}]^T \quad (16)$$

and \mathbf{F}_0 is the state transition matrix – previously defined in Equation (4). \mathbf{w}_k represents zero-mean white Gaussian process noise with covariance matrix given by:

$$\mathbf{Q}_k = \begin{bmatrix} \frac{1}{3} \delta t^3 & \frac{1}{2} \delta t^2 & 0 & 0 \\ \frac{1}{2} \delta t^2 & \delta t & 0 & 0 \\ 0 & 0 & \frac{1}{3} \delta t^3 & \frac{1}{2} \delta t^2 \\ 0 & 0 & \frac{1}{2} \delta t^2 & \delta t \end{bmatrix} \tilde{q} \quad (17)$$

Here, δt is the sampling interval and \tilde{q} is the process noise intensity level ([20], pg. 181 and [22], pg. 269). It dictates how closely the object adheres to the CV model: a value $\tilde{q} = \mathbf{Q}_{22}/\delta t = 1$ ([22], pg. 270) produces CV motion.

2) *Measurement Model*: Sensors collect x and y data and the measurement model is defined as:

$$\mathbf{z}_k = \mathbf{H}_k^{obs} \mathbf{x}_k + \mathbf{v}_k \quad (18)$$

where:

$$\mathbf{z}_k = [\mathbf{z}_{k,1} \quad \mathbf{z}_{k,2} \quad \cdots \quad \mathbf{z}_{k,n_s}]^T \quad (19)$$

$$\mathbf{H}_k^{obs} = [\mathbf{H}_1 \quad \mathbf{H}_2 \quad \cdots \quad \mathbf{H}_{n_s}]^T$$

\mathbf{z}_k is the measurement vector at time k made up of measurements collected by each sensor, $\mathbf{z}_{k,s}$. \mathbf{H}_s is the observation matrix associated with sensor s and identical to the matrix \mathbf{H}_0 (see Equation (10)). In this work, only x - y components

of the object trajectory are observed. \mathbf{v}_k is the measurement noise and drawn from a zero-mean white Gaussian distribution with parameters: $\mathcal{N}(\mathbf{v}_k|\mathbf{0}, \mathbf{R}_k)$. \mathbf{R}_k is given by:

$$\mathbf{R}_k = \mathbb{I}_{n_s \times |\mathbf{z}_k|} \sigma_R^2 \quad (20)$$

where $\mathbb{I}_{n \times m}$ is an $n \times m$ identity matrix. This definition of \mathbf{R}_k assumes that each sensor returns the same type of observation vector (see (19)), although note that it is possible for the sensors to collect different types of measurements. σ_R is the standard deviation of the multivariate Gaussian distribution that describes the measurement noise, \mathbf{v}_k .

C. Scenarios

The proposed algorithm is tested in an artificial scenario with the goal of demonstrating that the method is able to calibrate sensors effectively to the baseline. This is a preliminary test and therefore a basic setup is considered: two co-located radars and a fusion centre receiving updates from them every $\delta t = 0.5s$. Data is simulated of a single NCV object travelling in two dimensional space – i.e. only x and y coordinates and their associated rates of change, \dot{x} and \dot{y} , are measured by the radars. The simulation is run for a total of $t = 2000s$. Clutter and false alarms are not considered and, for all cases, probability of detection, p_d , is assigned the value 1. In this work, $n_s = 2$ and it is assumed that sensor $s = 2$ possesses the spatio-temporal biases, whilst sensor $s = 1$ is treated as the reference sensor. Standard deviation of measurement noise variance, σ_R , is selected to be $10m$ following those assigned by comparable works [23], [24], [25], whilst the process noise intensity, \tilde{q} , is set to 1 .¹ The initial object state is $[10 \ 10 \ 10 \ 10]^T$ and the initial object state tracker estimate is $[10 \ 5 \ 10 \ 5]^T$. Both radars are placed and fixed at co-ordinates $(0,0)$. Four distinct cases for the single object tracking are analysed. These are as follows:

- 1) single sensor;
- 2) two centralised sensors, correctly registered;
- 3) two centralised sensors, with relative spatio-temporal bias, and the proposed method for bias correction;
- 4) two centralised sensors, with bias, and no correction.

In cases 3) and 4) the temporal bias is assigned to $1 \times \delta t$ and angular bias to 10° . Sampling rate for all cases is $\delta t = 0.5s$.

IV. RESULTS

The results that follow have been averaged over 100 Monte Carlo (MC) runs and the temporal and angular biases jointly estimated using the maximum a posteriori (MAP) estimate of the likelihood function. The RMSE is the chosen metric for evaluation of tracking performance. RMSE is calculated for Euclidean distance over a range of σ_R values and \tilde{q} levels over the final quarter of the full simulation time, t (to allow for filter stabilisation). Figure (1) provides log plots of the RMSE of Euclidean distance as σ_R and \tilde{q} vary, respectively. Each data point represents the average over all 100 MC trials. A new

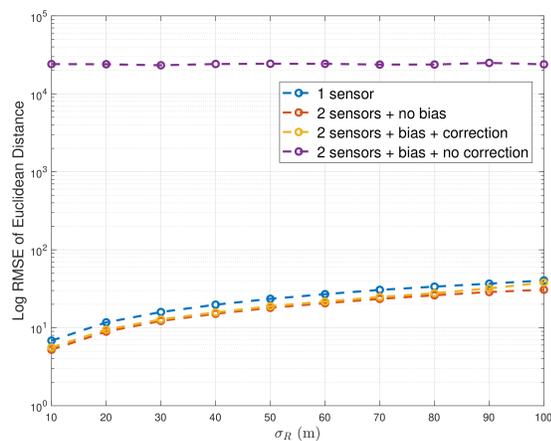
¹Note that this \tilde{q} value is the one used to generate tracking performance results against measurement noise standard deviation, σ_R .

object trajectory and set of sensor measurements is generated in each trial. Both plots provide a legend to link the data with the different test cases (for reference, see Section III-C). From these graphs, we can see that the worst performance, by far, is demonstrated by the configuration of two sensors with relative registration errors that are not corrected. This clearly shows how important it is to address the registration problem. Using the proposed method for correction, performance of the sensors is very close to that of the benchmark scenario, and outperforms the single sensor for all tested values of σ_R and \tilde{q} . For example, at the lowest \tilde{q} value of 10 the corrected case differs from the benchmark case by $\approx 0.4m$; whilst the single sensor case differs by $\approx 1.6m$ ($\approx 4 \times$ the difference of the corrected case); and the no correction case is out by $\approx 5,500m$. Whereas, when \tilde{q} is at its greatest tested value of 80, the correction case differs by $\approx 1.5m$ from the benchmark; the single sensor by $\approx 2.6m$; and the uncorrected case by $\approx 42,000m$. A similar trend can be seen in the measurement noise graph. However, all scenarios display an increase in RMSE with increasing σ_R and \tilde{q} .

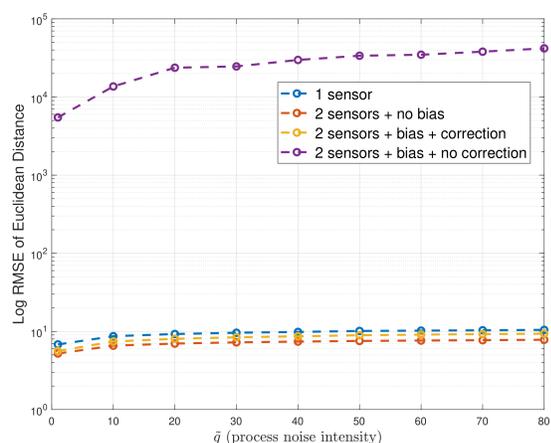
Although the key result has been demonstrated with this method and sensors are calibrated to the baseline, it is important to note that the issue of scalability must be addressed in the future. Due to the use of grid-based search, the number of grid points N (i.e. number of ASKFs required in the offspring process) relates to the number of registration errors, $|\psi_k^i|$, present as $N = \prod_{j=1}^{|\psi_k^i|} \chi_j$. χ_j represents the number of hypotheses made per error j . This shows that with every additional error, an additional dimension appears on the grid. A particle filter method may improve scalability.

V. CONCLUSION & FUTURE WORK

This article demonstrates the successful application of a method for simultaneous joint spatio-temporal alignment of sensors and object tracking. The registration parameters are continuously estimated as the object tracker runs, based on the performance of the tracker itself. The metric for the tracker performance is a KF likelihood function. It is used to update grid weights in the parent process of the HBM. Simulation results demonstrate how vital it is to have correctly calibrated sensors, and also show that tracking with multiple sensors is more accurate than tracking with only a single sensor. The work described in this article acts as a foundation for further investigation of the sensor alignment problem. The method can be extended to non-linear systems where, rather than collecting and processing data in the Cartesian frame, data is collected and processed in the polar frame. This is more realistic for radar systems. A new likelihood function would be required for the non-linear tracker. Varying probability of detection and introducing false alarms will also make the model more realistic. Increasing the number of sensors and targets is another useful avenue of investigation. Additionally, resampling and propagation can be introduced to the grid-based search method. A suitable resampling technique may allow for faster selection of a bias state - although it may add computational complexity. The usefulness of this method for



(a)



(b)

Fig. 1: 1(a) shows performance comparison for cases 1)-4) for increasing σ_R and 1(b) for increasing \tilde{q} .

non-co-located sensors can be explored, as well as for time-varying spatio-temporal biases and fractional temporal biases.

VI. ACKNOWLEDGMENT

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant number EP/S000631/1; and the MOD University Defence Research Collaboration (UDRC) in Signal Processing. Further thanks to Leonardo MW Ltd.² for their continual support of this project.

REFERENCES

- [1] W. Koch, "Tracking and Sensor Data Fusion," Tech. Rep., 4. [Online]. Available: <http://www.springer.com/series/8445>
- [2] S. Rangaswamy, "An Overview of Network Time Protocol." [Online]. Available: <https://www.researchgate.net/publication/353119594>
- [3] A. Burguera, F. Bonin-Font, and G. Oliver, "Towards robust image registration for underwater visual SLAM," *VISAPP 2014 - Proceedings of the 9th International Conference on Computer Vision Theory and Applications*, vol. 3, pp. 539–544, 2014.
- [4] C. Guo, X. Tong, S. Liu, S. Liu, X. Lu, P. Chen, Y. Jin, and H. Xie, "High-precision attitude estimation method of star sensors and gyro based on complementary filter and unscented kalman filter," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives*, vol. 42, no. 3W1, pp. 49–53, 2017.
- [5] Y. L. Zhao, X. Z. Ma, X. D. Gao, L. Yun, and S. Q. Fu, "Data fusion algorithm for asynchronous radar networks under blanket jamming," in *Applied Mechanics and Materials*, vol. 157–158, 2012.
- [6] E. Taghavi, R. Tharmarasa, T. Kirubarajan, Y. Bar-Shalom, and M. McDonald, "A practical bias estimation algorithm for multisensor-multitarget tracking," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 1, pp. 2–19, 2016.
- [7] S. Fortunati, A. Farina, F. Gini, A. Graziano, M. S. Greco, and S. Giompapa, "Least squares estimation and cramer-Rao type lower bounds for relative sensor registration process," *IEEE Transactions on Signal Processing*, vol. 59, no. 3, pp. 1075–1087, 2011.
- [8] D. Yunlong, H. You, and W. Guohong, "Exact Maximum Likelihood Registration Algorithm," *3rd International Conference on Computational Electromagnetics and Its Applications Proceedings*, pp. 85–88, 2004.
- [9] K. A. Kramer, S. C. Stubberud, and J. A. Geremia, "Target registration correction using the neural extended kalman filter," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 7, pp. 1964–1971, 2010.
- [10] A. Kurobe, Y. Sekikawa, K. Ishikawa, and H. Saito, "CorsNet: 3D Point Cloud Registration by Deep Neural Network," *IEEE Robotics and Automation Letters*, vol. 5, no. 3, pp. 3960–3966, 2020.
- [11] B. Ristic, D. E. Clark, and N. Gordon, "Calibration of multi-target tracking algorithms using non-cooperative targets," *IEEE Journal on Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 390–398, 2013.
- [12] S. Jiang, W. Pu, and Z.-q. Luo, "Optimal Asynchronous Multi-Sensor Registration in 3 Dimensions," *2018 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2018 - Proceedings*, pp. 81–85, 2018.
- [13] D. Cormack, I. Schlangen, J. R. Hopgood, and D. E. Clark, "Joint Registration and Fusion of an Infrared Camera and Scanning Radar in a Maritime Context," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 2, pp. 1357–1369, 2020.
- [14] A. M. Zaslavsky and S. James Press, "Hierarchical Bayesian Modeling Subjective and Objective Bayesian Statistics: Principles, Models, and Applications," Tech. Rep., 2003.
- [15] T. Alghamdi, K. Elgazzar, and T. Sharaf, "Spatiotemporal Prediction Using Hierarchical Bayesian Modeling," *2020 International Conference on Communications, Signal Processing, and their Applications (ICCSPA)*, pp. 1–6, 2020.
- [16] O. J. Urizar, M. S. Baig, E. I. Barakova, C. S. Regazzoni, L. Marcenaro, and M. Rauterberg, "A hierarchical bayesian model for crowd emotions," *Frontiers in Computational Neuroscience*, vol. 10, no. JULY, 7 2016.
- [17] A. S. Nadig and B. Potetz, "A hierarchical Bayesian model for pattern recognition," in *Proceedings of the International Joint Conference on Neural Networks*, 2012.
- [18] D. Cormack and J. R. Hopgood, "Message Passing and Hierarchical Models for Simultaneous Tracking and Registration," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 9251, no. c, pp. 1–13, 2021.
- [19] C. S. Lee, D. E. Clark, and J. Salvi, "SLAM with dynamic targets via single-cluster PHD filtering," *IEEE Journal on Selected Topics in Signal Processing*, vol. 7, no. 3, pp. 543–552, 2013.
- [20] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond The Kalman Filter: Particle Filters For Tracking Applications*. Boston, MA: Artech House, 2004.
- [21] M. Toussaint, "Lecture Notes: Gaussian identities," Tech. Rep., 2011.
- [22] Y. Bar-Shalom, X.-R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*. John Wiley & Sons, Inc., 2001.
- [23] S. Bu, C. Zhou, and G. Zhou, "Simultaneous spatiotemporal bias and state estimation for asynchronous multi-sensor system," *The Journal of Engineering*, vol. 2019, no. 19, pp. 5824–5828, 10 2019.
- [24] M. Li, Z. Jing, H. Pan, and P. Dong, "Joint registration and multi-target tracking based on labelled random finite set and expectation maximisation," *IET Radar, Sonar and Navigation*, vol. 12, no. 3, pp. 312–322, 2018.
- [25] Y. Bar-Shalom, "Multisensor target tracking performance with bias compensation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 42, no. 3, pp. 1139–1149, 2006.

²<https://uk.leonardocompany.com/en/home>

Detection of Human Target Location Under Simulated Randomized Rubble Using Global Fresnel's Reflection Coefficient

Amit Sarkar

School of Electrical Sciences
Indian Institute of Technology Bhubaneswar
 Khordha, Odisha, India
 as58@iitbbs.ac.in

Debalina Ghosh

School of Electrical Sciences
Indian Institute of Technology Bhubaneswar
 Khordha, Odisha, India
 deghosh@iitbbs.ac.in

Abstract—There has been a lot of cases when humans get buried under a pile of rubble created by the collapsing of man-made structures and sometimes due to natural causes. In such cases, it's imperative that they are extracted as soon as possible. Ultra-wideband impulse radar is being used for such purposes, for detecting micro as well as macro motions created by the target. This paper presents a simulated realistic rubble-target model that will incorporate real-life electrical and spatial properties of the rubble and human thorax. The global reflection coefficient is derived for the entire multi-layered rubble and target environment using Fresnel's coefficients and validated using Green's function with a mean absolute error of 0.6013 % and correlation of 97.37 %. Based on the thorax movement, radar is able to detect the location of the target using the variance approach. Considering the model simulation parameters, rubble properties, and attenuation suffered by the signal when travelling through the rubble, the maximum detection range of the target was set to 3 m. When the target was 3 m away, the maximum mean absolute error in the target range was obtained at 28 cm which is within the dimensions of an average human thorax.

Index Terms—Fresnel's coefficient, Green's function, human thorax, multi-layered rubble, rubble-target model, ultra-wideband impulse radar, variance.

I. INTRODUCTION

Over the past few years, radar has been actively used as the technology to detect buried survivors. So, in order to understand how a radar locates a person breathing within the rubble, we need to develop a system of a theoretical approach which can describe the rubble, target, and the complete radar-target environment. Upon inspecting the working of such a system, we can have a generalized understanding of how radar can behave under different varying conditions.

Current as well as past studies pertaining to radar applications, have most of the time focussed on human activity detection and classification [1], fall detection of elderly [2], gesture recognition [3], etc. Apart from these areas, ultra-wideband (UWB) radar is widely used in through-wall target detection applications. In [4], attenuation, electric permittivity, and conductivity were estimated for brick and concrete walls at 900 MHz band by taking measurements of through-wall transmission and comparing it with the already established

multi-ray model. As the knowledge of wall effects on signal propagation was steadily increasing, many research groups started focusing on target detection through-wall. Now, in order to effectively locate the target through-wall, the effect of the wall needs to be nullified. The authors in [5] used various filtering methods to mitigate wall effect for detecting a target in sitting position and obtain its heartbeat information using commercially available radar which was successfully in sync with a pulse sensor attached to its finger. Based on the nature of the target environment during the collapse of any structures, just wall-based analysis won't be enough. As a result, some of the literature has been proposed on target detection under a pile of uniform rubble made of bricks and concrete. In [6], they designed an UWB radar at 500 MHz center frequency which was able to detect respiration of target at 2 m distance. In almost all the cases, the rubble thickness is not enough and is composed of only brick and concrete. When a structure collapses, the rubble generated is a complete random mess of a variety of materials and not just concrete. With a variety of materials present, internal multiple reflections occur, which impacts the performance of target detection. So, with its absence, there's a type of uniformity that doesn't represent the nature of rubble. Hence, a more detailed analysis is required on this front.

In this paper, in Section II we modeled the rubble as rectangular slabs with randomized complex permittivity, conductivity, and thickness. We placed a human target model in between the slabs. We modeled the human target by defining various layers of the body as rectangle slabs and varied the width of the overall slab, giving the impression that the person is breathing. Then we analyzed the global reflection coefficient through the rubble-target model using Fresnel's reflection coefficient and Green's function in Section III. Then in Section IV we obtained the global received signal accuracy between the two methods and used variance statistics to capture the location of the human target and obtain target detection capability under varying ranges and layers of rubble.

II. MULTI-LAYER RUBBLE-TARGET MODEL

We have used a sinusoidal modulated Gaussian RF pulse as shown in Figure 1a, having a frequency bandwidth from 1 GHz to 2 GHz with the center frequency at 1.5 GHz. The power spectral density (PSD) of the input signal w.r.t. the FCC mask is shown in Figure 1b. It shows that our input signal is compliant with the FCC norms.

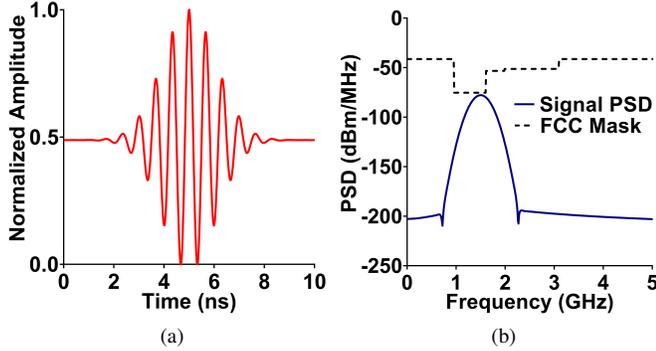


Fig. 1: (a) Input Gaussian pulse (b) PSD of input signal.

We have used an L-band radar input signal as most of the commercial radars available in the market work in both L and S-band. Signal attenuation in this band is low compared to higher frequency bands and hence is suitable for rubble with high conductive materials.

As we know in reality the composition of a rubble is mostly random in nature. In a non-uniform random rubble architecture, modeling a wave propagation becomes tricky. When the wave passes through the slab, if the slab is homogeneous, we can calculate the angle of transmitted signal considering the electrical parameters are known (which are not known to us in a real scenario). For each slab that the wave had already gone through, there would be a number of internal reflections, which would be hard to figure out as the slabs would be randomly oriented. Thus, in this paper, the nature of rubble and the target present inside it, as a result of the collapse of man-made structures, has been considered as a layered model to simplify the analysis of wave propagation through it. A generalized model of rubble created during the collapse of man-made structures is shown in Fig. 2.

Different materials comprising the rubble were organized in a homogeneous layered format. The entire rubble-target model was divided into three parts: top rubble layers, human tissue layers, and bottom rubble layers. These layers were then placed on top of a medium-dry ground. Concrete, brick, and wood were chosen as they form the backbone for any man-made construction. They are most likely the items to be found in the rubble left behind after the collapse of such man-made structures. For our analysis, the electrical and spatial properties of the layers were randomized to accommodate the properties of these materials, such that they mimic real-life rubble conditions to an extent. The electrical properties of these materials were obtained from [7].

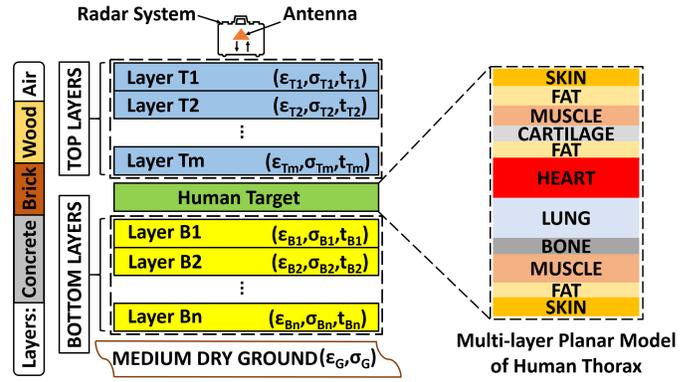


Fig. 2: Multi-layer planar rubble-target model.

For the human thorax model, the human tissue layers were derived from a cross-section of the Duke’s anatomy [8] passing through the heart. Due to the UWB nature of the radar signals, the human tissues have a dispersive nature, which is accounted for by considering the Cole-Cole model whose parameters are computed from Gabriel *et al.* [9].

III. ANALYSIS OF GLOBAL REFLECTION COEFFICIENT THROUGH RUBBLE-TARGET MODEL

The UWB signal was used to penetrate the rubble-target model all the way to the ground and back to the source. During its travel through different layers, it experiences reflection, transmission, and absorption. At each interface, there is an exchange of energy from one layer to the next. Upon accounting for all these exchanges, we were able to obtain the ultimate reflection captured from the entire model. This reflected signal takes into account the permittivity, conductivity, and thickness of all the layers. Here we have considered and analyzed two methods to obtain global reflection from the entire rubble-target model.

A. Fresnel’s Reflection Coefficient

In order to obtain the global reflection coefficient from a multi-layer planar model, we need the help of both reflection and transmission coefficient at each interface. For our analysis, we had considered slabs of thickness t_i with a refractive index of n_0, n_1, n_2 , and so on. The reflected signals from the top interface after undergoing multiple reflections as r_1, r_2, r_3 , and so on, which when accumulated gives the global reflection from the model.

These individual reflections contain multiple reflections from each slab. Upon further simplification, we obtain the global reflection coefficient by incorporating the reflection and transmission coefficient for each layer iteratively as stated in [10].

$$\hat{\Gamma}_i = \frac{\rho_i + \hat{\Gamma}_{i+1}e^{-2jk_it_i/\cos\theta_i}}{1 + \rho_i\hat{\Gamma}_{i+1}e^{-2jk_it_i/\cos\theta_i}} \quad (1)$$

Where, $\rho_1 = r_{01}$. Considering there are $K + 1$ interfaces, $i = K, K - 1, \dots, 1$ and initialized by $\hat{\Gamma}_{K+1} = \rho_{K+1}$. ρ is the reflection coefficient obtained from an interface between

two slabs. Here, r_{01} is the reflection from the topmost slab where the subscript indicates the layer numbers. $e^{-jkt/\cos\theta_i}$ is the phase shift inside the slab, k is the wavenumber, θ_i is the angle of incidence of the incoming signal.

As the signal passes through layers of rubble, it experiences attenuation. This attenuation is dependent on the electrical properties of the media. The attenuation constant is expressed as follows.

$$\alpha_i = w * \sqrt{\left(\frac{\mu_0 \epsilon_0 \epsilon'_{r_i}}{2}\right)} \left[\sqrt{1 + \left(\frac{\sigma_i}{w \epsilon_0 \epsilon'_{r_i}}\right)^2} - 1 \right] \quad (2)$$

where w is angular frequency at 1.5 GHz, μ_0 is vacuum permeability, ϵ_0 is vacuum permittivity, ϵ'_{r_i} is real part of complex permittivity, and σ_i is conductivity for i layers.

The global reflection coefficient after incorporating attenuation is expressed below.

$$\Gamma_i = \hat{\Gamma}_i * e^{-2\alpha_i t_i} \quad (3)$$

The attenuation was considered for each layer and because the signal had to travel a distance of t_i , two times for each layer, the factor of 2 was added in the above equation.

B. Planar Multi-Layer Green's Function

If we had an x-axis polarized electromagnetic wave propagating in z-direction through a multilayered structure, then the solution of that waves Maxwell's equations will be given by Green's function. Then using the analysis from [11], we can obtain the spatial domain Green's function at the origin with the help of Green's function in its spectral-domain as mentioned below.

$$G_{xx}^{\uparrow}(0, w) = \frac{1}{4\pi} \int_0^{+\infty} [R_n^{TE} - R_n^{TM}] e^{-2\Gamma_n h_n} k_{\rho} dk_{\rho} \quad (4)$$

Where R_n^{TE} and R_n^{TM} are the transverse electric and magnetic global reflection coefficients respectively [12]. $\Gamma_n (= \sqrt{k_{\rho}^2 - k_n^2})$ is the wave number in vertical direction, k_n is the free space propagation constant defined by relations $k_n^2 = -\zeta_n \eta_n$, $\zeta_n = iw\mu_n$, and $\eta_n = \sigma_n + iw\epsilon_n$. Here, ϵ_n , σ_n , and μ_n are complex permittivity, conductivity, and permeability respectively. h_n is the thickness of each layer.

IV. RESULTS AND DISCUSSION

In order to mimic real-life rubble scenarios, we randomized the real part of the electrical permittivity which the rubble may be composed of. In our analysis, the distance from the top of the rubble to the antenna was fixed at 0.1 m, as the radar is usually covered in a rugged box which is placed on top of the rubble. The distance from the top of the rubble to the top skin layer of the human was fixed such that the overall one-way range from the antenna to the human was user-defined within the bracket of 1-5 m. Maintaining the user-defined range, the thickness of the rubble layers were also randomized for both the top and bottom layers. All the simulations shown below had been performed in MATLAB using custom scripts.

A. Global Received Signal Accuracy

We compared the global reflection signal obtained from Fresnel's coefficients with the Green's function, keeping the entire rubble-target parameters unchanged. The parameters of the model which resulted in Fig. 3 are mentioned as follows. For the top three layers of the rubble, we had permittivity of (2.9, 5.08, 4.52), conductivity of (7.6, 56.8, 32.9) mS/m , and thickness of (0.31, 0.36, 0.23) m respectively. And for the bottom three layers of rubble, we had permittivity of (1.82, 4.82, 5.63), conductivity of (9.2, 48.8, 52.6) mS/m , and thickness of (0.32, 0.27, 0.24) m respectively. These parameters were chosen at random but within the range of properties of rubble materials.

Fig. 3 shows the overlapped global reflected signal from the rubble-target model obtained from the Fresnel's coefficients and Green's function. It also shows the comparison of the two methods in terms of their magnitudes and phase angle. The magnitude comparison is almost identical and the phase comparison is almost identical too, except for a brief notch around 1.6 GHz frequency. As the notch was kind of an impulse in nature, there was no significant impact on the overall wave nature when compared to the wave obtained from the Green's function.

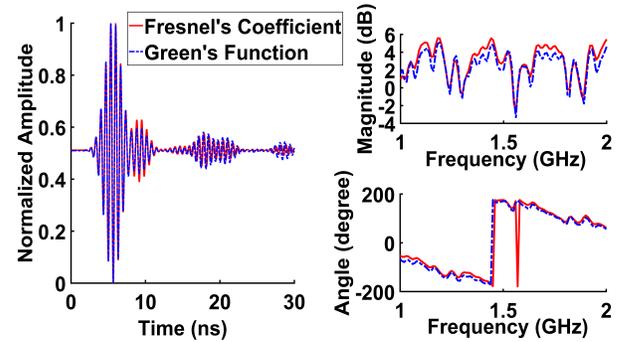


Fig. 3: Comparison of received signals using Fresnel's coefficient and Green's function.

The mean absolute error (MAE) between the two methods for this case was 0.6013 % and the correlation coefficient was 97.37 %.

B. Target Detection Accuracy

We now look forward to efficiently detect the location of the breathing target within the rubble. In the presence of a breathing human, the received signal will move to-and-fro about the target location in the time domain. This constant movement will create a variance about the target location. The details of target detection using variance statistics are described in detail in [13].

Based on the nature of the collapse of any man-made structures, there may be human beings trapped at any particular location within the rubble. Considering the simulation parameters mentioned in Table I, we calculated the maximum range of the radar.

TABLE I: Simulation parameters used for detecting maximum radar range.

System Parameters	Value
Average Transmitted Power	-41.7 dBm
Antenna Gain	7 dBi
Center Frequency	1.5 GHz
Human Thorax RCS Area	1.84 sq.m.
Minimum Detector SNR	5 dB
Minimum Detectable Signal	-95 dBm
Maximum Radar Range	3.752 m

For considering the RCS of the human thorax, we approximated the thorax region into an elliptical cylinder [14]. We opted for a gain of 7 dBi, based on an UWB Vivaldi antenna fabricated in our lab. To calculate the minimum detectable signal (MDS), we used 5 dB as the minimum detector SNR with the receiver bandwidth of 10^7 Hz. The value of -95 dBm as MDS is at par with commercial radar system receivers. With the help of the radar range equation, we obtained the maximum range of 3.752 m using the above system parameters. By looking into these parameters, we can see that the signal after exiting the transmitter antenna, can afford an attenuation of around 60 dB when interacting with the rubble-target model. If the signal on its travel through the environment attenuates more than 60 dB, then the received signal level will be less than MDS and the entire signal will be buried under the noise floor. So, in order to find the total loss faced by the signal in passing through the environment, we will take help of attenuation constant α , which is already mentioned in (2). The loss faced by the signal when travelling the top layers is defined as follows.

$$A(l) = e^{2\alpha(l)t_T(l)} \quad (5)$$

$$Loss(l) = 10 * \log_{10} A(l) \text{ dB} \quad (6)$$

$$Total_Loss = \sum_{l=1}^{T_m} Loss(l) \text{ dB} \quad (7)$$

Where $l = 1, 2, \dots, T_m$ is the number of layers situated above the target, $t_T(l)$ is the thickness of the top layers, and $A(l)$ is the attenuation of those individual layers. Based on the random selection of electrical and spatial properties of the rubble, Fig. 4 describes the total loss of the signal when travelling from the radar to the target and back to the radar. We simulated the environment with a one-way target range from 1-5 m along with a total number of layers (N) from 5 to 10.

From Fig. 4, we can observe that a cap of 60 dB attenuation limits the radar to detect the target efficiently within a range of 3.4-3.6 m depending on the number of layers of rubble. For our analysis, let's round off the maximum range to 3 m. So, we will analyze further how the MAE of breathing target detection changes with change in target range. In order to understand the nature of MAE, we looked into attenuation loss per unit length which is expressed as follows.

$$Unit_Loss(l) = \frac{\sum_{l=1}^{T_m} Loss(l)t_T(l)}{\sum_{l=1}^{T_m} t_T(l)} \text{ dB/m} \quad (8)$$

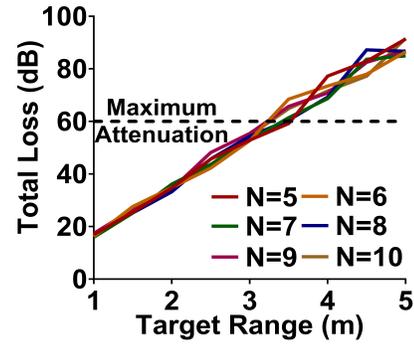


Fig. 4: Total loss w.r.t. target range for different number of layers.

We had defined $Loss(l)$ earlier as the attenuation faced by the signal in the top layers l . Here, unit loss gives us a perspective of on an average how much loss is occurring per meter downrange. Fig. 5 shows the change in MAE in range of breathing target w.r.t. change in the target range for a different number of layers of rubble. Here, the variance method was used to detect the breathing of the target and in turn its location. The target range was chosen from 1-3 m with the number of rubble layers from 5-10. When simulating for each scenario, the algorithm was run 100 times and then the average detection performance was considered for that particular case. From Fig. 5 we observed that as the target range was increasing, MAE also kept increasing. Possible explanation for this phenomenon is due to the increase in wave attenuation and dispersion as the range increases, the unit loss for each slab also increases and difference in detection capability between the two methods will start to increase further and further and hence the MAE will go on increasing. From Fig. 5 we can see that as the number of layers increases, both the MAE in range and unit loss decreases. For a particular range, if the number of layers increases, then the width of each layer is supposed to decrease. As a result, the wave losses will decrease and the detection performance as well as MAE in range will improve.

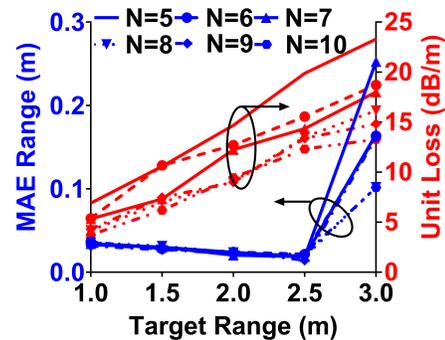


Fig. 5: Analysis of MAE and Unit Loss w.r.t. target range for different number of layers.

Looking into both the scenarios, we observed that for this particular set of data, the maximum MAE recorded was 28 cm

(this value was noted down manually from the plot and can be approximately visualized from Fig. 5). As the minimum average human thorax dimension is greater than 30 cm, the radar will be able to detect the breathing target uniquely if there is a nearby target.

V. CONCLUSION

In this paper, we focussed on the performance of radar in detecting the location of a breathing human buried under a pile of rubble. We simulated an environment of rubble-target model with real-life electrical and spatial parameters of rubble and human tissue layers. Our main goal was to add randomness to rubble characteristics so that we could analyze how a radar would be able to perform under different scenarios. We compared Green's function to that of the signal obtained from the Fresnel's coefficients to ensure its reliability. Upon randomizing the parameters of the rubble-target model, we obtained the MAE in determining the location of the breathing human target and analyzed the nature of MAE with a change in the target range. As the analysis was averaged over 100 times using random electrical and spatial properties of the rubble, we hope that real-life values of the rubble will be covered. For the target range of 3 m, the maximum MAE in range came out to be 28 cm, whose value is within an acceptable range.

ACKNOWLEDGMENT

The authors are thankful to the Fund for Improvement of S&T infrastructure in universities & higher educational institutions (FIST), R&D Infrastructure Division, Department of Science & Technology, and Government of India for the funding for upgrading the Microwave and Optical Engineering laboratory and to the University Grants Commission (UGC) for the Junior Research Fellowship (JRF). This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

- [1] D. Kumar, A. Sarkar, S. R. Kerketta, and D. Ghosh, "Human Activity Classification Based On Breathing Patterns Using IR-UWB Radar," 2019 IEEE 16th India Council International Conference (INDICON), Rajkot, India, 2019, pp. 1-4.
- [2] B. Jakanovic and M. Amin, "Fall detection using deep learning in range-Doppler radars," IEEE Trans. Aerospace & Electron. Syst., vol. 54, no. 1, pp. 180-189, 2017.
- [3] S. Ahmed and S. H. Cho, "Hand Gesture Recognition Using an IR-UWB Radar with an Inception Module-Based Classifier," Sensors, vol. 20, no. 2, pp. 564, 2020.
- [4] D. Pena, R. Feick, H. D. Hristov, and W. Grote, "Measurement and modeling of propagation losses in brick and concrete walls for the 900-MHz band," in IEEE Transactions on Antennas and Propagation, vol. 51, no. 1, pp. 31-39, 2003.
- [5] A. Sarkar and D. Ghosh, "Through-Wall Heartbeat Frequency Detection Using Ultra-Wideband Impulse Radar," 2019 International Conference on Range Technology (ICORT), Balasore, India, 2019, pp. 1-5.
- [6] H. Lv, G. H. Lu, X. J. Jing, and J. Q. Wang, "A new ultra-wideband radar for detecting survivors buried under earthquake rubbles," in Microw. Opt. Technol. Lett., vol. 52, no. 11, pp. 2621-2624, 2010.
- [7] "Effects of Building Materials and Structures on Radiowave Propagation Above About 100 MHz," in International Telecommunications Union Recommendation ITU-R P.2040-1, 2015.
- [8] M. Cavagnaro, E. Pittella, and S. Pisa, "UWB pulse propagation into human tissues," in Phys. Med. Biol., vol. 58, no. 24, pp. 8689-8707, 2013.
- [9] D. Andreuccetti, R. Fossi, and C. Petrucci, "An Internet resource for the calculation of the dielectric properties of body tissues in the frequency range 10 Hz - 100 GHz," IFAC-CNR, Florence (Italy), 1997. Based on data published by C. Gabriel *et al.* in 1996. [Online]. Available: <http://niremf.ifac.cnr.it/tissprop/>. Accessed: 1st December, 2020.
- [10] S. J. Orfanidis, "Electromagnetic Waves and Antennas," 2013. [Online]. Available: <https://www.ece.rutgers.edu/~orfanidi/ewa/>. Accessed: 28th March, 2021.
- [11] E. Slob and J. Fokkema, "Coupling effects of two electric dipoles on an interface," Radio Sci., vol. 37, pp. 1073, 2002.
- [12] S. Maiti, S. K. Patra, and A. Bhattacharya, "A Modified Plane Wave Model for Fast and Accurate Characterization of Layered Media," in IEEE Transactions on Microwave Theory and Techniques, vol. 65, no. 9, pp. 3492-3502, 2017.
- [13] A. Sarkar and D. Ghosh, "Detection of Multiple Humans Equidistant From IR-UWB SISO Radar Using Machine Learning," in IEEE Sensors Letters, vol. 4, no. 1, pp. 1-4, 2020.
- [14] B. R. Mahafza, "Radar Systems Analysis and Design Using MATLAB", 3rd ed. Boca Raton, U.S.A.: Chapman and Hall/CRC, 2016.

Semi-supervised Domain Adaptation via adversarial training

Antonin Couturier*
 Senior Data Scientist
 Thales UK
 Glasgow, Scotland
 first.last@uk.thalesgroup.com

Anton-David Almasan*
 Data Scientist
 Thales UK
 Glasgow, Scotland
 first.last@uk.thalesgroup.com

Abstract—Whilst convolutional neural networks (CNN) offer state-of-the-art performance for classification and detection tasks in computer vision, their successful adoption in defence applications is limited by the cost of labelled data and the inability to use crowd sourcing due to classification issues.

Popular approaches to solve this problem use the expansive labelled data for training. It would be more cost-efficient to learn representations from the unlabelled data whilst leveraging labelled data from existing datasets, as empirically the performance of supervised learning is far greater than unsupervised-learning.

In this paper we investigate the benefits of mixing Domain Adaptation and semi-supervised learning to train CNNs and showcase using adversarial training to tackle this issue.

Index Terms—Domain Adaptation, Semi-supervised learning

I. INTRODUCTION

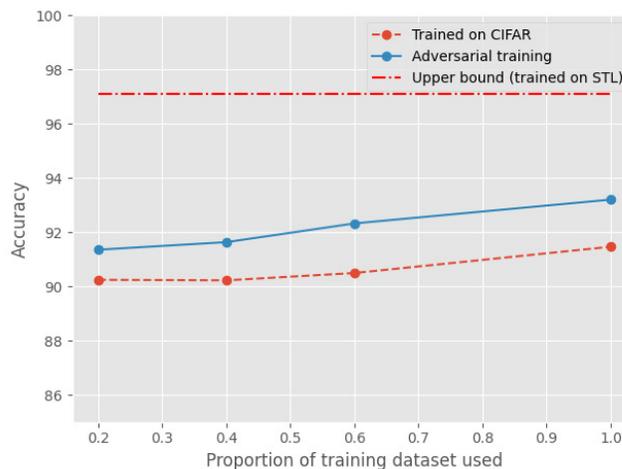
Convolutional Neural Networks offer state-of-the-art performance for classification and detection tasks in computer vision [1]. One aspect limiting their adoption and success in defence application is the cost of labelled data. The performance of a CNN usually degrades when moving from an academic dataset to real-life data. Moreover, whilst academia uses crowd sourcing to cheaply generate labels for large datasets, such a mechanism is not suitable when the images are classified.

Popular approaches to this problem usually imply training the CNN on a large dataset, and then fine-tune it on the existing labels from the problem at hand. Given the cost of labelling such data, the focus should be to use it for validation of the models rather than training. Therefore, it would be more cost-efficient to learn representations from the unlabelled data whilst leveraging labelled data from existing datasets, as empirically the performance of supervised learning is far greater than unsupervised-learning.

While Domain Adaptation may exhibit itself as the preferred paradigm to bridge the gap that separates labelled and unlabelled data, a vast amount of literature is oriented towards researching costly generative models. The complexity of such models render a large percentage of them unusable as real time video solutions, as defence production deployment is often constrained in size and available computing power. Furthermore, generative models have trouble dealing with complex background information such as the ones present

in natural images. Semi-supervised learning, on the other hand, can employ complex optimisation frameworks for both generative and discriminative models and usually takes the form of adversarial training.

Fig. 1. Graphical illustrations of results



II. RELATED WORK

Empirical risk minimisation works under the assumption that the joint probability distribution over inputs and targets between training and testing set is stable. In practice, algorithms trained using one dataset are deployed in real-life scenario with different sensors and scenes than ones used for training. For instance, it has been noted that ImageNet is biased towards texture [2], and that most academic vision datasets over-represent certain geographic areas due to how the data was collected.

Transfer learning in general aims at leveraging features learnt on one dataset to speed-up the training on a different dataset. Domain Adaptation in particular assumes that whilst the distribution changes between training and testing datasets, the task remains the same. Statistical learning theory offers some learning bounds for certain cases [3]. In general, Domain Adaptation is an empirical science where the validity of a method relies on empirical success rather than theoretical soundness.

* Equal contribution

In general, Domain Adaptation methods fall into three categories - discrepancy, adversarial, and reconstruction-based methods. Reconstruction-based methods encompass (variational) auto-encoders [4], [5] - we consider a compression problem where a backbone encodes a target, and a decoder reconstructs the target from the encoding, with the idea that the encoding should work for both domains. Adversarial-based methods [6] are usually made of two networks - a generator that generates images from a given distribution, and a discriminator that predicts whether the generated image is from one distribution or another. Finally, discrepancy-based methods *e.g* [7] tackle the domain shift by minimising intra and inter class discrepancy.

Much of the research done to advance Unsupervised Domain Adaptation employs Adversarial Discriminative models. [8] uses multiple adversarial Domain classifier submodules to assess the capability of an encoder to learn domain invariant latent features at multiple levels. A domain classifier uses a Scale Reduction Module (SRM) in conjunction with a Gradient Reversal layer (GRL) [9] to output a Bernoulli variable that models the probability of a feature map belonging to either source or target domain. Empirically, the SRM takes the form of a 1×1 convolution aimed at reducing the data dimensionality channel-wise, while the GRL is used to pass a negative gradient signal from the domain discriminator to the encoder in order to achieve mini-max (adversarial) optimisation. Moreover, a visual attention map [10] can be used as a pixel-wise weighting to leverage saliency variability within the feature map.

The goal of semi-supervised learning is to train a classifier using both labelled and unlabelled data. To do so, it minimises a loss function made of supervised and unsupervised losses [11].

III. METHODOLOGY

A. Training baseline models

Resnet50 [12] is used as the backbone for all the models. Following the training strategy from [13], we use increased training epochs, dropout on the fully connected layer, decreased weight decay, and cosine learning rate annealing with late restart as advised in [14]. Whilst we use increased training epochs, we also apply validation-based early stopping [15] as a regularisation method. Our data augmentation strategy is also different. For simplicity, we rely only on normalisation, resizing, cropping, and flipping. We use cross-entropy for the classification loss.

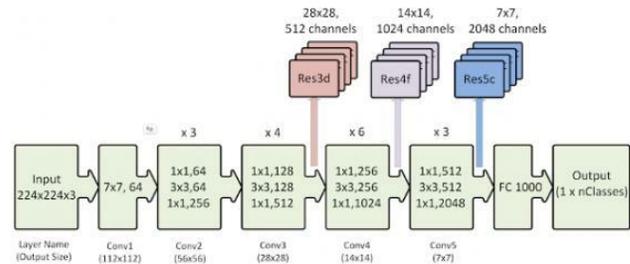
An important practical consideration is the sample efficiency of training *i.e.*, what is benefit of increasing the number of samples with respect to a given performance metric? To tackle this question we train baseline models using different sample sizes - 20, 40, 60, and 100% of the original training dataset sizes, and validation on the full testing set.

We use two datasets A and B. We train the CNN on dataset A at different sample sizes and test the network on the full testing set from dataset A, and the full testing set from dataset B. That way we can measure the cost of moving

from one environment to another. We also train a network on dataset B in order to have an upper-bound performance (corresponding to ideal performance). We therefore end up with 5-baseline networks. Those baseline networks are compared to our method using the same 4 sampling sizes - the subsamples are the same for all the methods. We do not repeat the experiment per sampling as in practice there is little variance ($8.36e - 06$) in accuracy when using the 20% sampling size and repeating the experiments 10 times.

B. Training semi-supervised models

Fig. 2. Illustration of Resnet50 architecture



Domain discriminator submodules [8] are used at three encoding levels of the Resnet50 encoder: Res3d, Res4f and Res5c (see Figure 2). Our SRM submodules are applied on feature maps with 512, 1024 and 2048 channels and have an output channel dimension of 256. Consequently, the different feature maps are sampled at different spatial dimensions, inversely proportional to the respective channel size.

Having domain discriminators at various levels allows for a more robust learning scheme as it prevents the encoder from only optimising deep latent features for domain invariance. Applying domain discriminators at various spatial sizes allows for alignment of latent features on a coarse to fine basis. The earlier domain discriminators thus act as a means of regularisation.

A Mean Squared Error (MSE) loss is computed for domain predictions at the three levels for the training sets of both datasets A and B, and validation accuracy is calculated on the test set of dataset B. Therefore the final loss when training the semi-supervised model is a weighted sum of cross-entropy classification loss (supervised) and the MSE for domain adaptation (unsupervised). Similarly to baseline models, we use different sample sizes of the original training dataset to assess the impact on semi-supervised feature learning.

IV. EXPERIMENTS

A. Datasets

Two simple and well-studied datasets are used to evaluate our method - CIFAR-10 [16] and STL-10 [17]. STL-10 is inspired by CIFAR and was designed to assess unsupervised

feature learning. It contains 5000 labelled images - that we use to train a baseline model, and 100,000 unlabelled images that we use for unsupervised learning. We evaluate the performance on the 7,200 test images. CIFAR-10 contains 50,000 training images, and 10,000 testing images. We assess the Domain Adaptation from CIFAR-10 to STL-10. In practice it is CIFAR-9 and STL-9 as the datasets different in one class.

All the models are trained on a GPU cluster for 350 epochs each. We use stochastic gradient descent with warm restart and momentum as optimiser, a 0.1 learning rate with linear scaling with respect to batch size, and no scaling for other hyper-parameters.

B. Baseline results

TABLE I
BASELINE RESULTS ON CIFAR AND STL

Training set	CIFAR-10	STL-10
CIFAR10 - 20%	94.07%	90.25%
CIFAR10 - 40%	95.21%	90.23%
CIFAR10 - 60%	95.89%	90.50%
CIFAR10 - 100%	96.66%	91.47%
STL10 (5 folds)		97.12%

Baseline results are shown in I, as expected the best validation results for CIFAR and STL correspond to trained models on full data. When focusing on STL, there is a 5.6 points difference between the best performing model trained on CIFAR, and the upper-bound performance. We observe a correlation in performance between subsets when comparing validation on CIFAR and STL ($\rho = 0.82$).

C. Adversarial training results

TABLE II
RESULTS ON STL USING ADVERSARIAL TRAINING

Training set	STL-10
Unlabelled STL + CIFAR10 - 20%	91.36%
Unlabelled STL + CIFAR10 - 40%	91.64%
Unlabelled STL + CIFAR10 - 60%	92.33%
Unlabelled STL + CIFAR10 - 100%	93.21%

All experiments illustrate an increase in accuracy when leveraging both of the datasets. This is to be expected as the model is able to incorporate domain discriminative learning signals from the source distribution as well as the target distribution during training and thus obtain more generalising capability.

D. Attention maps

By compressing the feature channel-wise through summation there is an expectation that salient features will become evident and could be extracted through thresholding [18]. Thus, by applying a softmax over this saliency map we can use it as a weighting inside the domain discriminator. However, there was no increase in accuracy when using this technique, most probably because of the simplicity of the datasets used. With one object per image at a 32 x 32 resolution, there may

not be enough spatial variability for attention map weighting to bring any improvements. As attention maps scale with the size of the input image, we expect their effect to be more apparent for high resolution images.

Below are two examples to show this scaling. In Figure 3 we have a thermal image of size 512 x 640 from the FLIR-ADAS dataset [19], alongside its respective attention map. Figure 4 illustrates a random high resolution 3000 x 1688 RGB image of a car from the internet and its attention map. We can thus see that for lower resolution images, attention map weighting can act as an object or hotspot detector, while being able to capture more fine grained features only in the high resolution image. This renders attention map weighting more suitable for object detection tasks, as opposed to classification tasks which naturally do not need high resolution imagery.

Fig. 3. Illustration of thermal image from FLIR dataset and its attention map

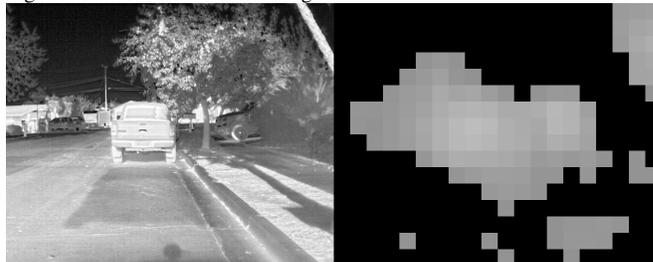


Fig. 4. Illustration of high resolution visual image and its attention map



V. FUTURE WORK

Following the experiments on Domain Adaptation via semi-supervised learning, we wish to see how more complex unsupervised learning methods can be combined with supervised learning techniques, in what we define as a Joint Training Framework (JTF). A JTF aims to use a general unsupervised learning method as a regularising factor during training of an algorithm under full supervision. The intuition behind a JTF is that the lack of labelling forces a model to create robust latent representations which can be used as priors for the supervised optimiser. Thus, we aim to combine the superior performance of supervised learning with the generalising capability of self-supervision within a complex multiple domain dataset.

STL and CIFAR are toy datasets, and the performance gap is quite small. We intend to apply this methodology for thermal and visual imagery, with thermal imagery being the target domain. We expect a higher performance gap using TI and TV and therefore more room to display the benefits of our method.

REFERENCES

- [1] T. J. Sejnowski, "The unreasonable effectiveness of deep learning in artificial intelligence," *Proceedings of the National Academy of Sciences*, vol. 117, no. 48, pp. 30 033–30 038, 2020.
- [2] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel, "Imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness," *arXiv preprint arXiv:1811.12231*, 2018.
- [3] D. McNamara and M.-F. Balcan, "Risk bounds for transferring representations with and without fine-tuning," in *International Conference on Machine Learning*. PMLR, 2017, pp. 2373–2381.
- [4] M. A. Kramer, "Nonlinear principal component analysis using autoassociative neural networks," *AICHE journal*, vol. 37, no. 2, pp. 233–243, 1991.
- [5] D. P. Kingma and M. Welling, "An introduction to variational autoencoders," *arXiv preprint arXiv:1906.02691*, 2019.
- [6] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Processing Magazine*, vol. 35, no. 1, pp. 53–65, 2018.
- [7] G. Kang, L. Jiang, Y. Wei, Y. Yang, and A. G. Hauptmann, "Contrastive adaptation network for single-and multi-source domain adaptation," *IEEE transactions on pattern analysis and machine intelligence*, 2020.
- [8] Z. He and L. Zhang, "Multi-adversarial faster-rcnn for unrestricted object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 6668–6677.
- [9] Y. Ganin and V. Lempitsky, "Unsupervised domain adaptation by back-propagation," in *International conference on machine learning*. PMLR, 2015, pp. 1180–1189.
- [10] Y. Zheng, D. Huang, S. Liu, and Y. Wang, "Cross-domain object detection through coarse-to-fine feature adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 13 766–13 775.
- [11] A. Rasmus, H. Valpola, M. Honkala, M. Berglund, and T. Raiko, "Semi-supervised learning with ladder networks," in *Proceedings of the 28th International Conference on Neural Information Processing Systems-Volume 2*, 2015, pp. 3546–3554.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [13] I. Bello, W. Fedus, X. Du, E. D. Cubuk, A. Srinivas, T.-Y. Lin, J. Shlens, and B. Zoph, "Revisiting resnets: Improved training and scaling strategies," *arXiv preprint arXiv:2103.07579*, 2021.
- [14] I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016.
- [15] L. Prechelt, "Early stopping-but when?" in *Neural Networks: Tricks of the trade*. Springer, 1998, pp. 55–69.
- [16] A. Krizhevsky, G. Hinton *et al.*, "Learning multiple layers of features from tiny images," 2009.
- [17] A. Coates, A. Ng, and H. Lee, "An analysis of single-layer networks in unsupervised feature learning," in *Proceedings of the fourteenth international conference on artificial intelligence and statistics*. JMLR Workshop and Conference Proceedings, 2011, pp. 215–223.
- [18] B. X. Nguyen, B. D. Nguyen, G. Carneiro, E. Tjiputra, Q. D. Tran, and T.-T. Do, "Deep metric learning meets deep clustering: An novel unsupervised approach for feature embedding," in *BMVC*, 2020.
- [19] "Flir thermal dataset for algorithm training," <https://www.flir.co.uk/oem/adas/adas-dataset-form/>, accessed: 2021-07-30.

Fast Givens Rotation Approach to Second Order Sequential Best Rotation Algorithms

Faizan Khattak, Stephan Weiss, Ian K. Proudler

Department of Electronic & Electrical Engineering, University of Strathclyde, Glasgow G1 1XW, Scotland

{faizan.khattak,stephan.weiss,ian.proudler}@strath.ac.uk

Abstract—The second order sequential best rotation (SBR2) algorithm is a popular algorithm to decompose a parahermitian matrix into approximate polynomial eigenvalues and eigenvectors. The work horse behind SBR2 is a Givens rotation interspersed by delay operations. In this paper, we investigate and analyse the application of a fast Givens rotation in order to reduce the computation complexity of SBR2. The proposed algorithm inherits the SBR2’s proven convergence to a diagonalised and spectrally majorised solution for the polynomial eigenvalues. We provide some analysis and examples for the execution speed of this fast Givens-based SBR2 compared to a standard SBR2 implementation.

I. INTRODUCTION

For broadband signals $\mathbf{x}[n] \in \mathbb{C}^M$ acquired by an M -element sensor arrays in discrete time $n \in \mathbb{Z}$, the space-time covariance matrix $\mathbf{R}[\tau] = \mathcal{E}\{\mathbf{x}[n]\mathbf{x}^H[n-\tau]\}$ captures the complete second order statistics of the data. The explicit lag parameter $\tau \in \mathbb{Z}$ is capable of characterising the correlation between sensor signals by relative time delays, and hence bears information on aspects such as the angle of arrival of a particular source signal. This is different from the narrowband case, where e.g. AoA is resolved simply by phase shifts, and it suffices to consider the instantaneous covariance $\mathbf{R}[0]$. To generalise narrowband optimal solutions, which often are based on the eigenvalue decomposition of $\mathbf{R}[0]$, to the broadband case, techniques have been developed to diagonalise $\mathbf{R}[\tau]$ for every value of τ . Because its z -transform $\mathbf{R}(z) = \sum_{\tau} \mathbf{R}[\tau]z^{-\tau}$, or abbreviated $\mathbf{R}(z) \bullet \circ \mathbf{R}[\tau]$, is a polynomial matrix, such techniques are referred to as polynomial matrix EVDs [1]–[3].

A polynomial matrix EVD exists in the case of an analytic $\mathbf{R}(z)$ that emerges from unmultiplexed data [2], [4], and two main families of algorithms with proven convergence have arisen over the last decade — sequential matrix diagonalisation (SMD, [5], [6]) and the second order sequential best rotation algorithm (SBR2, [1], [7]). Since applications such as subband coding [8], [9], beamforming [10], source separation [11], [12] or speech enhancement [13], [14] depend on low computation cost, various efforts have been directed at numerical [15]–[19] and implementational enhancements [19], [20].

Amongst the numerical enhancements of polynomial matrix EVD algorithms, the Givens rotation has attracted particular

We gratefully acknowledge support through Commonwealth Scholarship PKCS-2020-518. This work was also supported in parts by the Engineering and Physical Sciences Research Council (EPSRC) Grant number EP/S000631/1 and the MOD University Defence Research Collaboration in Signal Processing.

attention. The cyclic-by-row approach in [15] limited the number of rotation steps to only implement an approximate EVD within the SMD family of algorithms. Divide-and-conquer schemes [18], [19] have been aiming at reducing the spatial dimension by breaking $\mathbf{R}[\tau]$ into sub-blocks, and hence reducing the cost of matrix multiplications.

In this paper, we employ the idea of fast Givens rotations. Typically a similarity transform by a Givens rotation requires the modification of two rows and two columns of the matrix to be transformed. It is well known that square root- and division-free approaches [21]–[23] lead to simplifications and effectively split the unitary Givens rotation into a simple diagonal matrix and a simplified matrix that will only require to have the multiplications and additions of a Givens operation. In [24], this simplification is extended to successive real-valued Givens rotations and is beneficially applied to EVD and QR calculations. Here, we employ a complex-valued extension of this approach in the context of the SBR2 algorithm to reduce the SBR2’s complexity.

Therefore, in the following, Sec. II will first briefly outline the SBR2 algorithm, followed by the introduction of fast Givens rotations — both as single operations and in sequence — in Sec. III. This approach then drives a fast SBR2 version introduced in Sec. IV, which is evaluated in a numerical example and simulations in Sec. V. Finally, conclusions are drawn in Sec. VI.

II. POLYNOMIAL EVD ALGORITHMS

A. Polynomial EVD

The cross-spectral density (CSD) matrix $\mathbf{R}(z)$ satisfies the parahermitian property $\mathbf{R}^P(z) = \mathbf{R}^H(1/z^*) = \mathbf{R}(z)$, and admits a parahermitian matrix EVD (PhEVD) $\mathbf{R}(z) = \mathbf{Q}(z)\mathbf{\Lambda}(z)\mathbf{Q}^P(z)$, if $\mathbf{R}(z)$ is analytic and the measurement vector $\mathbf{x}[n]$ does emerge from multiplexed data [2], [4]. In this case, the factor $\mathbf{Q}(z)$ is paraunitary, i.e. $\mathbf{Q}(z)\mathbf{Q}^P(z) = \mathbf{I}$, and contains analytic eigenvectors in its columns. The corresponding analytic eigenvalues form the diagonal parahermitian matrix $\mathbf{\Lambda}(z)$.

A polynomial eigenvalue decomposition [1] is a modified version of the PhEVD,

$$\mathbf{R}(z) \approx \mathbf{U}(z)\mathbf{\Gamma}(z)\mathbf{U}^P(z), \quad (1)$$

where the potentially transcendental factors $\mathbf{Q}(z)$ and $\mathbf{\Lambda}(z)$ of the PhEVD are replaced by a polynomial paraunitary $\mathbf{U}(z)$ and a Laurent polynomial, diagonal, and parahermitian $\mathbf{\Gamma}(z)$.

Specifically, the analytic $\Lambda(z)$ is replaced by a spectrally majorised $\Gamma(z) = \text{diag}\{\gamma_1(z), \dots, \gamma_M(z)\}$, such that on the unit circle

$$\gamma_{m+1}(e^{j\Omega}) \geq \gamma_m(e^{j\Omega}) \quad \forall \Omega \in \mathbb{R}, \quad m = 1, \dots, (M-1). \quad (2)$$

This spectral majorisation yields eigenvalues that are different from the PhEVD if the analytic eigenvalues intersect. In this case, the polynomial EVD in (1) will require a high polynomial order for a satisfactory approximation, but is important in applications such as subband coding, where the majorisation property in (2) guarantees the maximisation of the coding gain [7], [25].

The PEVD can be calculated by a number of algorithms with guaranteed convergence. A family of second order sequential best rotation (SBR2) algorithms [1], [7], [26] aims to iteratively eliminate the largest off-diagonal component at every iteration step, whereby spectral majorisation has been shown to be enforced [27]. More recently, sequential matrix diagonalisation (SMD) algorithms [5], [6] iteratively diagonalise a time-shifted version of the space-time covariance, which converges faster per iteration but at an overall higher computational cost than SBR2.

B. Second Order Sequential Best Rotation Algorithm

SBR2 is a generalisation of the Jacobi method, whereby successive Givens rotations are interspersed by delay operations in order to move maximum off-diagonal elements to the lag zero component, where they subsequently are eliminated.

Starting from $\mathbf{S}_0(z) = \mathbf{R}(z)$, at the i th iteration SBR2 finds the maximum off-diagonal element

$$\{\tau_i, m_i, n_i\} = \arg \max_{\substack{\tau, m, n \\ m \neq n}} |s_{i-1, m, n}[\tau]|, \quad (3)$$

where $s_{i-1, m, n}[\tau]$ is the element in the m th row and n th column of $\mathbf{S}_{i-1}[\tau]$, whereby $\mathbf{S}_{i-1}[\tau] \circ \bullet \mathbf{S}_{i-1}(z)$. A delay (or advance, for $\tau_i < 0$) operation

$$\Delta_i(z) = \text{diag}\{\underbrace{1, \dots, 1}_{n_i-1}, z^{-\tau_i}, \underbrace{1, \dots, 1}_{M-n_i}\} \quad (4)$$

then ensures that $s_{i-1, m_i, n_i}[\tau_i]$ as well as $s_{i-1, n_i, m_i}[-\tau_i] = s_{i-1, m_i, n_i}^*[\tau_i]$ are moved to lag zero by the similarity transform $\mathbf{S}_{i-\frac{1}{2}}(z) = \Delta_i(z)\mathbf{S}_{i-1}(z)\Delta_i^P(z)$.

The energy of the terms $s_{i-1, m_i, n_i}[\tau_i]$ and $s_{i-1, n_i, m_i}[-\tau_i]$ is then transferred onto the diagonal via a Givens rotation \mathbf{V}_i ,

$$\mathbf{V}_i = \begin{bmatrix} \mathbf{I}_1 & & & & \\ & c_i & & e^{j\gamma_i} s_i & \\ & & \mathbf{I}_2 & & \\ & -e^{-j\gamma_i} s_i & & c_i & \\ & & & & \mathbf{I}_3 \end{bmatrix}, \quad (5)$$

with elements $c_i = \cos(\theta_i)$ and $s_i = \sin(\theta_i)$ only at the intersection of the m_i th and n_i th rows and columns, and \mathbf{I}_i , $i = 1, 2, 3$, identity matrices of appropriate dimensions. The values of θ_i and γ_i depend on $\mathbf{S}_{i-\frac{1}{2}}[0]$, with

$\mathbf{S}_{i-\frac{1}{2}}[\tau] \circ \bullet \mathbf{S}_{i-\frac{1}{2}}(z)$ [1], [5]. When completing the i th SBR2 iteration step, therefore

$$\mathbf{S}_i(z) = \mathbf{V}_i \mathbf{S}_{i-\frac{1}{2}}(z) \mathbf{V}_i^H. \quad (6)$$

If after I iterations, the off-diagonal components are suppressed below some given threshold or a maximum number of iterations has been reached, w.r.t. (1) we obtain $\Gamma(z) = \mathbf{S}_I(z)$ and

$$\mathbf{U}(z) = \mathbf{V}_I \Delta_I(z) \cdots \mathbf{V}_2 \Delta_2(z) \cdot \mathbf{V}_1 \Delta_1(z). \quad (7)$$

Note that in every iteration step, while energy at lag zero is transferred into the diagonal, Givens rotations are also applied at lags $\tau \neq 0$, where some energy may leak back into the off-diagonal components. Nonetheless, SRB2 has been proven to converge [1], [5], [27]. Since the Givens operation is the central algebraic operation in SBR2, in the following we will investigate a fast version derived for the real-valued case in [24].

III. FAST GIVENS ROTATION

A. Givens Rotation

Within SBR2, the Givens rotation is applied in (6), which alternatively in the time domain requires

$$\mathbf{S}_i[\tau] = \mathbf{V}_i^H \mathbf{S}_{i-\frac{1}{2}}[\tau] \mathbf{V}_i. \quad (8)$$

If the support of $\mathbf{S}_{i-\frac{1}{2}}[\tau]$ is restricted to $|\tau_i| \leq T_i$, such that $\mathbf{S}_{i-\frac{1}{2}}[\tau] = \mathbf{0} \quad \forall |\tau| > T_i$, then we require to perform $2T_i + 1$ such similarity transforms. Due to the parahermitian nature of $\mathbf{S}_{i-\frac{1}{2}}(z)$, T of these transforms will be redundant. With the exception of the lag zero matrix $\mathbf{S}_{i-\frac{1}{2}}[0]$, the remaining coefficients $\mathbf{S}_{i-\frac{1}{2}}[\tau]$ will generally not be Hermitian matrices for $\tau \neq 0$. However, it is only the Hermitian zero-lag matrix $\mathbf{S}_{i-\frac{1}{2}}[0]$ where a pair of off-diagonal components must be eliminated.

If initially we focus on a single operation out off the above $T + 1$ similarity transforms for (8) and drop all subscripts, we typically face the problem $\mathbf{S}' = \mathbf{V}^H \mathbf{S} \mathbf{V}$. The similarity transform based on a Givens rotation will modify the m th and the n th rows and columns of \mathbf{S} to obtain \mathbf{S}' , and therefore require approximately $4M$ complex-valued multiply accumulate (MAC) operations, which translates to $16M$ real-valued MACs.

B. Fast Approach

For a faster implementation of the Givens rotation, termed a fast Givens rotation (FGR), [24] contains two interesting ideas. The first is a reduction of computations for a single Givens rotation; a second step, which will be elaborated in Sec. III-C, exploits subsequent savings if several Givens rotations are iterated. The basic idea is to factorise the orthogonal Givens operation into two non-orthogonal ones, where as many elements as possible are set to unity, hence not requiring explicit multiplications.

We first consider a Givens rotation that is applied to a square matrix \mathbf{B}_0 such that $\mathbf{B}_1 = \mathbf{V}_1^H \mathbf{B}_0 \mathbf{V}_1$, with \mathbf{V}_1 a

Givens rotation matrix as in (5). Then, according to [24] it is possible to find diagonal matrices \mathbf{D}_0 and \mathbf{D}_1 , such that $\mathbf{B}_0 = \mathbf{D}_0 \mathbf{A}_0 \mathbf{D}_0$, and $\mathbf{B}_1 = \mathbf{D}_1 \mathbf{F}_1 \mathbf{A}_0 \mathbf{F}_1^H \mathbf{D}_1$. One particular and numerically stable choice according to [24] is such that \mathbf{D}_1 and \mathbf{D}_0 only differ in the m th and n th elements, leading to

$$\mathbf{F}_1 = \begin{bmatrix} \mathbf{I}_1 & & & & \\ & 1 & & f_{1,1} & \\ & & \mathbf{I}_2 & & \\ f_{1,2} & & & 1 & \\ & & & & \mathbf{I}_3 \end{bmatrix}. \quad (9)$$

The angles ϑ and γ that determine \mathbf{V}_1 (i.e. (5)) would be calculated from $\mathbf{A}_0 = \mathbf{D}_0^{-1} \mathbf{B}_0 \mathbf{D}_0^{-1}$ in the standard Givens fashion, whereby [24] advocates the selection

$$p = \frac{|d_m^2 a_{mm}^2 - d_n^2 a_{nn}^2|}{\sqrt{(d_m^2 a_{mm}^2 - d_n^2 a_{nn}^2)^2 - 4d_m^2 d_n^2 |a_{mn}|^2}}, \quad (10)$$

with $0 \leq p \leq 1$, a_{mn} the element in the m th row and n th column of \mathbf{A}_0 , and d_i the i th diagonal element of \mathbf{D}_0 , such that e.g.

$$\cos \vartheta = \sqrt{\frac{1+p}{2}}. \quad (11)$$

Overall, we may select $\mathbf{D}_0 = \mathbf{I}$. Then for \mathbf{F}_1 to take the form in (9), we require that \mathbf{D}_1 matches \mathbf{D}_0 , but that the m th and n th elements are modified to $d_m \cos \vartheta$ and $d_n \cos \vartheta$, respectively. For the complex-valued case, it can be shown that $\gamma = \angle a_{mn}$ [1]. Further, we have

$$f_{1,1} = e^{j\gamma} \frac{d_n \sin \vartheta}{d_m \cos \vartheta}, \quad f_{1,2} = -e^{-j\gamma} \frac{d_m \sin \vartheta}{d_n \cos \vartheta} \quad (12)$$

for the simplified matrix \mathbf{F}_1 .

The application of \mathbf{F}_1 requires only half the number of MACs compared to \mathbf{V}_1 , but the above approach involves some overheads since $\mathbf{V}_1 = \mathbf{D}_1 \mathbf{F}_1 \mathbf{D}_0^{-1}$, even if $\mathbf{D}_0 = \mathbf{I}$. Nonetheless, computational saving arise [24], and the technique becomes even more powerful in case several Givens rotations need to be executed successively, such as part of an EVD or QR decomposition [28].

C. Successive Application of Fast Givens Rotations

If a number of Givens rotations \mathbf{V}_k , $k = 1, \dots, K$, need to be executed successively, then note from above that [24]

$$\mathbf{V}_K \dots \mathbf{V}_2 \mathbf{V}_1 = \mathbf{D}_K \mathbf{F}_K \dots \mathbf{F}_2 \mathbf{F}_1 \mathbf{D}_0^{-1}. \quad (13)$$

Particularly with $\mathbf{D}_0 = \mathbf{I}$, the evaluation is simple, and the only overhead is to track the modifications of \mathbf{D}_k , $k = 1, \dots, K$ which requires a multiplication of the m th and n th element, as indicated above.

IV. FAST GIVENS ROTATION-BASED SBR2

A. Modified Fast SBR2 Algorithm

Even though the SBR2 algorithm does not comprise of a simple consecutive application of Givens rotations, we can apply the idea of (13). This is due to the fact that the interspersed

Algorithm 1: Fast Givens Rotation-Based SBR2

- 1: inputs: $\mathbf{R}(z)$, δ_{\max} , I_{\max}
- 2: initialise: $\mathbf{S}'_0(z) = \mathbf{R}(z)$; $\mathbf{D}_0 = \mathbf{I}$; $\mathbf{U}'_0 = \mathbf{I}$; $i = 0$;
- 3: **repeat**
- 4: $i \leftarrow i + 1$;
- 5: find maximum off-diagonal element δ of $\mathbf{D}_{i-1} \mathbf{S}'_{i-1}(z) \mathbf{D}_{i-1}^H$ via (3);
- 6: determine $\Delta_i(z)$;
- 7: calculate $\mathbf{S}'_{i-\frac{1}{2}}(z) = \Delta_i(z) \mathbf{S}'_{i-1}(z) \Delta_i^P(z)$;
- 8: determine $f_{i,1}$ and $f_{i,2}$ based on $\mathbf{S}'_{i-\frac{1}{2}}[0]$;
- 9: calculate $\mathbf{S}'_i(z) = \mathbf{F}_i \mathbf{S}'_{i-\frac{1}{2}}(z) \mathbf{F}_i^H$;
- 10: calculate $\mathbf{U}'_i(z) = \mathbf{F}_i \Delta_i(z) \mathbf{U}'_{i-1}(z)$;
- 11: update \mathbf{D}_i based on \mathbf{D}_{i-1} and $\mathbf{S}'_{i-\frac{1}{2}}[0]$;
- 12: **until** $(|\delta| < \delta_{\max}) \vee (i \geq I_{\max})$.
- 13: outputs: $\mathbf{U}(z) = \mathbf{D}_i \mathbf{U}'_i(z)$ and $\Gamma(z) = \mathbf{D}_i \mathbf{S}'_i(z) \mathbf{D}_i^H$

delay operations $\mathbf{D}(z)$ in (7) are diagonal matrices, and hence substituting $\mathbf{V}_\ell = \mathbf{D}_\ell \mathbf{F}_\ell \mathbf{D}_{\ell-1}^{-1}$, $\ell = 1, \dots, I$, into (7) leads to

$$\mathbf{U}(z) = \mathbf{D}_I \mathbf{F}_I \mathbf{D}_{I-1}^{-1} \Delta_I(z) \dots \mathbf{D}_1 \mathbf{F}_1 \mathbf{D}_0^{-1} \Delta_1(z) \quad (14)$$

$$= \mathbf{D}_I \mathbf{F}_I \Delta_I(z) \dots \mathbf{F}_2 \Delta_2(z) \mathbf{F}_1 \Delta_1(z) \mathbf{D}_0^{-1}. \quad (15)$$

Therefore, the overall operation now consists of a sequence of delay operations $\Delta_\ell(z)$ and simplified matrix operations \mathbf{F}_ℓ , whereby a multiplication with \mathbf{F}_ℓ only requires approximately half the MACS compared to a matrix multiplication with \mathbf{V}_ℓ . Some book keeping is required to update \mathbf{D}_0 , \mathbf{D}_1 all the way up to \mathbf{D}_I , but these changes only ever affect two components of these diagonal quantities at a time.

This results in the modification of the SBR2 algorithm to its fast Givens rotation-based version, with its algorithmic steps outlined in Algorithm 1. The modifications are reflected in the variables $\mathbf{S}'_i(z)$ and $\mathbf{U}'_i(z)$, which differ from $\mathbf{S}_i(z)$ and $\mathbf{U}_i(z)$ in the standard SBR2 algorithm. The steps 8–11 and 13, to determine, apply and propagate the diagonal correction matrix, apply the reduced-cost matrix \mathbf{F}_ℓ , and to finally correct the output also differ from the standard SBR2 method.

B. Convergence

For the convergence of Algorithm 1, the essential detail is that the maximum search in step 5 within $\mathbf{D}_{i-1} \mathbf{S}'_{i-1}(z) \mathbf{D}_{i-1}^H = \mathbf{S}_{i-1}(z)$ is performed over the same quantity as in the SBR2 algorithm in (3). Therefore the convergence proof of the SBR2 algorithm in [1], [7] holds equally for the fast Givens rotation-based SBR2 version. This also implies that Algorithm 1 is guaranteed to converge to a spectrally majorised solution as defined in (2) [27].

Note that it is not necessary to multiply out $\mathbf{D}_{i-1} \mathbf{S}'_{i-1}(z) \mathbf{D}_{i-1}^H$ in Step 5 of Algorithm 1 explicitly; a maximum modulus search can first be performed over the temporal dimension of $\mathbf{S}'_{i-1}[\tau]$, and the matrix-valued result $\mathbf{A}_{\max, \tau}$ can be weighted by $\mathbf{D}_{i-1} \mathbf{A}_{\max, \tau} \mathbf{D}_{i-1}^H$, which due to the Hermitian nature of $\mathbf{A}_{\max, \tau}$ and the exclusion of diagonal terms only takes $\frac{1}{2}M(M-1)$ MACs. The remainder of the

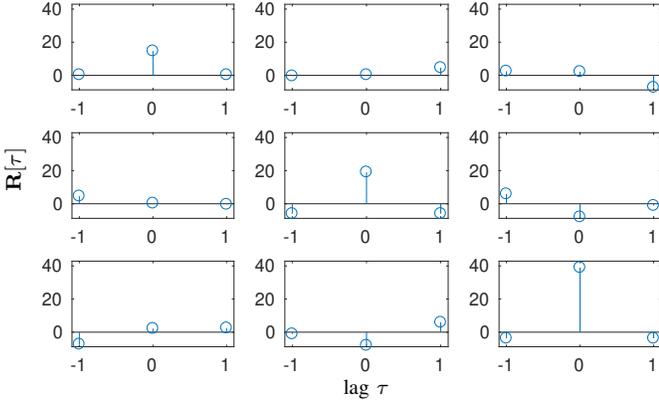


Fig. 1. Plots showing $\mathbf{R}[\tau] \in \mathbb{R}^{3 \times 3}$ used as a numerical example; each subplot represents one polynomial entry of $\mathbf{R}[\tau]$.

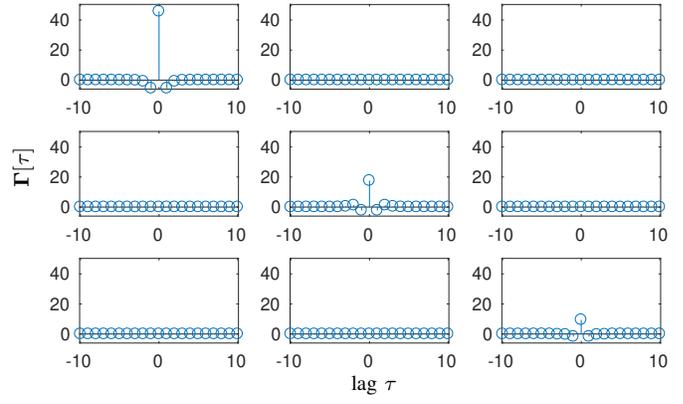


Fig. 2. Approximately diagonalised matrix $\mathbf{\Gamma}[\tau]$, obtained from $\mathbf{R}[\tau]$ in Fig. 1 by the fast Givens rotation-based SBR2 algorithm.

maximum search can then be performed over either the lower left or upper right triangular part of $\mathbf{D}_{i-1} \mathbf{A}_{\max, \tau} \mathbf{D}_{i-1}^H$.

V. SIMULATION AND RESULTS

A. Numerical Example and Convergence

As a numerical example, we utilise the matrix $\mathbf{R}(z) : \mathbb{C} \rightarrow \mathbb{C}^{3 \times 3}$ of order 2, such that $\mathbf{R}(z) = \mathbf{R}[-1]z + \mathbf{R}[-0] + \mathbf{R}[1]z^{-1}$. For the matrix-valued coefficients, we have

$$\mathbf{R}[0] = \begin{bmatrix} 14.7 & 0.3 & 2.2 \\ 0.3 & 19.1 & -8.0 \\ 2.2 & -8.0 & 39.0 \end{bmatrix} \quad (16)$$

$$\mathbf{R}[1] = \begin{bmatrix} 0.3 & 4.6 & -7.3 \\ -0.4 & -6.0 & -1.1 \\ 2.5 & 5.9 & -3.7 \end{bmatrix}. \quad (17)$$

Further note that $\mathbf{R}[-1] = \mathbf{R}^H[1]$. This matrix is also characterised in Fig. 1.

Operating on the above $\mathbf{R}(z)$, the fast Givens rotation-based SBR2 algorithm converges in $I = 238$ iterations, and for a maximum off-diagonal component threshold $\delta_{\max} = 10^{-5}$ yields the eigenvalues shown in Fig. 2. The extracted eigenvalues $\mathbf{\Gamma}[\tau] \circ \bullet \mathbf{\Gamma}(z)$ are Laurent polynomial that decay in both positive and negative lag directions, and Fig. 2 only provides values for the central lags $|\tau| \leq 10$. The spectral majorisation of the eigenvalues as defined in (2), and to which according to [27] and Sec. IV-B the fast Givens rotation-based SBR2 algorithm is guaranteed to converge, is demonstrated in Fig. 3. These power spectral density terms are non-negative real and ordered in power, satisfying (2).

In comparison, for the above $\mathbf{R}(z)$, the standard SBR2 algorithm [1], [7] converges in $I = 235$ iterations. The obtained eigenvalues $\mathbf{\Gamma}_{\text{SBR2}}(z)$ are near-identical to those of the proposed algorithm in Figs. 2 and 3, with an error $\sum_{\tau} \|\mathbf{\Gamma}_{\text{SBR2}}[\tau] - \mathbf{\Gamma}[\tau]\|_F^2$ of -96.9dB , whereby $\|\cdot\|_F$ is the Frobenius norm. The small difference in the number of iterations and in the obtained eigenvalues—which according to [2], [4] are unique—is likely due to the numerical differences between the two algorithm versions, even though the fast Givens approach is claimed to be robust to over-and

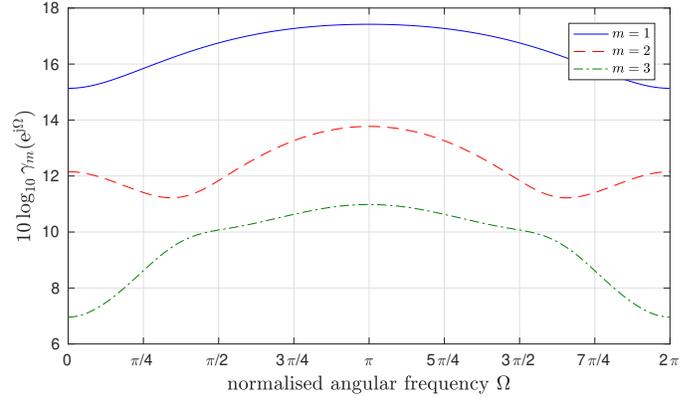


Fig. 3. Eigenvalues $\mathbf{\Gamma}(z)$ as obtained by the fast Givens rotation-based SBR2 algorithm evaluated on the unit circle, $z = e^{j\Omega}$.

underflow [24]. The extracted eigenvalues differ by a somewhat larger amount due to their ambiguity w.r.t. multiplications by arbitrary allpass functions [2], [16], and are therefore not shown here.

B. Computational Complexity

In a standard Matlab implementation, matrix-valued operations are favoured and often there is no relation between the sparseness of a matrix and the execution time for its multiplication. Therefore, the run time difference (averaged over 500 runs) for the example in Sec. V-A is only 98.1 ms for the standard SBR2 and 92.6 ms for the proposed fast Givens rotation-based SBR2 version. To investigate the potential of the fast Givens approach for speeding up an implementation in a non-Matlab environment, the fast and standard Givens rotation operations have been explicitly implemented in C and operated from within Matlab through pre-compiled MEX files.

Using Matlab’s profiler, the MEX-routines are specifically called to perform $I = 150$ iterations on a matrix $\mathbf{R}(z)$ calculated from a ground truth with a diagonal $\mathbf{\Gamma}(z)$ of order 100 and an arbitrary paraunitary matrix $\mathbf{U}(z)$ determined from randomised elementary paraunitary operations [29] of order 50. The execution time averaged over 5000 runs for different

TABLE I
EXECUTION TIME COMPARISON BETWEEN STANDARD AND FAST GIVENS
ROTATION-BASED SBR2 IMPLEMENTATIONS, SHOWING MEAN
PLUS/MINUS ONE STANDARD DEVIATION.

method	computation time / [ms]			
	$M = 3$	$M = 5$	$M = 10$	$M = 20$
standard	1.03 ± 0.02	2.67 ± 0.04	5.18 ± 0.29	14.56 ± 0.22
FGR	0.99 ± 0.03	2.35 ± 0.04	4.66 ± 0.26	13.03 ± 0.26

spatial dimensions M is summarised in Tab. I. The execution includes memory allocation, maximum searches and various book keeping, and therefore is not as dramatic a reduction as direct comparison in MAC operations might suggest. Nevertheless, as M increases, a substantial gap between the run times of the standard and proposed SBR2 implementations emerges, with the computation time shrinking in excess of 10% for the largest matrix size of $M = 20$ that is employed here.

VI. CONCLUSION

This paper has exploited a fast Givens rotation trick to reduce the number of multiply-accumulate operations, particularly when operating on a sequence of Givens rotations. This trick has been adopted for a polynomial matrix eigenvalue decomposition technique known as the second order sequential best rotation algorithm, where the interspersing by delay elements can be absorbed. The modified algorithm will minimise the same cost function as the standard SBR2 algorithm, which in every iteration step will eliminate the maximum off-diagonal component. We have shown that due to the maintenance of the cost function and maximum search, the proposed algorithms inherits the convergence proof and properties of the standard SBR2 algorithm to a diagonalised and spectrally majorised solution for the polynomial eigenvalues. Particularly as the spatial dimension—i.e. the number of sensors recording the data—increases, the computational savings can become significant.

REFERENCES

- [1] J. G. McWhirter, P. D. Baxter, T. Cooper, S. Redif, and J. Foster, "An EVD Algorithm for Para-Hermitian Polynomial Matrices," *IEEE Trans SP*, **55**(5):2158–2169, May 2007.
- [2] S. Weiss, J. Pestana, and I. K. Proudler, "On the existence and uniqueness of the eigenvalue decomposition of a parahermitian matrix," *IEEE Trans SP*, **66**(10):2659–2672, May 2018.
- [3] S. Weiss, I. K. Proudler, and F. K. Coutts, "Eigenvalue decomposition of a parahermitian matrix: Extraction of analytic eigenvalues," *IEEE Trans SP*, **69**:722–737, Jan. 2021.
- [4] S. Weiss, J. Pestana, I. Proudler, and F. Coutts, "Corrections to "on the existence and uniqueness of the eigenvalue decomposition of a parahermitian matrix"," *IEEE Trans SP*, **66**(23):6325–6327, Dec. 2018.
- [5] S. Redif, S. Weiss, and J. McWhirter, "Sequential matrix diagonalization algorithms for polynomial EVD of parahermitian matrices," *IEEE Trans SP*, **63**(1):81–89, Jan. 2015.
- [6] J. Corr, K. Thompson, S. Weiss, J. McWhirter, S. Redif, and I. Proudler, "Multiple shift maximum element sequential matrix diagonalisation for parahermitian matrices," in *IEEE SSP*, Gold Coast, Australia, pp. 312–315, June 2014.
- [7] S. Redif, J. McWhirter, and S. Weiss, "Design of FIR paraunitary filter banks for subband coding using a polynomial eigenvalue decomposition," *IEEE Trans SP*, **59**(11):5253–5264, Nov. 2011.
- [8] S. Weiss, S. Redif, T. Cooper, C. Liu, P. Baxter, and J. McWhirter, "Paraunitary oversampled filter bank design for channel coding," *EURASIP J. Advances Signal Processing*, vol. 2006, pp. 1–10, 2006.
- [9] N. Moret, A. Tonello, and S. Weiss, "Mimo precoding for filter bank modulation systems based on PSVD," in *IEEE 73rd Vehicular Technology Conference*, May 2011.
- [10] S. Weiss, S. Bendoukha, A. Alzin, F. Coutts, I. Proudler, and J. Chambers, "MVDR broadband beamforming using polynomial matrix techniques," in *EUSIPCO*, Nice, France, pp. 839–843, Sep. 2015.
- [11] S. Redif, S. Weiss, and J. McWhirter, "Relevance of polynomial matrix decompositions to broadband blind signal separation," *Signal Processing*, **134**:76–86, May 2017.
- [12] S. Weiss, C. Delaosa, J. Matthews, I. Proudler, and B. Jackson, "Detection of weak transient signals using a broadband subspace approach," in *Int. Conf. Sensor Signal Processing for Defence*, Edinburgh, UK, Sept. 2021.
- [13] V. W. Neo, C. Evers, and P. A. Naylor, "Speech enhancement using polynomial eigenvalue decomposition," in *2019 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz, NY, pp. 125–129, Oct. 2019.
- [14] A. Hogg, V. Neo, S. Weiss, C. Evers, and P. Naylor, "A polynomial eigenvalue decomposition music approach for broadband sound source localization," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, New Paltz, NY, Oct. 2021.
- [15] J. Corr, K. Thompson, S. Weiss, J. McWhirter, and I. Proudler, "Cyclic-by-row approximation of iterative polynomial EVD algorithms," in *Sensor Signal Processing for Defence*, Edinburgh, Scotland, pp. 1–5, Sep. 2014.
- [16] J. Corr, K. Thompson, S. Weiss, I. Proudler, and J. McWhirter, "Row-shift corrected truncation of paraunitary matrices for PEVD algorithms," in *EUSIPCO*, Nice, France, pp. 849–853, Sep. 2015.
- [17] —, "Reduced search space multiple shift maximum element sequential matrix diagonalisation algorithm," in *IET/EURASIP Intelligent Signal Processing*, London, UK, Dec. 2015.
- [18] F. Coutts, J. Corr, K. Thompson, I. Proudler, and S. Weiss, "Divide-and-conquer sequential matrix diagonalisation for parahermitian matrices," in *Sensor Signal Processing for Defence Conference*, London, UK, pp. 1–5, Dec. 2017.
- [19] F. K. Coutts, I. K. Proudler, and S. Weiss, "Efficient implementation of iterative polynomial matrix evd algorithms exploiting structural redundancy and parallelisation," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 66, no. 12, pp. 4753–4766, Dec. 2019.
- [20] S. Kasap and S. Redif, "Novel field-programmable gate array architecture for computing the eigenvalue decomposition of para-hermitian polynomial matrices," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 22, no. 3, pp. 522–536, Mar. 2014.
- [21] R. W. Stewart, R. Chapman, and T. S. Durrani, "The Square Root In Signal Processing," in *Real-Time Signal Processing XII*, J. P. Letellier, Ed., vol. 1154, International Society for Optics and Photonics. SPIE, pp. 89 – 101, 1989.
- [22] J. Götze and U. Schwiigelshohn, "A square root and division free givens rotation for solving least squares problems on systolic arrays," *SIAM Journal on Scientific and Statistical Computing*, vol. 12, no. 4, pp. 800–807, 1991.
- [23] M. Moonen and I. Proudler, "Generating 'fast qr' algorithms using signal flow graph techniques," in *Conference Record of The Thirtieth Asilomar Conference on Signals, Systems and Computers*, vol. 1, pp. 410–414, Nov. 1996.
- [24] W. Rath, "Fast Givens rotations for orthogonal similarity transformations," *Numerische Mathematik*, vol. 40, no. 1, pp. 46–56, 1982.
- [25] P. Vaidyanathan, "Theory of optimal orthonormal subband coders," *IEEE Transactions on Signal Processing*, vol. 46, no. 6, pp. 1528–1543, Jun. 1998.
- [26] Z. Wang, J. G. McWhirter, J. Corr, and S. Weiss, "Multiple shift second order sequential best rotation algorithm for polynomial matrix EVD," in *23rd European Signal Processing Conference*, , pp. 844–848, Nice, France, Sep. 2015.
- [27] J. G. McWhirter and Z. Wang, "A novel insight to the SBR2 algorithm for diagonalising para-hermitian matrices," in *11th IMA Conference on Mathematics in Signal Processing*, Birmingham, UK, Dec. 2016.
- [28] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, Maryland: John Hopkins University Press, 1996.
- [29] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs: Prentice Hall, 1993.

Target Detection and Recognition of Ground Penetrating Radar using Morphological Image Analysis and Graph Laplacian Regularisation

Jun Dong¹, Vladimir Stankovic¹, Nigel Davidson²

¹Department of Electronic and Electrical, University of Strathclyde, Glasgow

²The Defence Science and Technology Laboratory, Kent
{jun.dong, vladimir.stankovic}@strath.ac.uk

Abstract—Ground Penetrating Radar (GPR) is often used for detecting non-intrusively buried targets, in road engineering, manufacturing, and in military fields. Based on transmitting high frequency electromagnetic waves, GPR generates 3D data of the underground structure enabling accurate and fast target detection. However, after inverse Fourier Transform, the 3D GPR images are often out-of-focus and contain high measurement noise. This calls for advanced signal and image processing methods to improve signal-to-noise ratio, isolate the most discriminative features, and perform target detection and localisation. Using a vehicle-mounted GPR array data provided in the 2020 UDRC GPR data challenge, we show that morphological image analysis and semi-supervised learning via graph Laplacian regularisation can detect different types of targets buried at various depths with very low false alarm rate.

I. INTRODUCTION

Ground Penetrating Radar (GPR) uses high frequency radio waves for detection of the structure of underground objects, based on the difference in electrical properties between the target object and surrounding medium. GPR has been widely used for detection of metallic and non-metallic objects [1] in road engineering, manufacturing, archaeology, and has also become popular for detection of buried targets in military fields, such as land mines and Improvised Explosive Devices (IEDs) [2], [3]. Indeed, the GPR array [2], [3] has shown significant advantages over competing technologies as a non-destructive, remote sensing technique that provides high resolution 3-D data, which helps to make the detection and recognition of targets accurate and fast.

GPR is a non-invasive geophysical technique used for detecting objects underground or analysing the structure in visually opaque materials, based on measuring propagation of emitted high-frequency electromagnetic waves. Though the research and application of GPRs have long history [4], [5], some of the major challenges of using GPR data to detect underground objects are still present. This is due to low-resolution features and out-of-focus GPR images, high measurement noise, and interference, leading to very low signal to noise ratio (SNR). This calls for advanced signal processing techniques to improve SNR and construct discriminative features.

Since a typical output of a GPR system is a spatio-temporal GPR image, various signal and image processing

methods can be used for extracting useful information. Traditional popular approaches are based detection of hyperbolas in GPR images, e.g., using Hough transform [6], template matching [7], or Viola-Jones learning algorithm with Haar-like features [8]. However, these approach suffer from false alarms or are complex and require a high number of parameters [2]. In this paper, we propose a low-complexity system, by transforming the time slices of the original dataset acquired by a GPR system into a binary image by performing morphological image analysis to construct distinct features of potential targets, and then classify the constructed features into target/non-target classes using semi-supervised learning based on label propagation over graphs. Morphological image processing [9] is a collection of non-linear operations used to analyse geometrical structures. Morphological operations rely only on the relative ordering of pixel values, and not on their numerical values, and therefore are especially suited to processing binary images.

Classification via Graph Laplacian Regularisation (GLR) has been widely used to classify image and time-series signals, especially when the number of labelled signals that can be used for training is small [10], [11]. In this paper, we use normalised GLR (identified in [12] as the best performing semi-supervised classification method for seismic signals among several tested methods based on regularisation-on-graphs), to identify the time slices which contain information about the real targets in order to reduce the false alarm rate and locate the real targets.

In particular, this paper presents a novel post-processing detection method for a stepped frequency continuous wave (SFCW) GPR system. After signal pre-processing, normalisation, and inverse Fourier Transform, the detection of potential targets is performed by carrying out morphological analysis on each time slice. Afterwards, target recognition is performed via GLR-based semi-supervised learning on all potential targets.

The advantage of the proposed method over recent alternatives, such as [2] and [3], is its semi-supervised nature that requires a very small labelled dataset for training (namely, 3.5% in our simulations), while deep-learning based approaches [2], [3] report results with about two thirds or more of available data used for training. Furthermore, the proposed approach automatically learns system parameters from the data and thus, in contrast to prior work [7], [6], does

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant number EP/S000631/1 and the MOD University Defence Research Collaboration in Signal Processing.

not have many parameters to tune the model. Our simulation results show that the proposed method has a very low false alarm rate. Its low reliance on labelled data, low complexity, and the fact that it can be implemented in real time, make the algorithm a practical solution for future GPR array target detection systems.

This paper is organized as follows: In Section II we describe pre-processing of the GPR data and the method of generating features. In Section III we provide brief background on the morphological image processing and graph signal processing, present the proposed detection and recognition algorithm, including data normalisation and morphology detection method for the constructed features. Section IV presents the simulation results, and Section V contains our conclusions.

II. DATASET USED AND SIGNAL PRE-PROCESSING

In this work, we use the data captured by a vehicle-mounted SFCW GPR array, which contains 41 transmit-receive channels. The GPR channels were evenly spaced 7.5 cm apart, covering 3m width in total. The GPR system was mounted on a vehicle, driven along test lanes collecting data at 5 cm intervals. Therefore, a 3D dataset was collected as the GPR array was driven along the lanes. The data was made available as part of a UDRC challenge on automatic target detection for 3D GPR data¹.

The test lanes included a variety of targets buried at different depths. Multiple test fields were used, and in this paper we focus on a test field that was 100 meters long and shares the same width with the radar antenna array, facilitating for the vehicle to travel through the whole test field and collect a comprehensive dataset. The test field contains 54 targets, randomly located across the whole field.

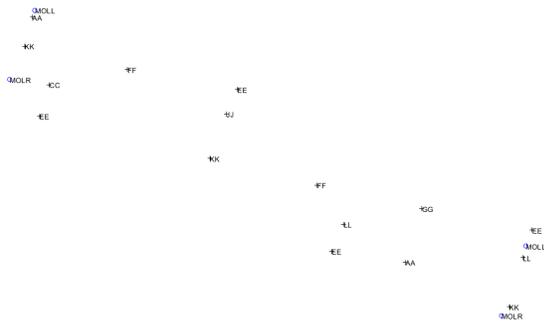


Fig. 1. A part of the plan view of the test lane.

The plan view of the test lane, which displayed the layout for targets and markers distribution in the test lane, is shown in Fig. 1. The test lane is a rectangular field with 10 landmarks marking its boundary. Blue circles represent the positions of the landmarks of the test lane. The vehicle with the mounted radar was moving within the marked field from one side to another. The cross marks represent

¹<https://udrc.eng.ed.ac.uk/sites/udrc.eng.ed.ac.uk/files/attachments/Automatic%20Target%20Detection%20for%203D%20Ground-Penetrating%20Radar%20Data%20Challenge.pdf>.

the target locations, which are considered as ground-truth target positions. The labels such as ‘FF’ and ‘EE’ near the corresponding cross marks indicate the target types; all targets of the same type were buried at a common depth.

As described above, the GPR array collected the data along the test lane at the 5 cm interval, thus, the data in the original format contains 2000 time slices. To facilitate analysis, we first reshaped the dataset, into a $34 \times 41 \times 2000$ matrix, so that each column represents one radar Transmitter/receiver channel (41 in total). Since the data was collected in the frequency domain due to the characteristic of the SFCW system, to perform analysis in the spatial domain, an inverse fast Fourier transform (IFFT) was carried out along the columns of the dataset, i.e., along the data returned by every array channel, to retain the channel-specific information. Then we process data time slice by time slice, i.e., frame by frame, where each frame is of size 34×41 pixels.

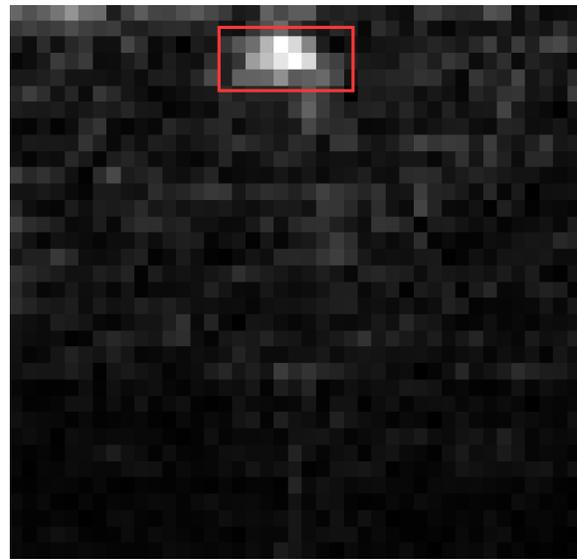


Fig. 2. The magnitude data of the 254th time slice (i.e., frame) of the dataset.

A typical time slice, after IFFT is shown in Fig. 2. Each column corresponds to one channel and the vertical axis shows the depth derived from the propagation time, increasing from the top down. The figure shows a section view of the test lane where the vertical direction can be regarded as the depth under the surface. Significantly high values can be observed in the red rectangular box indicating the presence of a target. However, in practice, the targets are often not visible due to measurement noise and clutter. It can also be seen that the targets are pixelated due to the out-of-focus problem common to GPR images [4]. Thus, feature construction methods that are robust to noise and blur are required to construct discriminative features necessary for accurate target detection.

III. METHODOLOGY

This section describes the proposed methodology. In particular, given a 2-D GPR image (e.g., Fig. 2), we use morphological image analysis to construct features and Graph

Laplacian Regularisation (GLR)-based label propagation, as advanced semi-supervised classifier. This has been shown to perform well when the number of labelled samples is low and dataset is noisy [11], [12].

All matrices will be denoted by capital bold letters, such as \mathbf{X} , whose entry in the i -th row and j -th column will be $x_{i,j}$. Vectors will be denoted by lower-case bold letters, e.g., \mathbf{x} , with the i -th element being x_i . Sets are denoted by calligraphic font, e.g., \mathcal{S} , and $|\mathcal{S}|$ denotes cardinality of \mathcal{S} .

A. Morphological Detection

As shown in the example in Fig. 2, the targets are characterised by high pixel intensity and the shape. Morphological image analysis is often used to identify the shape or morphology of an object in the image. Thus, in order to extract the shape of the objects, we introduce a morphological probe [15]. This method requires defining a small shape or template called structuring element, which is slid along the image and compared with the corresponding neighbourhood of pixels [15], [9].

Morphological operations rely only on the relative ordering of the pixel values, and not on their numerical values [15]. Thus, we need to generate a binary image from the original GPR image time slice to extract the features and increase the SNR, which is done using the following equation:

$$\mathbf{P} = \mathbf{G} - w \cdot \overline{\mathbf{G}}, \quad (1)$$

where, \mathbf{G} is the original 2D time slice data matrix (i.e., the GPR image whose example is shown in Fig. 2), $\overline{\mathbf{G}}$ is its mean, w is a scaling parameter set to 6 in our simulations, and \mathbf{P} is the resulting normalised matrix. Let $p_{i,j}$ be the element (i.e., pixel) of the normalized matrix \mathbf{P} in the i -th row and j -th column. Then, we feed $p_{i,j}$ into Heaviside step function, that is:

$$b_{i,j} = H(p_{i,j}), \quad (2)$$

where $b_{i,j}$ is the element of the binary matrix, i.e., the morphological detection result, \mathbf{B} in the i -th row and j -th column.

Based on the target shapes in the training datasets, we experimented with different structuring element shapes and finally adopted a “line” structure element, i.e., 3 pixels in a line as a template. The detection is performed by sliding this 3-pixel line template along the image. The detection result will return a positive detection only if the structuring element “fits” the target, which means if all the 3 pixels in the structuring element calculated by Eq. (2) are larger than 0. This process would generate a binary image \mathbf{B} , as shown in Fig. 3 for the example GPR image from Fig. 2. In \mathbf{B} , the pixels with value 1 are identified as potential targets. In the meantime, since the relative location of the time slice and the radar channel is known, the coordinate of potential targets can be calculated. However, as it will be shown in the next section, our simulation results show that this detection result is not reliable enough, as the false alarm rate is very high, that is, due to high measurement noise many false targets

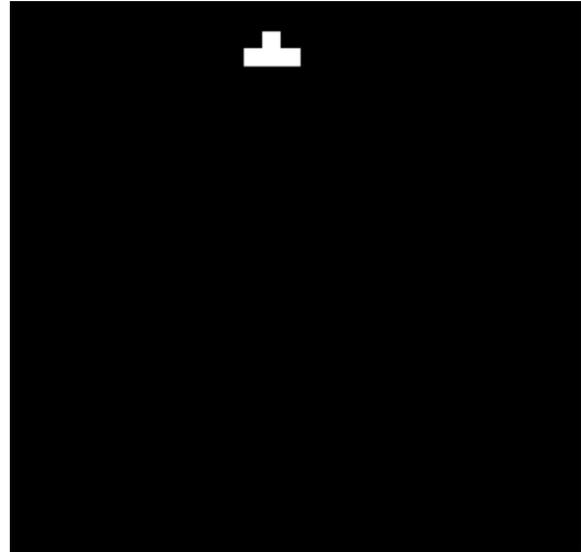


Fig. 3. Morphological detection result for a time slice.

are returned. Therefore, a classifier based on GLR is used to increase the accuracy as described next.

B. GLR-based classification

The potential targets isolated using morphological image analysis as described in the previous subsection, are fed into a GLR-based classifier.

Semi-supervised learning via label propagation on graphs is widely used for time-series signal classification when the number of labelled signals available for training is very small and hence insufficient to develop good statistical models of the data [10], [11]. The main idea is to represent the classification labels as a graph signal [16] where highly correlated samples are connected with high-weight edges, and then restoring the unknown graph signal samples (corresponding to the test data) using label propagation [10], e.g., via GLR.

Specifically, first, the features extracted from the original time slice data using morphological analysis are embedded into an undirected, connected graph, $\mathcal{G} = (\mathcal{V}, \mathbf{A})$, where \mathcal{V} is the set of vertices and \mathbf{A} is the graph adjacency matrix [16]. Each vertex in \mathcal{G} , $\nu \in \mathcal{V}$ corresponds to one column in the GPR image and is characterised by the corresponding feature vector, e.g., intensity of the corresponding column in the GPR binary image after morphological image processing.

The graph needs to capture well the correlation between the constructed features. Let $f_k(i)$ represent the k -th feature assigned to Node ν_i , that is, the intensity of k -th pixel in Column i in the processed GPR image. Then, we set the (i, j) entry in \mathbf{A} , $a_{i,j}$, i.e., the weight of the edge between vertices i and j , as:

$$a_{i,j} = \exp \left\{ - \sum_{k=1}^K \frac{(f_k(i) - f_k(j))^2}{2\sigma_k^2} \right\}, \quad (3)$$

where σ_k represents the graph kernel bandwidth associated to the k -th feature, and K is the feature vector dimension, which is the length of the GPR image \mathbf{B} column (in pixels).

We assign to each vertex, ν_i a discrete graph signal s_i that carries the class label of the corresponding event. That is, for Vertex ν_i

$$s_i = \begin{cases} +1, & \text{if } \nu_i \text{ belongs to Target Class and } i \leq n \\ -1, & \text{if } \nu_i \text{ does not belong to Target Class and } i \leq n \\ 0, & \text{for } n + 1 < i \leq N, \end{cases} \quad (4)$$

where n and N , respectively, represent the number of training samples (i.e., labelled columns) and the total number of image columns in the dataset.

If graph \mathcal{G} captures well the correlation between the constructed features, then the vertices with the same label will be connected by high-weight edges, that is, the graph signal \mathbf{s} will be smooth with respect to \mathcal{G} and we can apply “label propagation” [10], for example, via GLR or normalised GLR, to extrapolate the missing labels (that are initialised to zero in (4)). GLR extrapolates the missing labels by finding the smoothest graph signal under constraints given in the first two lines of Eq. (4) [16].

The combinatorial graph Laplacian matrix is given by $\mathcal{L} = \mathbf{D} - \mathbf{A}$, where \mathbf{D} is the diagonal matrix, given by $d_{i,i} = \sum_j a_{i,j}$, and its normalised form is $\mathbf{L} = \mathbf{D}^{(-1/2)} \mathcal{L} \mathbf{D}^{(-1/2)}$. To perform GLR we define the objective function as:

$$\begin{aligned} \tilde{\mathbf{s}} &= \arg \min_{\mathbf{s} \in \mathbb{R}^N} (\mathbf{s}^T \mathbf{L} \mathbf{s}) \\ \text{s.t. } \tilde{\mathbf{s}}_{1:n} &= \mathbf{s}_{1:n}, \end{aligned} \quad (5)$$

which has the close-form solution given by [17]:

$$\tilde{\mathbf{s}}_{n+1:N} = \mathbf{L}_{n+1:N, n+1:N}^\dagger (-\mathbf{s}_{1:n}^T \mathbf{L}_{1:n, n+1:N})^T, \quad (6)$$

where \dagger denotes matrix pseudo-inverse.

IV. RESULT AND EVALUATION

In this section we present and analyse our simulation results. In all our simulations, we use the dataset described in Section II. After detection via morphological image analysis, 1162 potential targets are identified. 40 column vectors with potential targets are used for training the semi-supervised classifier, and the remaining column vectors are used to form the test set.

We evaluate the performance of the proposed system using the probability of detection and probability of false alarm as performance measures. These criteria are widely used in the similar detection problems [18]. We also used the Receiver Operating Characteristic (ROC) curves to evaluate the classifier performance.

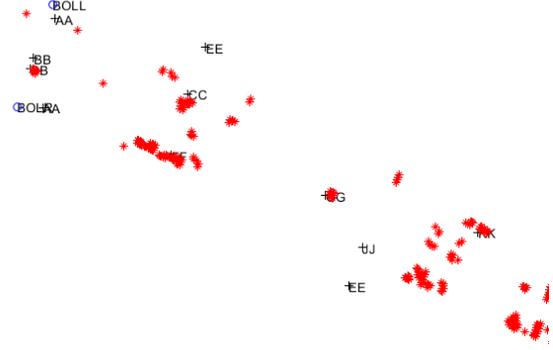


Fig. 4. Detection result before classification.

As part of our study, first we test the performance after morphological analysis only (without semi-supervised learning). The plan view of the test lane showing morphological detection result is shown in Fig. 4. The red asterisks represent the predicted positions and the cross marks represent the actual target locations, i.e., the ground-truth. One can see many correct “hits” (where the red asterisk and the cross marks overlap), but also many false positives (the red asterisk is far from the target real position), and missed targets.

Aiming to reduce the false positive rate, we use semi-supervised graph classifier to distinguish the true targets. We use the adjacency matrix as given in Eq. (3) and learn the optimal σ_k 's from the training data as in [11]. That is, using the training labels, we find the optimal σ_k 's that maximise the difference between the weights of the edges that connect nodes that belong to the same class and the weights of the edges connecting nodes that belong to the opposite class.

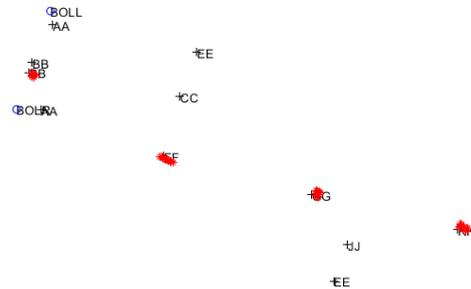


Fig. 5. Detection result after classification.

The plan view of the test lane showing morphological detection result after classification is shown in Fig. 5. Compared to Fig. 4, one can see that most of the false positives are removed while the correct “hits” are kept. High false negative rates are mainly due to very low SNRs in the original dataset, resulting in the absence of prominent features; therefore those targets were unable to be detected

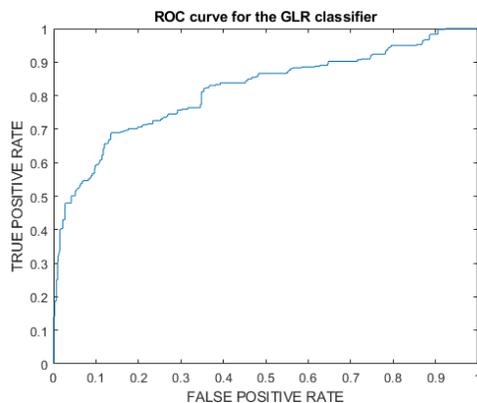


Fig. 6. ROC curve for the proposed GLR classifier.

by the morphological analysis. Since the semi-supervised classifier takes morphological analysis results as input, it cannot reduce the false negatives. Note that, these targets are mainly shallow targets, most likely due to low resolution (the out-of-focus problem), high level of interference, and echo of radio waves when they hit the surface [4].

TABLE I
RATING CRITERIA FOR THE TEST .

	After classification	Before classification
Target detection reports	144	1162
False reports	20	965
Probability of detection	86.1%	17%
Probability of false alarm	13.9%	83%

Table I shows the quantitative performance of the overall proposed system. The results are reported based on the fact that if the predicted position falls within a 0.3m radius circle whose center is the real target position, then the predicted result would be counted as true, otherwise it would be counted as false. In Table I, the target detection reports represent how many true positive results have been returned by the classifier. The false report represents the number of positive predictions that are in fact false (false positives). Probability of detection and probability of false alarm show the proportion of true reports and false reports, respectively.

Fig. 6 shows the ROC curve of the GLR classifier for the true target class, which gives the true positive rate against the false positive rate. The larger the area under the ROC curve the better the performance of the classifier.

Semi-supervised classification has significantly reduced the false alarm rate and in turn largely improved the performance. This performance additionally shows the morphological detection and GLR classification method have a satisfactory detection rate and a relatively low false positive rate.

V. CONCLUSION

We presented a novel target detection method for data collected by ground penetrating radar. We showed that by performing morphological feature detection and graph-based semi-supervised classification, we can recognise several types of targets buried in the soil, at various depths with

a very low false positive rate. Future work will focus on further improving the results, especially for targets at lower depths via de-blurring and further optimising the classifier performance.

ACKNOWLEDGEMENT

The authors would like to thank Prof. Ian Proudler for proofreading the initial draft of the paper and providing his comments.

REFERENCES

- [1] Daniels, D. J. "Ground penetrating radar," *Encyclopedia of RF and microwave engineering*. 2004.
- [2] M. Pham and S. Lefèvre, "Buried object detection from B-Scan ground penetrating radar data using Faster-RCNN," *IGARSS 2018 - 2018 IEEE International Geoscience and Remote Sensing Symposium*, 2018, pp. 6804-6807, doi: 10.1109/IGARSS.2018.8517683.
- [3] V. Kafedziski, S. Pecov and D. Tanevski, "Detection and classification of land mines from Ground Penetrating Radar data using Faster RCNN," *2018 26th Telecommunications Forum (TELFOR)*, 2018, pp. 1-4, doi: 10.1109/TELFOR.2018.8612117.
- [4] C. Ozdemir, S. Demirci, E. Yiğit, and B. Yılmaz, "A review on migration methods in B-scan ground penetrating radar imaging," *Mathematical Problems in Engineering*, Hindawi, vol. 2014, 2014.
- [5] A. Benedetto, F. Tosti, L.B. Ciampoli, and F. D'amico, Fabrizio, "An overview of ground-penetrating radar signal processing techniques for road inspections," *Elsevier Signal Processing*, pp. 201–209, vol. 132, 2017.
- [6] C. G. Windsor, L. Capineri, and P. Falorni, "A data pair-labeled generalized Hough transform for radar location of buried objects," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, pp. 124–127, 2014.
- [7] G. Terrasse, J.-M. Nicolas, E. Trouve, and E. Drouet, "Automatic localization of gas pipes from gpr imagery," in *EUSIPCO-2016*, pp. 2395–2399, 2016.
- [8] C. Maas and J. Schmalzl, "Using pattern recognition to automatically localize reflection hyperbolas in data from ground penetrating radar," *Computers & Geosciences*, vol. 58, pp. 116–125, 2013.
- [9] P. Soille, *Morphological Image Analysis: Principles and Applications*, Springer Science & Business Media, 2013.
- [10] M. Belkin and P. Niyogi, "Using manifold structure for partially labeled classification", *NeurIPS*, 2003.
- [11] M. Ye, V. Stankovic, L. Stankovic, and G. Cheung, "Robust deep graph based learning for binary classification," *IEEE Trans. Signal Inf. Process. Networks*, Nov. 2020.
- [12] J. Li, L. Stankovic, S. Pytharouli, and V. Stankovic, "Automated platform for microseismic signal analysis: Denoising, detection and classification in slope stability studies", *IEEE Trans. Remote Sensing*, to appear, 2021.
- [13] G. Cardone, G. Cincotti, P. Gori, and M. Pappalardo, "Optimization of wide-band linear arrays," *IEEE Trans. Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 48, no. 4, pp. 943–952, July 2001.
- [14] H. L. Van Trees, *Detection, Estimation and Modulation Theory: Optimum Array Processing*. New York: Wiley, 2002.
- [15] "Morphological Image Processing", *Cs.auckland.ac.nz*, 2021. [Online]. Available: <https://www.cs.auckland.ac.nz/courses/compsci773s1c/lectures/ImageProcessing-html/topic4.htm>. [Accessed: 30- Apr- 2021].
- [16] A. Ortega, P. Frossard, J. Kovacevic, J. M. F. Moura and P. Vandergheynst, "Graph Signal Processing: Overview, challenges, and applications," in *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, May 2018.
- [17] S. Boyd and P. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [18] X. Shi, D. Cheng, Z. Song and C. Wang, "A real-time method for landmine detection using vehicle array GPR," *2018 17th International Conference on Ground Penetrating Radar (GPR)*, 2018, pp. 1-4, doi: 10.1109/ICGPR.2018.8441584.

Object Detection in EO/IR and SAR Images Using Low-SWAP Hardware

Richard O. Lane
QinetiQ
Malvern, UK

Adam J. Wragge
QinetiQ
Malvern, UK

Wendy J. Holmes
QinetiQ
Malvern, UK

Stuart J. Bertram
QinetiQ
Malvern, UK

Tim Lamont-Smith
QinetiQ
Malvern, UK

Abstract—This paper studies the performance of object detection algorithms applied to electro-optic (EO), infrared (IR), and synthetic aperture radar (SAR) data. First, we describe the simulation of EO and IR images containing vehicles and people and the merging of measured environmental SAR scenes with target image chips. This produces a large data set for training three deep learning algorithms: RetinaNet, EfficientDet, and YOLOv5. The algorithms were trained with a powerful elastic compute cloud (EC2) instance. Performance on simulated data at inference time, in terms of speed and accuracy, was tested on the EC2 instance and a low size weight and power (SWAP) single board computer. YOLOv5 was the most accurate algorithm and the fastest on the EC2 instance but the slowest on the low-SWAP device. RetinaNet and EfficientDet and produced operationally useful throughput on the low-SWAP device for surveillance applications, with RetinaNet having the higher accuracy. This is believed to be the first time the same algorithms have simultaneously been tested on EO, IR, and SAR data, and compared in a unified framework.

Keywords—*electro-optic (EO) data, infrared (IR) data, low-SWAP hardware, object detection, synthetic aperture radar (SAR)*

I. INTRODUCTION

The armed forces operate surveillance platforms with visible-band electro-optic (EO) and infrared (IR) cameras, and synthetic aperture radar (SAR). Recent progress in machine learning (ML) means more processing tasks can be automated than previously possible. Object detection is of particular interest. However, much research relies on large powerful processing hardware, making use of cloud processing and data centres. This type of system can be used in strategic settings, where data is sent to a central point, but there is a need for low size, weight, and power (SWAP) solutions where on-board processing is required, especially for air platforms. On-board processing has the benefit of reducing dependency on tactical comms channels, which may have low bandwidth or be denied in combat situations. For example, rather than transmitting a high data-rate continuous video stream, just the metadata associated with detected objects could be transmitted. We compare the performance of three object detection algorithms applied to EO, IR, and SAR data on low-SWAP hardware. The remainder of this paper is organised as follows. Section II describes algorithm and hardware selection. Sections III and IV describe the data collection process for video (EO/IR) and SAR. Section V compares performance of the algorithms. Finally, conclusions are given in section VI.

II. ALGORITHM AND HARDWARE SELECTION

The aim of object detection is to process an image to produce bounding box coordinates for each object, a class label, and a confidence score. Post processing can be applied, such as thresholding confidence scores to trade true detection

This work was funded by Dstl under the Serapis Lot 4 (assured information infrastructure) programme and the QinetiQ Fellow scheme.

rates against false alarms, or tracking bounding boxes between successive video frames to provide smoother estimates.

A literature review was conducted to find object detection algorithms appropriate for application on low-SWAP hardware [1]. Three algorithms were identified as being particularly suitable: You Only Look Once (YOLO), RetinaNet, and EfficientDet. These algorithms are designed to attain high accuracy while having low computational cost.

YOLO was specifically designed for fast processing [2]. It was followed by various versions up to YOLOv5 [3][4][5][6]. The YOLO convolutional neural net has 53-layers with residual connections. The system predicts the class and “objectness” for each bounding box, which are modelled as offsets from learnt anchor points, instead of absolute image coordinates. The structure can trade speed and accuracy by scaling image size. Hierarchical labelling allows combination of detection and image datasets during training. The various versions of YOLO have widely been used as baselines for comparison of other algorithms. This paper examines the small (s) and extra-large (x) versions of YOLOv5. YOLO is a large model. This is of importance when deploying to low-specification embedded boards because of limited memory, less comms overhead to export new models through wireless update, and less energy consumption [1].

RetinaNet was designed to be fast, while overcoming accuracy issues in earlier models by using a focal loss function to focus on hard-to-classify examples [7]. The structure of the neural net is a feature pyramid network (FPN), containing convolution and max pooling layers. This outputs feature information at different scales. The FPN backbone is the well-known ResNet model, with 50 or 101 layers, and the input image can be scaled to a resolution of 400-800 pixels square, providing methods for trading accuracy and speed. According to some benchmarks, RetinaNet is more accurate but slower than YOLOv3. However, YOLOv4 is faster and has a higher accuracy than RetinaNet [5].

The EfficientDet series of models was designed to allow trade-offs between accuracy and speed [8]. It is based on the principal that convolutional neural nets should scale width, depth, and resolution simultaneously to achieve the best performance for a given compute power. Separate optimum scale factors were found for each of these parameters. The structure of the net includes a bi-directional FPN, allowing feature information at different scales to flow up and down. Eight models are available from D0 (fastest but least accurate) to D7 (most accurate but slowest). In the present paper, the D3 variant was used for EO/IR data and D2 for SAR. According to some benchmarks, EfficientDet has a higher accuracy than RetinaNet for the same processing speed. The maximum accuracy achievable by EfficientDet is higher than YOLOv4.

However, for the same accuracy, EfficientDet is slower than YOLOv4, giving a trade between accuracy and speed [5].

In this paper, the post-processing step of non-maximum suppression was applied for RetinaNet and EfficientDet, with a maximum intersection-over-union value of 0.3 to prevent multiple predicted boxes being produced for a single object. This step is inherently part of the YOLOv5 implementation. The performance of the above algorithms was analysed independently to verify whether the above conclusions, made by the authors of each algorithm, hold.

A review of low-SWAP edge computing devices was conducted [1]. The aim was to select a particular single board computer (SBC) to demonstrate operation of the algorithms. An SBC is a complete computer built on a single circuit board, with microprocessor(s), memory, input/output, and other features required of a functional computer. SBCs have a high level of integration and low component counts. This gives them a lower-SWAP than comparable multi-board computers. The Nvidia Jetson Xavier NX was selected as it has a good trade-off compared to other options: small size (70 mm x 45 mm); low mass (24 g); medium power consumption (10-15 W); high processing ability (6000 GFLOPS with 384 CUDA cores); low cost; and good support. The NX board comes in two form factors. The production module is smaller than a credit card and is what would be mounted on a sensing platform. The developer kit extends the production module with extra components for easy connection to peripherals. The developer kit was used for tests in this paper, as the aim of this work was to rapidly test a proof of concept, rather than develop a productionized system for deployment. It is expected bespoke hardware solutions could further improve the power efficiency of the system. Computation of power efficiency for each algorithm was out of scope for this work.

III. VIDEO DATA AND SIMULATION METHODS

It can be expensive to gather the data volume required by deep learning algorithms. A military trial requires extensive organisation of vehicles, sensors, and personnel. It may not be possible to gather much data about adversaries. Annotation of data can be laborious. Simulations are increasingly important in ML as they allow large data volumes to be created and annotated automatically at low cost. Good results can be achieved by training models on purely synthetic data [9][10].

Virtual Battlespace 3 (VBS) is a commercial synthetic environment simulator to simulate realistic EO and IR images and was used to simulate all such data described in this paper. To ensure simulations are representative of real-life scenarios it is necessary to select an environment that includes real data features. VBS includes the Rahmadi terrain, a fictional island with desert-like conditions with shrubs, trees, and a cluster of buildings. This is a suitable environment as it allows study of object detection in open and built-up areas (where partial obscuration by buildings is likely), and the littoral environment. Fifteen areas on the island, covering these different types of background, were defined as regions in which objects of interest could randomly be placed. Ten objects were simulated, including armoured personnel carriers (APCs), infantry fighting vehicles (IFVs), tanks, 4x4s, and people. The specific objects are BMP3 (tracked IFV), BMPT, (tracked armoured fighting vehicle), Bumerang (wheeled amphibious APC), Kurganets-25 (tracked: APC and IFV variants), T14 (main battle tank), T15 (tracked IFV), T90S

(battle tank), Gaz Tiger (4x4 vehicle), and Infantry Leader (person). The mix of objects includes a combination of visually distinct vehicles, such as the BMP3 and Gaz Tiger, and highly similar vehicles, such as the APC and IFV variants of the Kurganets-25. This variety allows a range from easy to difficult object detection and classification tasks to be studied.

Three types of scenario were studied: empty scenes with no objects of interest, single objects, and multiple objects. In all cases, simple and complex backgrounds (including buildings, trees, and roads) were included. Simulation details were selected to produce images representative of EO/IR cameras on the Watchkeeper un-crewed aerial vehicle (UAV), based on openly published information [11], but the algorithms have utility on a much wider range of platforms, including small UAVs. For each scenario, location, and object, the sensor platform flew a 360° circle around the object with images captured every 10°. The start angle was chosen randomly to ensure exact multiples of ten were not always present. Sensor platforms are not expected operationally to have the ability to fly a complete circuit of objects. The flight path was chosen to maximise viewing angle diversity so reported results are not dependent on specific angles – objects are easiest to recognise from oblique angles, where the front, side, and top of the object can be viewed simultaneously. The 10° interval was chosen as a compromise between data storage/processing and realism. Video feeds over short time periods contain frames captured at similar angles, which provides little benefit over a single image. Objects begin visually to be more different for wider-angle changes. The sensor platform was set to have an altitude of 3048 m (10,000 ft) and a ground distance of 5 km from the object of interest. The camera was set to point at a random location near the object(s). Objects were not put at the image centre to prevent algorithms falsely learning objects always appear there. The camera field of view (FOV) was set to 0.003 or 0.006. In VBS, this is the angle in radians from image centre to edge. The 0.006 setting is equivalent to a full-width FOV 0.69°, which is near the maximum zoom of the DCoMPASS camera [11]. The 0.003 setting, at a distance of 5 km, produces images similar to what could be obtained with 0.006 at 2.5 km. In multi-object scenarios, the orientation of each object was set randomly so all objects do not point in the same direction. The minimum spacing between objects was set to 10 m, just larger than the longest vehicle. The maximum spacing was scaled with the field of view to ensure that the majority of the vehicles appear within the image for all viewing angles. The maximum was set at 30 m for an FOV of 0.006.

The above process was repeated to produce separate sets of EO and IR images. The image resolution for both cameras was set to 1366x768, a standard resolution available in VBS. This is the closest resolution to the DCoMPASS EO camera, 1394x1040. The DCoMPASS IR camera resolution is lower at 640x512. The two camera resolutions were kept the same in simulations so that differences in performance can be attributed to imaging modality rather than resolution. The effect of resolution on performance should be a subject for future study. IR simulations required further work to produce realistic images. When vehicles are first placed in a scene, they are cold by default and relatively featureless. After a while, detail begins to appear. To deal with this, camera capture was not initiated until 90 s after the beginning of the simulation. There is a trade-off in image quality against simulation time. Longer waiting times improve the detail on vehicles. To

produce high-quality images, the VBS render resolution was set at 114% of the captured image resolution, the smallest setting above 100%. These additional settings were applied to the IR and EO data to ensure consistency between the sets.

All simulation settings and processing steps were defined using the Status Quo Function (SQF) language, a scripting language used by VBS to run simulations automatically. This was preferable to the alternative of manually defining and running simulations. Although manual definition may allow a more nuanced set-up of object locations and scenarios, it does not allow for the high volume and variety of data generation necessary for developing a good training database. The VBS simulation has a record of the location and extent of each object in the 3D world. The eight corners of the 3D bounding box in world coordinates were converted to locations in 2D image coordinates. The min and max X and Y coordinates of these points were recorded to produce a 2D bounding box for the object in the image. This enables boxes of different sizes and aspect ratios to be produced as needed.

Annotated images were manually reviewed. In several cases, objects were hidden either partially or fully behind environmental features, such as buildings or trees. Obscured objects have two effects. The first is that if images with hidden objects were presented to a classifier during the training stage, the algorithm would “learn” that objects sometimes appear behind buildings. This would cause spurious detections to be made on unseen data with buildings, even where there is no evidence an object is present. The second effect is present when measuring classifier performance. If an object is hidden, the classifier is likely to fail to detect it. However, the ground-truth annotation has a record that an object is present and the classifier output would be marked as incorrect. According to a recent review [12] “very little work has been devoted to occlusion handling for generic object detection” and “the occlusion problem is far from being solved”. To avoid these problems, images were only kept in the data where more than half of the objects are “detectable”. Detectable was defined as where at least half of the object is visible, since recognition rates drop rapidly with occlusion proportion, and performance with half or less of an object visible is generally very poor [13]. In [14] objects are considered to exhibit “heavy occlusion” if less than 65% is visible. Determination of which images to keep was done manually.

Using the above simulation procedure, 24188 images were produced in total, 12020 for EO, and 12168 for IR. The data were split so that 11 simulation locations were used for training and four different locations for testing. Each object had approximately 997 examples in the training database. The images were saved in the lossless compression PNG format. Future work should consider whether using a lossy compression format like JPG would affect the results.

IV. SAR DATA AND SIMULATION METHODS

Comprehensive SAR data sets with annotated objects in large scenes are not available. Therefore a data set was constructed by combining available background scene data with target vehicle information from the public release of the moving and stationary target acquisition and recognition (MSTAR) data set [15][16]. The output was a new data set of large scenes containing multiple target vehicles.

The background scene data was selected as follows. QinetiQ owns a number of SAR datasets collected under

various MOD programmes. A particular set of images, with aerial landscape views containing no targets, was identified for use. This dataset was collected during field trials in 2010 using the PodSAR radar mounted on a Tornado aircraft. The images primarily consist of fields, trees, hedges, bushes, and roads, with a few buildings or structures. In radar terminology, this type of data is known as “clutter” to distinguish it from targets of interest. Seven large images were identified, ranging in size from 1752 to 23940 pixels wide and 4024 to 14204 pixel high, with pixel spacing 0.225 m to 0.407 m, radar resolution 0.3 m to 0.466 m, and area 0.678 to 51.986 km². To ensure that no information about the background can leak into decisions about whether a target is present in the scene, the first six images were used in training and the final image used in testing. The test image has a higher contrast than the training images. It also exhibits a meandering beam pattern as the strip-map collection took place over a distance of about 9 km. These differences ensure that reported results are not over-trained to particular image characteristics. As the test image is very wide, it was split into nine non-overlapping sub images before further processing.

The public release version of the MSTAR dataset used for this project contains target images for: one of three T72 main battle tanks (MBTs); one of three BMP2 APCs; or a single BTR70 APC. The BMP2 vehicles are different serial numbers of the same variant. The T72 MBTs are different variants. The images are 128 x 128 pixels in size and have a radar resolution of 0.3048 m (1 foot) and pixel spacing 0.202 m in range and cross range. Training data was collected at an elevation of 15° and test data was collected at 17°.

The background and target datasets were merged by inserting small target image chips into larger background images. The background images were manually analysed to find suitable areas where targets could be inserted. Before insertion, the following processes were applied to target image chips. First, rotation by 90° so shadows point down to match the background. Second, application of an alpha-channel transparency window. This is a full-amplitude circle centred on the chip centre, with a radius of 25 pixels, and a linearly decreasing alpha value (opaqueness) in a radial direction from the circle edge to the chip edge. This reduces insertion artefacts, making the final image more realistic. Third, the amplitude of the chip was scaled so the non-target mean matches the local mean of the background. In addition to target chips, clutter chips from the MSTAR dataset were randomly inserted into background scenes. This was done to ensure the classifier does not use any subtle insertion artefacts to help with the target detection process. Shadows in the MSTAR data are more pronounced than those in the QinetiQ clutter data. To ensure the two datasets match, noise of an appropriate power level was added to the target chips prior to insertion. Bounding box annotations were fixed to be 32x32 pixels in size, centred on the inserted chip centre.

The dynamic range of radar images is inherently very high, as radar returns from targets and buildings can be orders of magnitude brighter than those from fields and trees. In operational use, analysts adaptively adjust the brightness and contrast of the screen to understand the scene. For fields, higher contrast is required, but for buildings, lower contrast is required. In this work, image amplitudes were clipped so that pixels brighter than a certain threshold were set to the threshold. The thresholded image was then amplitude-scaled

to fit the dynamic range of the PNG format. The threshold was set to 98% for training images and 95% for test images. This partly accounts for differences in the images but also ensures the reported algorithm performance is not over-trained to a particular threshold.

The above process produces images containing background scenes and targets. The images have different dimensions from each other. The object detection algorithms in this paper scale images so they are a standard size before input. This is appropriate for EO and IR images because objects appear as different sizes depending on how close they are to the camera and the level of zoom. Examples of objects at multiple scales are required in training. However, the physics of SAR image generation creates pixels, and hence targets, at the same size regardless of the sensor-target distance. This means the physical size of objects is encoded in the image. This is useful information for discriminating between targets and would be lost if whole images were scaled on input to the neural net. To avoid this, large images were “tiled” to produce constant size images of 768x768 pixels in size. In total, there are 1000 training image files, each containing zero, or more, inserted target and clutter chips (850 images with targets and 150 without) and 315 test image files, each containing zero, or more, inserted target chips (268 images with targets and 47 without). The images all have a pixel spacing of 0.26 m and a radar resolution of 0.3 m. The training database contained an average of 2072 examples per class. The scale issue was addressed in [17] by using sliding windows instead of tiles to process large images, but that network is only applicable to fixed-scale SAR images and not EO/IR. There are an order of magnitude fewer images for SAR than the EO/IR datasets. This is due to the manual nature of the radar data background and target merging process. With further work, improvements could be made to allow a greater degree of automation and increase the data volume.

V. ALGORITHM PERFORMANCE

The algorithms were trained on a G4dn.xlarge EC2 instance on Amazon Web Services (AWS). For each algorithm, the weights were initialized from published pre-trained models trained on photographs. It was found that using pre-trained, rather than randomized, initial weights improved performance for all algorithms and datasets tested, including IR and SAR, even though the sensing modality of the pre-training data is different. This approach contrasts to [17], which only uses SAR data and no pre-training. A simple hyperparameter search was performed to obtain good values for the batch size, learning rate, and confidence threshold. The batch size was limited by the memory of the EC2 instances to 6, 2, 16, or 2 images for RetinaNet, EfficientDet, YOLOv5s, or YOLOv5x with EO/IR data, and 8, 4, 16, or 8 with SAR data. The training duration and number of epochs with the EO data, for each of the algorithms was: RetinaNet (13 hr, 17), EfficientDet (88 hr, 99), YOLOv5s (5 hr, 40) and YOLOv5x (15 hr, 20). Similar figures were obtained for the IR data. Training duration was an average of 12 times faster for the SAR data due to the smaller size of the images and dataset.

Each trained model was used to process test data on the EC2 instance and the low-SWAP NX board. A manual split of train/test data ensured no similar backgrounds in the train data were in the test data. For this reason automated cross-validation was not used. Several performance metrics were computed for each algorithm and data set, including precision,

recall, false alarm rate, and confusion matrices. Full results are in [18]. The primary metrics are mean average precision (mAP) and processing time per image. Results for the EC2 instance are shown on the left of Fig. 1. Both RetinaNet and EfficientDet have several larger or smaller variants of model than those tested here, allowing for trades between speed and accuracy. The line associated with each data point in the figure is an estimate of what different accuracies or speeds could be expected to be achieved with the variant models, using data given in [8]. The YOLOv5 performance curve is shown for the smallest (s) and largest (x) models. On the EC2 instance, YOLOv5 is more accurate and faster than the other algorithms for all datasets. The low mAP for EfficientDet in EO and IR data is driven by missed detections; the other algorithms have better detection performance. Example RetinaNet detections in EO and SAR data are shown in Fig. 2 and Fig. 3.

Results for the NX board are shown on the right of Fig. 1. RetinaNet is more accurate than EfficientDet for the same processing speed, which mirrors the EC2 results. The speed for both algorithms is approximately two times slower on the NX board than EC2. YOLOv5 is more accurate, but slower than the other algorithms by a factor of about 4 for the small (s) model and 35 for the larger (x) model. Unlike on AWS, the PyTorch implementation of YOLOv5 was unable to make use of the OpenMP interface on the NX board, which significantly impeded the algorithm throughput. It is thought that if this issue were resolved the algorithm could be sped up.

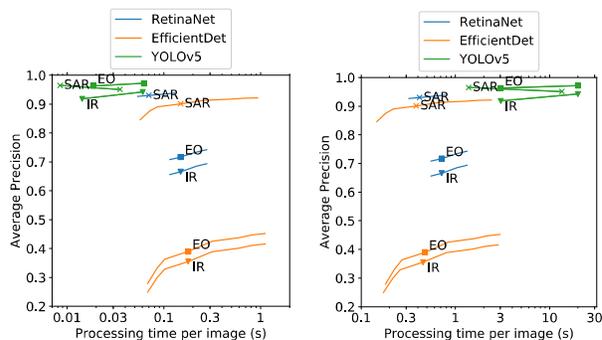


Fig. 1. Accuracy-speed trade-off. Left: G4dn.xlarge. Right: NX.

One aim of this work was to see whether algorithms could be optimised for speed on low-SWAP hardware. RetinaNet and EfficientDet are implemented in the Keras/TensorFlow framework, but they are not highly optimised for specific hardware. Nvidia TensorRT (TRT) provides run-time performance optimisations for tensor operations on specific hardware platforms like the NX board. These include specialised libraries and hardware and lower precision numbers for the calculations, sacrificing small amounts of accuracy for increases in speed. TRT is integrated into TensorFlow as a module called TF-TRT. This can be accessed in Keras, allowing models to fall back to native TensorFlow when required. This gives flexibility in model support at the expense of performance, compared to pure binary TRT models. Each EfficientDet or RetinaNet model was converted to an optimised 16-bit floating-point format using TF-TRT. This format is specific to the hardware where it is generated. Only a subset of operations in each model are supported by the library for optimisation, so it is possible to create neural nets with a significant proportion of nodes that cannot be converted to TRT functions. Additionally, TRT only

optimises subgraphs of a minimum size to reduce the overhead of swapping to/from optimised formats. This results in some otherwise optimisable steps not being optimised due to context. These limitations mean some models have no performance gain because there is little optimised code.

For each combination of EfficientDet/RetinaNet and EO/IR/SAR data, the accuracy and throughput of the algorithms at inference time was measured with both optimised and un-optimised versions of the models. The inference time per image is about ten times longer for the first image in a video stream, to allow various processes be loaded into memory. Thus, we analysed the truncated mean of the inference times for each stream. We found a 20-85% increase in throughput for EfficientDet but a greater than 80% decrease in throughput for RetinaNet, without a significant change in accuracy in either case. The decreased RetinaNet speed appears to relate to the amount of un-optimised steps and the overhead of repeatedly swapping between optimised and un-optimised parts of the network state. Due to time limitations and the fact that YOLOv5 is based on PyTorch rather than TensorFlow, optimisation of YOLOv5 was not attempted. However, a short experiment to optimise an off-the-shelf YOLOv4 model, which is based on TensorFlow, showed that image throughputs of 20-50 frames per second were possible with 416x416 size images. This demonstrates that the NX platform is capable of high throughput with an appropriate model.

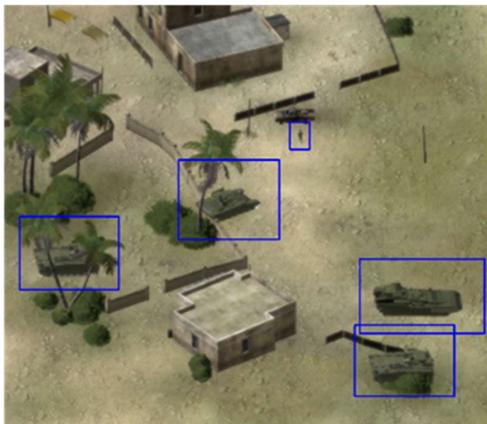


Fig. 2. Example detections in an EO image.

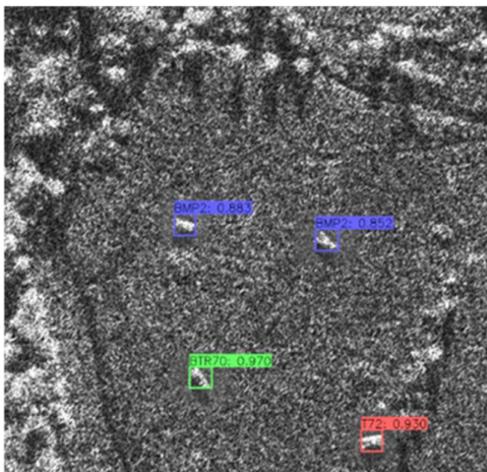


Fig. 3. Example detections in a SAR image.

VI. CONCLUSIONS

Three object detection algorithms, RetinaNet, EfficientDet and YOLOv5 have been analysed. They can be applied to any type of image as long as suitable training data is available. Successful application of the algorithms to simulated EO, IR and SAR data has been demonstrated. This is believed to be the first time the same set of algorithms has simultaneously been tested on these three sensor types and compared in a unified framework. When tested on an EC2 instance, YOLOv5 was the best algorithm in terms of both accuracy and speed. On the off-the-shelf Nvidia Jetson Xavier NX low-SWAP device RetinaNet and EfficientDet produced operationally useful throughput for surveillance applications on a UAV, with reasonable accuracy. YOLOv5 had the better accuracy but did not achieve high throughput due to implementation issues.

ACKNOWLEDGMENT

We thank Dylan Rivers for help with VBS simulations and set up of the object-detection processing pipeline, and Alan Blake for the supply of SAR data.

REFERENCES

- [1] A. Hadland, AI analysis using low-SWAP hardware: literature review, QinetiQ/20/02995, Sep. 2020.
- [2] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, You only look once: unified, real-time object detection, IEEE Conf. on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, Jun. 2016.
- [3] J. Redmon, A. Farhadi, YOLO9000: better, faster, stronger, IEEE Conf. on Computer Vision and Pattern Recognition, USA, Jul. 2017.
- [4] J. Redmon, A. Farhadi, YOLOv3: An incremental improvement, arXiv:1804.02767, Apr. 2018.
- [5] A. Bochkovskiy, C.Y. Wang, H.Y.M. Liao, YOLOv4: Optimal speed and accuracy of object detection, arXiv:2004.10934, Apr. 2020.
- [6] D. Thuan, Evolution of YOLO algorithm and YOLOv5: the state-of-the-art object detection algorithm, Bachelor's Thesis, Oulu University of Applied Sciences, 2021.
- [7] T.Y. Lin, P. Goyal, R. Girshick, K. He, P. Dollár, Focal loss for dense object detection, IEEE Int. Conf. on Computer Vision, Oct. 2017.
- [8] M. Tan, R. Pang, Q.V. Le, EfficientDet: scalable and efficient object detection, arXiv:1911.09070, Apr. 2020.
- [9] C.P. Moate, S.D. Hayward, J.S. Ellis, L. Russell, R.O. Timmerman, R.O. Lane, T.J. Strain, Vehicle Detection in Infrared Imagery using Neural Networks with Synthetic Training Data, Int. Conf. on Image Analysis and Recognition, Póvoa de Varzim, Portugal, Jun. 2018.
- [10] R.O. Lane, Electrical device classification using deep learning, Sensor Signal Processing for Defence, Sep. 2020.
- [11] Elbit Systems, DCoMPASS: digital compact multi-purpose advanced stabilized system, 2016.
- [12] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikäinen, Deep learning for generic object detection: a survey, Int. J. of Computer Vision, 128(2):261-318, Feb. 2020.
- [13] Z. Zhang, C. Xie, J. Wang, L. Xie, DeepVoting: a robust and explainable deep network for semantic part detection under partial occlusion, Conf. CVPR, pp. 1372-1380, Jun. 2018.
- [14] S. Zhang, J. Yang, B. Schiele, Occluded pedestrian detection through guided attention in CNNs, Conf. CVPR, pp. 6995-7003, Jun. 2018.
- [15] T.D. Ross, S.W. Worrell, V.J. Velten, J.C. Mossing, M.L. Bryant, Standard SAR ATR evaluation experiments using the MSTAR public release data set, Proc. SPIE 3370, Sep. 1998.
- [16] R. Schumacher, K. Rosenbach, "ATR of battlefield targets by SAR – classification results using the public MSTAR dataset compared with a dataset by QinetiQ, UK", NATO Symp. on Target Identification and Recognition Using RF Systems, Oslo, Norway, Oct. 2004.
- [17] S. Chen, H. Wang, F. Xu, Y.Q. Jin, Target classification using the deep convolutional networks for SAR images, IEEE Trans. Geoscience and Remote Sensing, 54(8):4806-4817, Aug. 2016.
- [18] R.O. Lane, W.J. Holmes, A.J. Wragge, S.J. Bertram, T. Lamont-Smith, AI Processing Final Report, QinetiQ/21/00278/1.2, May 2021.

Exponential Filters for Passive Underwater Acoustic Detections - A Global Processing Gain Perspective

Stéphane Blouin (Senior IEEE member)
 Defence R&D Canada (DRDC), Atlantic Research Centre
 Dartmouth, NS, Canada, B2Y 3Z7

stephane.blouin2@forces.gc.ca

Abstract—Multiple passive acoustic detectors exist for under/above-water applications. Implementations involving exponential filters have also been tested. However, the literature provides little guidance about their tuning, which may be non-intuitive. To alleviate this shortcoming, this publication reviews the key configuration parameters of a particular implementation of such a detector. Performance metrics are also proposed and their relationship to configuration parameters is established. The computation of the global processing gain is the main focus of this publication, and of interest to practitioners the impact of configuration parameters is studied. Examples demonstrate how the proposed closed-form expression adequately matches simulated results. Also, a novel and approximate output variance expression is proposed. Early results indicate the need for optimization among many performance metrics.

Index Terms—signal processing, detection, filter, processing gain, acoustics, configuration

I. INTRODUCTION

Passive acoustics has been of interest for military purposes since WWII [1], with modern applications inclusive of wildlife monitoring [2], [3]. Passive acoustics aims to “pull out” a signal of interest from ambient noise. Conventionally, the very first stage involves target detection, then followed by localization/tracking [4]. Detectors can also be used to assess network connectivity [5].

The most elementary passive acoustic detector is based on energy levels. To that effect, the SONAR equation [6] refers to a constant detection threshold, which in turns relates to performance. Usually, the type of acoustic signals encountered (narrow-band, broadband, transient) must be known *a priori* to better tune such detectors.

DRDC has developed and tested an underwater passive acoustic detection scheme based on exponential filters [7], [8]. This detector is also energy-based, but unlike many schemes it performs a noise auto-normalization. A variant of this scheme has been applied to cetacean detections [9]. Such a filtering scheme can react appropriately to environmental phenomena like squalls and rainfalls. However, the number of configuration parameters and their interactions render its implementation challenging. Also, references rarely treat interactions between successive signal processing steps from a performance standpoint, which practitioners highly value. This publication addresses both shortcomings by focusing on the global processing gain

with illustrative examples to convey results about the impact of some key configuration parameters.

Section II describes the detection scheme. An overview of global performance metrics steps and the derivation of processing gain are provided in Section III. Section IV presents global performance results. Simulated and real-implementation results are provided in Section V before summarizing and concluding with Sections VI and VII. To be self-contained, this publication reviews concepts that could be considered trivial by some readers before disclosing new material.

II. DETECTION SCHEME DESCRIPTION

\mathbb{R} and \mathbb{N} describe real and natural numbers, respectively. Letters f , N and T represent frequency (Hz), number of elements in a set, and duration. The letter $t \in \mathbb{R}^+$ ($k \in \mathbb{N}^+$) represents a positive continuous (discrete) time.

The block diagram of Figure 1 depicts key processing steps of the proposed detector. Processing steps of Figure 1 convert an analog continuous-time pressure signal $p(t) \in \mathbb{R}$ to a digital one $p(k) \in \mathbb{R}$, and ultimately to a binary value determining the presence or absence of a target (or acoustic source). Blue-font variables in Figure 1 are configuration parameters defined in Table I.

TABLE I: Configuration parameters.

Symbol	Meaning
f_s	A-D sampling frequency
N_{FFT}	number of FFT points
$overlap$	% overlap btw FFT samples
N_A	average sample size
γ_1, γ_2	filter time-constants
T	threshold value
$\%P$	% bins $> T$, target present
$\%N$	% bins $< T$, target absent

The first computational step of Figure 1 digitizes the time-varying and analog pressure signal $p(t)$ through an analog-to-digital conversion requiring a user-specified sampling frequency $f_s \in \mathbb{R}$ (s^{-1}), which leads to a sampling time T_s . The subsequent step is a holding mechanism, usually a zero-order hold. Not shown here is the quantization converting the original continuous magnitude to a finite value set whose size depends on the required precision.

The second block of Figure 1 applies shading and the Fast Fourier Transform (FFT). The Hanning window is

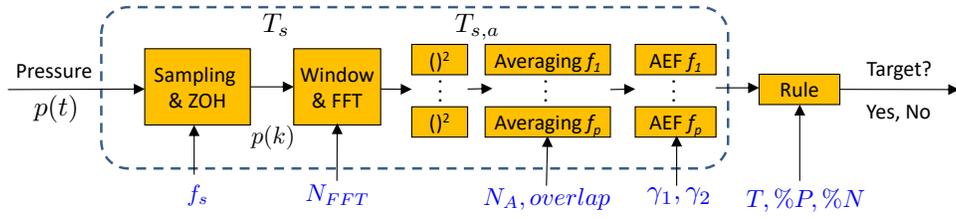


Fig. 1: Block-diagram representation of the detection scheme where blue-font variables are configuration parameters (AEF = Arrangement of Exponential Filters). Only processing steps within the dashed lines are treated here.

selected here but its effect on the global processign gain is ignored. As a reminder, the FFT uses $N_{FFT} \in \mathbb{N}$ data points $p(k), p(k + 1), \dots, p(k + N_{FFT} - 1)$ of signal $p(k)$ and decomposes it in a collection of complex values $a_1 + jb_1, \dots, a_n + jb_n$ quantifying the amplitude and phase of a set of n equally-spaced frequencies of resolution

$$f_{resolution} = \frac{f_s}{N_{FFT}} = \frac{1}{T_{int}}, \quad (1)$$

where T_{int} is the FFT integration time (s), i.e, the time period captured by a single FFT. The $()^2$ step of Figure 1 converts the array of n complex numbers $[a_1 + jb_1, \dots, a_n + jb_n]$ to n real values by performing the following non-coherent operation $c_p = a_p^2 + b_p^2$ with $p = 1, 2, \dots, n$, thus leading to an FFT power spectrum. Subsequent steps until the rule block of Figure 1 only deal with one frequency bin.

The averaging step in Figure 1 utilizes $N_A \in \mathbb{N}$ samples of FFT power spectra and a fixed percentage of overlap among consecutive FFT outputs. The overlap determines how many discrete-time samples are shared among two consecutive FFT's. Therefore, $T_{s,a}$, the sampling time (period between two consecutive data points) at the input (and output) of the moving-average step is given by

$$T_{s,a} = T_s \left(N_{FFT} \frac{100 - \%overlap}{100} \right), \quad (2)$$

thus impacted by parameters N_{FFT} and $\%overlap$.

The AEF block of Figure 1 stands for ‘‘Arrangement of Exponential Filters’’. It is beyond the scope of this paper to discuss the merit/inconvenience of a particular arrangement (ratio versus difference, for instance) of exponential filters. Each exponential filter is defined by

$$y(k') = \gamma u(k') + (1 - \gamma)y(k' - 1), \quad (3)$$

where $0 \leq \gamma \leq 1$ is the filter time-constant, $u(k')$ the measurement/input at the new discrete time index k' based on sampling period in (2), and $y(\cdot)$ is the filter output. Here, input signals are averaged power spectra. The current setup makes use of the ratio (in dB) of the output of two exponential filters with distinct time-constants γ_1 and γ_2 where $\gamma_1 > \gamma_2$, $\gamma_{min} = \min(\gamma_1, \gamma_2)$ and $\gamma_{max} = \max(\gamma_1, \gamma_2)$. Values for γ_1, γ_2 are usually chosen based on the nature of signals and noise to filter out.

The rule of Figure 1 involves three parameters $T, \%P$, and $\%N$ defined in Table I and omitted from this analysis. Overall, this detector has nine configuration parameters.

III. PERFORMANCE METRICS AND PROCESSING GAIN

The individual impact of the first four configuration parameters of Table I was analyzed in [10]. Due to page limits, only a subset of those results is presented here.

A. Performance metrics

This analysis uses performance metrics characterizing the stability, long-term (steady-state)/transitory behaviours, and key features of the proposed scheme.

Bounded-input-bounded-output (BIBO) is the main stability metric of interest. Two steady-state performance metrics, the processing gain PG and the cut-off frequency f_c , as well as one transient metric, the settling time t_s , were studied. Why are those parameters important? The larger PG is, the easier a target will be ‘‘pulled out of noise’’. The cut-off frequency f_c becomes important because the overall scheme acts as a low-pass filter and a lower f_c means that fewer high frequencies will remain. The settling time is relevant in the sense that ideally, the next processing step should not react based on short-lived and potentially misleading transient behaviours.

The global characteristics resulting from steps described in the dashed box of Figure 1 are $PG_{global}, f_{c,global}$, and $t_{s,global}$. Each processing step impacts the global performance of the detection scheme as indicated in Table II. Other features of importance are $f_{resolution}$ and T_a , the frequency resolution (shown in equation (1)) enabling the separation of two nearby frequencies and the length of time averaging (developed in [10]), respectively.

TABLE II: Influence of configuration parameters on global performance criteria

	f_s	N_{FFT}	$overlap$	N_A	γ_{min}	γ_{max}
$f_{c,global}$	✓	✓	✓	✓	—	✓
PG_{global}	—	✓	—	✓	✓	✓
$t_{s,global}$	✓	✓	✓	—	✓	—
$f_{resolution}$	✓	✓	—	—	—	—
T_a	✓	✓	✓	✓	—	—

Based on Table II, the number of configuration parameters is not overwhelmingly large. However, intricacies arise

because almost all configuration parameters affect many performance criteria, including the global processing gain whose key derivation steps are presented next.

B. Processing gain

For any of the processing step defined in Figure 1, we assume that a signal polluted by noise is the process input and that a noisy output signal is the response of the said process subject to such input signal. In this publication, the input-output processing gain PG means

$$PG = \frac{SNR_{output\ signals}}{SNR_{input\ signals}} = \frac{SNR_O}{SNR_I},$$

where each SNR (signal-power-to-noise-ratio) is

$$SNR = \frac{\bar{P}_s}{\bar{P}_n} = \frac{\bar{P}_{s+n} - \bar{P}_n}{\bar{P}_n} \approx \frac{RMS(P_s)}{RMS(P_n)}, \quad (4)$$

with the average power for signal “ s ” and noise “ n ”, \bar{P}_s and \bar{P}_n , and the Root Mean Square (RMS), thus giving

$$PG = \frac{\bar{P}_{s,O}/\bar{P}_{n,O}}{\bar{P}_{s,I}/\bar{P}_{n,I}} = \left(\frac{\bar{P}_{n,I}}{\bar{P}_{n,O}} \right) \times \left(\frac{\bar{P}_{s,O}}{\bar{P}_{s,I}} \right).$$

From the above development, when the processing step (FFT, moving/exponential averaging, etc.) does not induce a permanent input-to-output gain change for a particular input signal (excluding the noise component), then the processing gain reduces to the ratio of noise average powers $\bar{P}_{n,I}/\bar{P}_{n,O}$. In the opposite case, then a frequency-dependent expression can be derived from the processing step impulse response as shown in [11].

If $\bar{P}_{s,O}/\bar{P}_{s,I} \approx 1$ and in presence of a stationary and ergodic process generating a zero-mean noise, an assumption commonly found in literature, then the signal power reduces to its variance σ^2 [11, p.280], which then leads to

$$PG = \frac{\bar{P}_{n,I}}{\bar{P}_{n,O}} = \frac{\sigma_{n,I}^2}{\sigma_{n,O}^2} = \left(\frac{\sigma_{n,I}}{\sigma_{n,O}} \right)^2. \quad (5)$$

Both processing gain and SNR can be expressed in an absolute or dB scale. Note that distinct dB scalings apply when considering SNR versus $SaNR$ (signal-amplitude-to-noise-ratio) or signal powers versus amplitudes. Note that PG and SNR are interchangeable metrics for quantifying the benefits of a particular processing step.

IV. PERFORMANCE ANALYSIS

Only the PG_{global} performance metric is presented here in full as it is the most compact development. Each of the three main steps (FFT, averaging, and exponential filters) are described next before deriving PG_{global} .

A. FFT

As shown in [10], the maximal gain of the FFT processing can be expressed as follows (assuming $\sigma_{n,O}^2/\sigma_{n,I}^2 \approx 1$)

$$PG_{FFT}^{dB}(N_{FFT}) = 10 \log_{10} \left(\frac{N_{FFT}}{2} \right), \quad (6)$$

so a smaller gain is expected in reality.

B. Moving average

The proposed incoherent moving-average filter reduces the variations in the background noise power by averaging over N_A successive FFT's. The reduction in the output noise standard deviation over N_A FFT's, that is $\sigma_{N_A\ FFT's}$, relative to the output noise standard deviation of a single FFT, that is $\sigma_{single\ FFT}$, is given by

$$\sqrt{\left(\frac{\bar{P}_{n,O}}{\bar{P}_{n,I}} \right)_a} = \frac{\sigma_{N_A\ FFT's}}{\sigma_{single\ FFT}} = \frac{1}{\sqrt{N_A}}. \quad (7)$$

From Figure 1 the input signal of the incoherent averaging is made of power spectra, which explains why the $SaNR$ expression (7) is the basis for deriving PG_a , in combination with the fact that $\bar{P}_{s,O}/\bar{P}_{s,I} = 1$ for the moving-average filter [10]. Expression (7) holds for most windows and overlap of 50% or less [12]. With averaging performed on the input signal, the gain of the incoherent averaging can be expressed by [12, P.421]

$$PG_a^{dB}(N_A) = 10 \log_{10} \sqrt{N_A}. \quad (8)$$

C. Exponential filter

The capacity of a single exponential filter to reduce the noise variance strictly depends on the value of γ [13]

$$\left(\frac{\bar{P}_{n,O}}{\bar{P}_{n,I}} \right)_{exp} = \left(\frac{\sigma_{n,O}^2}{\sigma_{n,I}^2} \right)_{exp} = \frac{\gamma}{2 - \gamma}, \quad (9)$$

if time is allowed to grow indefinitely large. Using the exponential filter impulse response, one can show [10] that each exponential filter has unit process gain thus giving

$$PG_{exp}^{dB} = 10 \log_{10} \left(\frac{2 - \gamma}{\gamma} \right). \quad (10)$$

D. Exponential filter arrangement

Whether one considers the ratio or the difference of exponential filters, the output of such an arrangement needs to be looked at from the variance of combined signals. When the Ratio of Exponential Filters (REF) is considered, the approximate expression for variance of the ratio of two signals in [14] is exploited. Here, only step-like input signals of magnitude A feed the exponential filters such that the variance of the REF reduces to (see ref. [10])

$$\sigma_{s+n,O}^2 \left(\frac{y_1}{y_2} \right) \approx \frac{\sigma_{n,I}^2}{A^2} \left[\frac{\gamma_1}{2 - \gamma_1} - \frac{2cov(y_1, y_2)}{\sigma_{n,I}^2} + \frac{\gamma_2}{2 - \gamma_2} \right], \quad (11)$$

where y_1, y_2 are exponential filter output signals associated to γ_1, γ_2 . Figure 2 shows how expression 11 compares favorably to that of [14] and also to values directly computed from manufactured datasets when $\gamma_1 = 0.1$ and $A = 1$.

As $\sigma_{s+n,O}^2(y_1/y_2) \approx \sigma_{n,O}^2(y_1/y_2)$ (only dealing with step-like input signals), expression (5) with (11) becomes

$$PG_{REF} \approx A^2 \left[\frac{\gamma_1}{2 - \gamma_1} - \frac{2cov(y_1, y_2)}{\sigma_{n,I}^2} + \frac{\gamma_2}{2 - \gamma_2} \right]^{-1}. \quad (12)$$

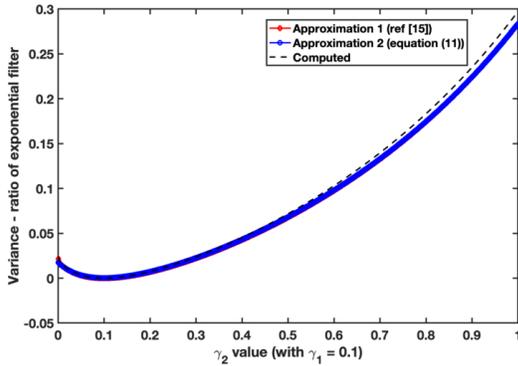


Fig. 2: REF output variance (assumes $A = 1$, $\gamma_1 = 0.1$).

E. Global characteristics

The global processing gain (omitting quantization and the window function) is derived from the sum of all individual processing gains derived in (6), (8) and (12)

$$PG_{global}^{dB} = PG_{FFT}^{dB}(N_{FFT}) + PG_a^{dB}(N_A) + \dots + PG_{REF}^{dB}(\gamma_1, \gamma_2), \quad (13)$$

thus showing the effect of four of the six main configuration parameters discussed earlier and present in Table II. Alternatively, one can use the SNR_{global}^{dB} metric instead.

V. SIMULATED EXAMPLES AND REAL IMPLEMENTATION

This section provides three simulated examples illustrating how to use the previous results to assess global processing gain and how to determine feasible detection ranges dependent upon said gain. The last example is an actual implementation of the scheme shown in Figure 1 on real acoustic data used to illustrate the effect of the N_{FFT} configuration parameter.

Example 1 *This example uses the setup in [8] with: $N_{FFT} = 1024$, $f_s = 2500$, $overlap = 50\%$, $N_A = 6$, and $\gamma_{max} = 0.1$, $\gamma_{min} = 0.001$, and $f_0 = 300$ Hz (thus $\omega = 2\pi f_0/f_s = 0.75$). Using those values in equation (13) gives (using the leftmost value of Figure 2 of 0.0129 and assuming $A = 1$ and $\gamma_1 = 0.1$)*

$$PG_{global}^{dB} = 10 \log_{10}(1024/2) + 10 \log_{10}(\sqrt{6}) + \dots + 10 \log_{10}(1/0.0129) = 49.8,$$

thus showing that contributions to the global processing gain are: 3.9 dB from averaging, 18.9 dB from the ratio of exponential filter, and 27 dB from the FFT.

As shown in Example 1, with small moving-average sample sizes, *i.e.*, $2 \leq N_A \leq 8$, and typical γ_1, γ_2 values, the main contribution to the global processing gain, PG_{global}^{dB} , remains N_{FFT} , the FFT data sample length.

The next example shows how to exploit the global processing gain generated from the scheme of Figure 1 for assessing detection ranges.

Example 2 *With the setup of Example 1, the proposed detector leads to a processing gain of about $PG_{global}^{dB} = G_p = 49.8$ dB. Let us consider a vessel emitting sound at a source level of $SL = 120$ dB [15] and located at a range R (in metres) from the receiver. With a geometric spreading loss of $PL = 20 \log_{10}(R)$ and assuming no absorption losses, the achievable detection range can be computed by solving the SONAR equation $SL - PL - NL + G_p = SE + DT$, where SE =Signal Excess and DT =Detection Threshold as per [16]. Noise level around 100 Hz is assumed to be $NL = 80$ dB, thus mimicking a wind speed of 10 m/s as per the approximation developed in [16]. Assuming an arbitrary sum $SE + DT = 9$ (dB), the resulting range is given by $R = 10^{(120-80+49.8-9)/20} \approx 11$ km.*

Example 2 illustrates how one could quantify the improvement in detection ranges resulting from changes in global processing gain, PG_{global}^{dB} , itself dependent upon configuration parameters N_{FFT} , N_A , γ_{min} and γ_{max} as per (13). Example 3 demonstrates the outcome of a simulated case using the processing steps shown in Figure 1.

Example 3 *As Example 1, this scenario uses the setup in [8] and the processing steps of Figure 1. In this simulation, the “signal” to be detected is a single tone centered at 300 Hz and of magnitude 1. The “noise” is uniformly distributed and also of magnitude 1. The signal plus noise makes up the signal $p(t)$ entering the process chain of Figure 1. The overall simulation lasts for 80 seconds with the signal only becoming active after 20 seconds. Figure 3 displays three SNR curves computed based on a few bins centered around 300 Hz. One curve is associated to the signal passing through the FFT and windowing steps (solid blue line), another curve measures the SNR after the FFT averaging (dashed line), and the last curve relates to the ratio of exponential filters (solid yellow line). As per (4), computed SNR values based on simulated data were obtained by taking the ratio of the output signal RMS values around the 300 Hz frequency bin (thus mostly the “signal”) over the output signal RMS values away from that frequency (thus mostly “noise”). Note that the FFT and averaging SNR values of Figure 3 are near the nominal values of Example 1 and derived based on expressions (6) and (8). The REF SNR is smaller than that found in Example 1 because here $A \approx 0.1$ thus bringing its contribution closer to 0 dB, which is also predicted by equation (12). Therefore, the analytic development of Section IV closely matches simulation results and enables rapid assessments of overall processing gain PG_{global}^{dB} .*

From all past examples, one could be misled thinking that solely increasing N_{FFT} to augment PG_{global}^{dB} is a logical approach. As shown next, augmenting N_{FFT} too much could be a mistake because it can have adverse effects on $f_{c,global}$ and $t_{s,global}$. Space limitations prevent a full treatment of this topic here, but Example 4 shows some of those potential detrimental effects.

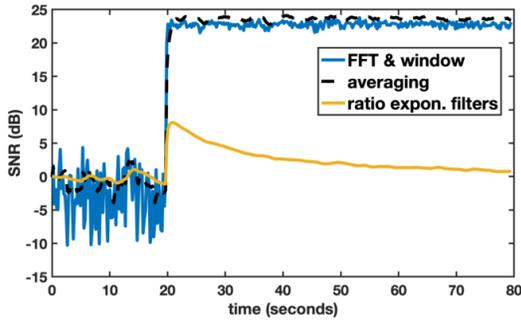


Fig. 3: SNR for major processing steps in Example 3.

Example 4 This example uses underwater acoustic data collected in Gascoyne Inlet (NU) on August 21 2019. The acoustic source consisted of a transducer transmitting three tones at 681, 789, and 1157 Hz. Tones were projected during five instances of varying lengths and then stopped. Configuration parameters were: $N_A = 6$, $\text{overlap} = 50\%$, $\gamma_1 = 0.1$, $\gamma_2 = 0.001$, and $f_s = 3,000$ sps. Other experimental details can be found in [17]. Figure 4 shows the output of the REF for the frequency bin nearest to 681 Hz for two distinct N_{FFT} values (1024 and 4096). The higher N_{FFT} value attenuated noise more efficiently, but the response became sluggish as observed by the slower decay caused by an increased sampling period as predicted by (2). This slow-decay in magnitude response could cause masking issues if a weaker acoustic event follows one that is louder. Such results suggest a need for optimizing parameters based on all performance metrics of Table II.

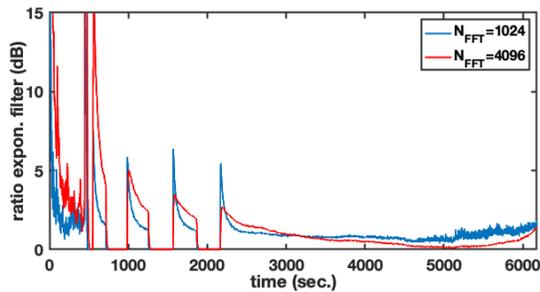


Fig. 4: N_{FFT} effect on ratio of exponential filters.

VI. SIGNIFICANCE OF RESULTS

This self-contained publication provides a procedure for assessing the global processing gain of arrangements of exponential filters. For ratios of exponential filters, a novel approximate output signal variance expression is derived. Also, examples demonstrate the close match between simulated cases and the predicted processing gain. The proposed technique enables SONAR design refinements without the need for simulations and provides a tool for rapidly assessing the impact of configuration parameters.

VII. CONCLUSIONS

This publication reviews the main processing steps of a detection scheme based on exponential filters. Three main steady-state and transient performance metrics are proposed but due to space limitations, only the expression for the global processing gain, that is the gain resulting from all processing steps, is presented here. Examples illustrate: (a) the quality of the proposed technique, (b) an approach exploiting these results for assessing detection ranges, and (c) some conflicting performance outcomes for key configuration parameters, thus indicating the need for optimization.

REFERENCES

- [1] R. A. Holler, A. W. Horbach, and J. F. McEachern, *The Ears of Air ASW: A History of U.S. Navy Sonobuoys*, Navmar Applied Science Corporation, 2008.
- [2] L. Sayuri Moreira Sugai, T. Sanna Freire Silva, J. Wagner Ribeiro JR., and D. Llusia, "Terrestrial passive acoustic monitoring: Review and perspectives," *BioScience*, vol. 69, no. 1, pp. 15–25, January 2019.
- [3] S. Blouin, "Impact of sensor network topology on acoustic communications for protecting marine mammals," in *2019 IEEE International Symposium on Robotic and Sensors Environments (ROSE)*, June 2019, pp. 1–7.
- [4] J.Y. Yu, D. Ustebay, S. Blouin, M. Rabbat, and M. Coates, "Distributed underwater acoustic source localization and tracking," in *Proceedings of the Asilomar Conference on Signals, Systems and Computers*, 2013.
- [5] M. M. Asadi, M. Khosravi, A. G. Aghdam, and S. Blouin, "Joint power optimization and connectivity control problem over underwater random sensor networks," in *2015 American Conference (ACC)*, 2015, pp. 2709–2714.
- [6] Michael A. Ainslie, *Principles of Sonar Performance Modelling*, Springer Praxis Books / Geophysical Sciences, 2010.
- [7] D. Watson, "Next generation autonomous systems - starfish cube array," Tech. Rep. DRDC Atlantic CR2011, Omnitech, 10 Akerley Blvd, Suite 1, Dartmouth, Nova Scotia, March 2011.
- [8] C.E. Lucas, G.J. Heard, and N. Pelavas, "DRDC starfish acoustic sentinel and phase gradient histogram tracking," Scientific Report DRDC-RDDC-2015-R025, Defence Research and Development Canada, February 2015.
- [9] J.D. Hood, D.G. Flogeras, and J.A. Theriault, "Improved passive acoustic band-limited energy detection for cetaceans," *Applied Acoustics*, no. 106, pp. 36–41, 2016.
- [10] S. Blouin, "Sentinel passive acoustic detector - part 1: Individual component analysis and dynamical modeling," Tech. Rep., Defence R&D Canada (DRDC), 2021, scientific report # R21-0507-05221, under review.
- [11] D. A. Abraham, *Underwater acoustic signal processing : modeling, detection, and estimation*, Modern acoustics and signal processing. Springer, Cham, Switzerland, 2019.
- [12] R.G. Lyons, *Understanding Digital Signal Processing*, Pearson Education, 2nd edition, 2004.
- [13] J. Oxaal, "Temporal averaging techniques reduce image noise," *EDN*, March 1983.
- [14] H. Seltman, "Approximations for mean and variance of a ratio," <http://www.stat.cmu.edu/hselman/files/ratio.pdf>, accessed on May 2021.
- [15] A. D. Waite, *Sonar for Practicing Engineers*, John Wiley & Sons, Ltd., third edition, 2002.
- [16] S. Blouin, "Simultaneous design of underwater acoustic sensor and communication networks," in *Proceedings of the International Conference on Underwater Networks & Systems, WUWNet, November, 2017*, pp. 16:1–16:5.
- [17] E. A. Silber, C. Binder, S. Blouin, S. Pecknold, G. Heard, and C. Lucas, "Canadian Arctic underwater sentinel experimentation project 2019 gascoyne inlet field trial quicklook report," Tech. Rep., Defence R&D Canada (DRDC), 2020, DRDC-RDDC-2020-R116.

An Approximate Likelihood Ratio Detector for QTMS Radar and Noise Radar

David Luong*, Bhashyam Balaji*[†], and Sreeraman Rajan*

**Department of Systems and Computer Engineering
Carleton University, Ottawa, Canada
{david.luong3, sreeraman.rajan}@carleton.ca*

[†]*Radar Sensing and Exploitation Section
Defence Research and Development Canada, Ottawa, Canada
bhashyam.balaji@drdc-rddc.gc.ca*

Abstract—The most important task of any radar is to detect targets. From this perspective, they are machines for distinguishing between two hypotheses: target absent and target present. The test statistic—or detector function—used by the radar is clearly of primary importance. In this paper, we explore the properties of a detector function for quantum two-mode squeezing radar and noise radar. This detector is derived from a second-order approximation of the likelihood ratio, and is attractive because it has a simple mathematical form. In order to analyze the performance of this detector function, we derive an expression for the receiver operating characteristic curve and verify it via simulations.

Index Terms—Quantum radar, QTMS radar, noise radar, detector function, likelihood ratio

I. INTRODUCTION

Modern radars have the ability to support a vast array of functions: ranging, tracking, landscape imaging, and so on. But the primary goal of radar has always been to detect targets. Therefore, at a fundamental level, all radars are designed to solve a decision problem: is a target present or absent? When the question is put this way, we see that a radar is a physical embodiment of the mathematical problem known as *hypothesis testing*.

The hypothesis testing procedure involves calculating a test statistic and comparing that to a threshold which is chosen to achieve a given significance level. Or, to put it in terms better suited to radar engineering, we calculate a *detector function* and choose the threshold to achieve a desired *probability of false alarm*. Indeed, the gold standard for evaluating radar detection performance is the *receiver operating characteristic* (ROC) curve, which gives the power of the hypothesis test as a function of the significance level—that is, the probability of detection as a function of the probability of false alarm.

When it comes to hypothesis testing, one of the most obvious choices of detector function is the *likelihood ratio* (LR). Indeed, possible applications of the LR to radar detection have been studied for a long time [1], [2]. However, these studies are very general and often give rise to complicated detector functions which are not always suitable for implementation in radars.

In this paper, we focus on the application of the LR to noise radar and its quantum-enhanced version, *quantum two-mode squeezing* (QTMS) radar. This work is particularly timely because of the great interest in quantum radar that was aroused recently by the experiments performed by the Wilson group at the University of Waterloo's Institute for Quantum Computing [3], [4], as well as a follow-up experiment performed by the Fink group at the Institute of Science and Technology Austria [5] which confirmed the leading results of the Wilson group. This interest stems from the fact that QTMS radar is the first type of quantum radar for which laboratory experiments at microwave frequencies have been performed. It is now understood that, from many points of view, QTMS radar is an enhanced version of standard noise radars as described in, e.g., [6]–[10]. QTMS radars, therefore, share many of the same advantages as noise radars, such as their low probability of intercept capabilities and their suitability to spectrum sharing [11]. In particular, the form of the LR is largely the same between the two types of radars, and the results in this paper hold equally for noise radars and QTMS radars.

The main contribution of this paper is an explicit formula for a simplified detector function for QTMS radar and noise radar that is based on a second-order approximation of the LR. Such a simplification is desirable because the exact LR is very complicated. We then derive a formula for the ROC curve of the approximate LR detector. Finally, we make use of simulations to confirm the ROC curve formula and elucidate certain properties of the approximate LR detector.

II. TARGET DETECTION USING QTMS/NOISE RADAR

Every noise radar or QTMS radar has three electromagnetic signals associated with it: the transmitted signal, a reference signal retained within the radar, and the received signal. The reference signal is supposed to be highly correlated with the transmitted signal so that any echo returning from a target is also correlated with the reference signal; this correlation is used to determine whether a target is present or not. Each physical signal can be mathematically described by two real-

valued time series, one representing the in-phase voltages of the signal and the other representing its quadrature voltages.

In QTMS radars and in many noise radars, the two voltage time series of the transmitted and reference signals are Gaussian white noise processes. In this paper, we will assume that all other noise sources inside and outside the radar, such as system noise or atmospheric noise, are modeled as additive white Gaussian noise (AWGN). It follows that the received signal will also be a Gaussian white noise process. We will denote the in-phase and quadrature voltages of the received signal by I_1 and Q_1 , respectively, and the in-phase and quadrature voltages of the reference signal by I_2 and Q_2 . (For the purposes of this paper, we do not need to introduce notation for the transmitted signal.) In accordance with the above discussion, each of these variables is a Gaussian white noise process. The four voltage signals are fully characterized by stating that they all have zero mean and by giving the 4×4 covariance matrix $E[xx^T]$, where $x = [I_1, Q_1, I_2, Q_2]^T$. In [12], we showed that the covariance matrix $E[xx^T]$ for noise radar signals can be written in block matrix form as

$$\Sigma = \begin{bmatrix} \sigma_1^2 \mathbf{I}_2 & \rho \sigma_1 \sigma_2 \mathbf{R}(\phi) \\ \rho \sigma_1 \sigma_2 \mathbf{R}(\phi)^T & \sigma_2^2 \mathbf{I}_2 \end{bmatrix} \quad (1)$$

where σ_1^2 and σ_2^2 are the received and reference signal powers, respectively, while ϕ is the phase shift between the signals, \mathbf{I}_2 is the 2×2 identity matrix, and $\mathbf{R}(\phi)$ is the rotation matrix

$$\mathbf{R}(\phi) = \begin{bmatrix} \cos \phi & \sin \phi \\ -\sin \phi & \cos \phi \end{bmatrix}. \quad (2)$$

The covariance matrix for QTMS radar signals has the same block matrix form, but with a reflection matrix instead of a rotation matrix:

$$\mathbf{R}'(\phi) = \begin{bmatrix} \cos \phi & \sin \phi \\ \sin \phi & -\cos \phi \end{bmatrix}. \quad (3)$$

We may assume without loss of generality that $\rho \geq 0$, because its sign can always be absorbed into $\mathbf{R}(\phi)$ or $\mathbf{R}'(\phi)$.

For the purposes of target detection, it is unimportant whether $\mathbf{R}(\phi)$ or $\mathbf{R}'(\phi)$ is used. To maintain conformity with our previous work on QTMS radar, we will use $\mathbf{R}'(\phi)$.

The correlation coefficient ρ characterizes the magnitude of the off-diagonal elements of (1) and is intimately related to the target detection problem. When the target is present, $\rho > 0$ because the received signal will be correlated (possibly quite weakly) with the transmitted signal, which is itself highly correlated with the reference signal. On the other hand, when the target is absent, $\rho = 0$ because the received signal is purely background noise. Distinguishing between the presence or absence of a target, therefore, amounts to testing the following hypotheses:

$$\begin{aligned} H_0 : \rho = 0 & \quad \text{Target absent} \\ H_1 : \rho > 0 & \quad \text{Target present} \end{aligned} \quad (4)$$

The source of the ‘‘quantum advantage’’ of QTMS radars over standard noise radars has been discussed in previous papers [11]–[13], but in the context of (4) we may state the case as follows: when a target is present, QTMS radars can achieve

higher values of ρ . Naturally, when there is no target, $\rho = 0$ for both QTMS radars and noise radars. Therefore, QTMS radars achieve a greater contrast between the two hypotheses, making them easier to distinguish.

In order to perform this hypothesis test, we must decide on a test statistic—or detector function—that allows us to distinguish between the two cases. Such a detector would be a generalization of matched filtering to the case where there are more than two time series. Previous work has focused on performing matched filtering between the complex signals $I_1 + jQ_1$ and $I_2 + jQ_2$ [6], or on estimating ρ directly [14]. In this work, we use the likelihood ratio test to derive a detector function.

There are three other parameters in (1) to deal with: σ_1 , σ_2 , and ϕ . None of these play a direct role in distinguishing whether or not there is a target. To keep our problem simple, we assume $\sigma_1 = \sigma_2 = 1$ and $\phi = 0$, and show via simulations that our results hold even when the assumption $\sigma_1 = \sigma_2 = 1$ is violated. Our assumption that $\phi = 0$ is in line with previous work [4], [5], [13], though in future work we will generalize to the case of unknown ϕ .

Under these assumptions, the mathematical problem that we need to solve may be stated as follows: let N independent samples be drawn from a multivariate normal distribution with zero mean and covariance matrix

$$\Sigma(\rho) = \begin{bmatrix} 1 & 0 & \rho & 0 \\ 0 & 1 & 0 & -\rho \\ \rho & 0 & 1 & 0 \\ 0 & -\rho & 0 & 1 \end{bmatrix}. \quad (5)$$

Given these samples, use a likelihood ratio test to decide whether $\rho = 0$ or $\rho > 0$. (Note that N is the number of integrated samples and is proportional to the integration time.) This matrix is for QTMS radars, but our results apply to standard noise radars as well if the negative signs are changed into positive signs.

III. AN APPROXIMATE LR DETECTOR

In this section, we derive an explicit formula for the LR detector function under the assumption that $\rho \ll 1$. This assumption simplifies the problem greatly, and it also has an important physical interpretation: ρ is small when the target is far away or when the radar cross section is small [15]. The case of large ρ is much less interesting because the target is then very easy to see, so the probability of detection will always be very high and there is no need to use a complicated detector. For this reason, we will focus on the case where $\rho \ll 1$.

For N samples drawn independently from a zero-mean multivariate Gaussian distribution with covariance matrix (5), the log-likelihood can be shown to be

$$\ell(\rho) = -\frac{N}{2} \left[\frac{\bar{P}_{\text{tot}} - 2\bar{D}_1\rho}{1 - \rho^2} + 2\ln(1 - \rho^2) + 2\ln(2\pi) \right] \quad (6)$$

where

$$P_{\text{tot}} \equiv I_1^2 + Q_1^2 + I_2^2 + Q_2^2 \quad (7)$$

$$D_1 \equiv I_1 I_2 - Q_1 Q_2 \quad (8)$$

and a line over any expression indicates the sample mean. For example,

$$\overline{I_1 I_2} = \frac{1}{N} \sum_{n=1}^N \overline{i_1^{(n)} i_2^{(n)}} \quad (9)$$

where $i_1^{(n)}$ and $i_2^{(n)}$ denote the n th samples of I_1 and I_2 , respectively.

The reason for naming the sum in (7) as P_{tot} is because it corresponds to the total power in the in-phase and quadrature components of both the received signal and the reference signal. The quantity D_1 appeared in [4] under the name ‘‘Detector 1’’ and was studied in a previous publication [13]. (Interestingly, the ‘‘digital receiver’’ in [5] can be shown to be an implementation of D_1 .)

Under the small- ρ assumption, we may expand (6) in powers of ρ , keeping terms up to second order:

$$\ell(\rho) \approx -\frac{N}{2} [\bar{P}_{\text{tot}} + 2 \ln(2\pi) - 2\bar{D}_1 \rho + (\bar{P}_{\text{tot}} - 2)\rho^2]. \quad (10)$$

The LR detector is defined as

$$\begin{aligned} \hat{D}_{\text{LR}} &\equiv -2[\ell(0) - \ell(\hat{\rho})] \\ &\approx N\hat{\rho}[2\bar{D}_1 - (\bar{P}_{\text{tot}} - 2)\hat{\rho}] \end{aligned} \quad (11)$$

where $\hat{\rho}$ is the maximum likelihood estimate of ρ . The factor of -2 allows us to exploit Wilks’ theorem later, when we derive an expression for the ROC curve.

It is a simple matter to maximize (10), leading to the following approximation of the maximum likelihood estimate $\hat{\rho}$:

$$\hat{\rho} \approx \frac{\bar{D}_1}{\bar{P}_{\text{tot}} - 2}. \quad (12)$$

By substituting this into (11), we obtain the desired LR detector function:

$$\hat{D}_{\text{LR}} \approx \frac{N\bar{D}_1^2}{\bar{P}_{\text{tot}} - 2}. \quad (13)$$

This equation is an approximation of the exact LR detector, correct up to second order in ρ .

IV. ROC CURVE FOR THE APPROXIMATE LR DETECTOR

In this section, we give an expression for the ROC curve of the approximate LR detector. This expression is valid when ρ is small and N is large. Our argument is a heuristic one, but we will show via simulations that our expression is a very good approximation.

Our approach is to determine the distribution of the detector function (13) under the two hypotheses (4). Once this is known, it is a simple matter to write down the ROC curve in terms of their cumulative density functions.

Under the null hypothesis $\rho = 0$, we may immediately state that the distribution of the LR detector in the limit $N \rightarrow \infty$ is $\hat{D}_{\text{LR}} \sim \chi_1^2$, a chi-square distribution with one degree of freedom. This is a consequence of Wilks’ theorem [16], applied to the case that the difference in the number of free parameters between H_1 and H_0 is 1.

We will now argue that, when $\rho > 0$, the approximate LR detector is approximately a noncentral chi-square random

variable. First, note that in the limit $N \rightarrow \infty$, the Central Limit Theorem can be invoked to show that \bar{D}_1 is normally distributed with mean 2ρ and variance $2(1 + \rho^2)/N$ [13]. Since the approximate LR detector was derived under the assumption that $\rho \ll 1$, we may simplify this and state that $\bar{D}_1 \sim \mathcal{N}(2\rho, 2/N)$. To simplify the analysis, we introduce the transformed random variable

$$X \equiv \sqrt{\frac{N}{2}} \bar{D}_1. \quad (14)$$

It follows from the properties of the normal distribution that $X \sim \mathcal{N}(\rho\sqrt{2N}, 1)$.

We will now show that $\bar{P}_{\text{tot}} \sim \mathcal{N}(4, 8/N)$. From (7) and (5), we immediately see that $E[P_{\text{tot}}] = 4$. Because \bar{P}_{tot} is the sample mean of N independent copies of P_{tot} , $E[\bar{P}_{\text{tot}}] = 4$. The variance of P_{tot} is more difficult to calculate, but using equation (13) of [17] (or a computer algebra system such as Mathematica), it can be shown that $\text{var}[P_{\text{tot}}] = 8(1 + \rho^2)$. Because $\rho \ll 1$, we may drop the ρ^2 term. We now know the mean and variance of P_{tot} , so we may once again invoke the Central Limit Theorem to state that $\bar{P}_{\text{tot}} \sim \mathcal{N}(4, 8/N)$ when N is large.

Now, let us define

$$Y \equiv \frac{\bar{P}_{\text{tot}}}{2} - 1. \quad (15)$$

It is evident that $Y \sim \mathcal{N}(1, 2/N)$ and that

$$\hat{D}_{\text{LR}} \approx \frac{N\bar{D}_1^2}{\bar{P}_{\text{tot}} - 2} = \frac{X^2}{Y}. \quad (16)$$

In the limit $N \rightarrow \infty$, the distribution of Y narrows and becomes concentrated near 1. It follows that, with very high probability, $\hat{D}_{\text{LR}} \approx X^2$. But X^2 is the square of a normally distributed random variable with variance 1, so we may identify it as a noncentral chi-square random variable with one degree of freedom and noncentrality parameter $(\rho\sqrt{2N})^2$. We have thus arrived at the desired result: when N is large and ρ is small,

$$\hat{D}_{\text{LR}} \sim \chi_1^2(2N\rho^2). \quad (17)$$

As expected, this reduces to the central chi-square distribution predicted by Wilks’ theorem when $\rho = 0$.

It is now a simple matter to calculate the ROC curve for the approximate LR detector. Since the cumulative density function for the χ_1^2 distribution is the regularized gamma function $P(1/2, x/2)$, the probability of false alarm for a given threshold T is given by the corresponding survival function $S(T)$, where

$$p_{\text{FA}} = S(T) = 1 - P\left(\frac{1}{2}, \frac{T}{2}\right). \quad (18)$$

Similarly, the survival function for the $\chi_1^2(2N\rho^2)$ distribution is given by the Marcum Q-function $Q_{1/2}(\rho\sqrt{2N}, \sqrt{x})$. Therefore, the ROC curve is given by

$$p_{\text{D}}(p_{\text{FA}}) = Q_{\frac{1}{2}}\left[\rho\sqrt{2N}, \sqrt{S^{-1}(p_{\text{FA}})}\right]. \quad (19)$$

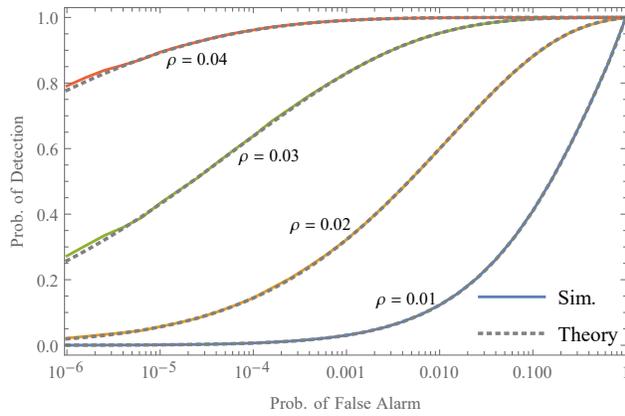


Fig. 1. Comparison of simulated ROC curves for the approximate LR detector with theoretical ROC curves calculated from (19) for $N = 10000$ and varying values of ρ .

V. SIMULATIONS

In this section, we present the results of simulating radar detection data and calculating the approximate LR detector from this data. Naively, this could be done by drawing N random vectors from a multivariate normal distribution. However, an inspection of (6), (7), and (8) shows that the LR detector depends on the detection data only through the sample covariance matrix

$$\bar{\mathbf{S}} = \frac{1}{N} \sum_{n=1}^N x_n x_n^T \quad (20)$$

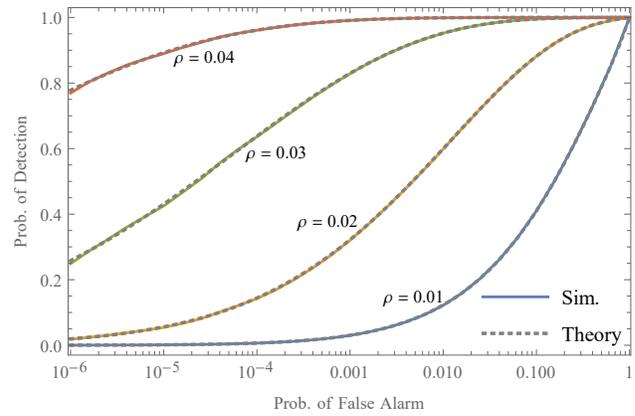
where x_n is the n th sample vector. Therefore, to save computational time, we use the approach outlined in [14]. Instead of drawing random vectors, we draw a single random matrix from the Wishart distribution $W_4(\Sigma, N)$ and normalize the result by N .

Our first simulation results are shown in Fig. 1. In this figure, we confirm the accuracy of the ROC curve formula (19) by comparing them to ROC curves obtained by simulation. It is clear that, although several approximations were made in deriving (19), the formula matches the simulation results extremely well.

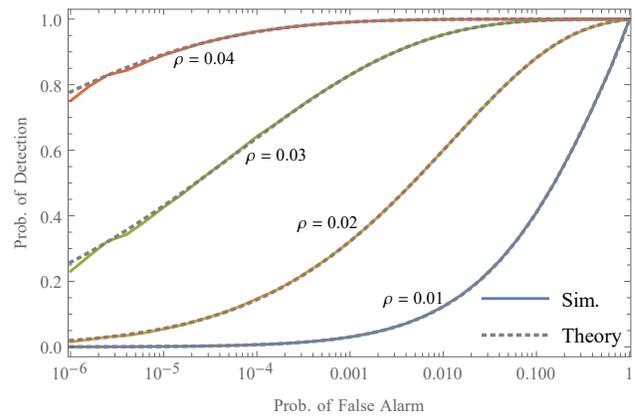
We stated in Sec. II that our approximate LR detector works even when the assumption $\sigma_1 = \sigma_2 = 1$ is violated. This is shown in Fig. 2, where we simulated ROC curves for the approximate LR detector when $(\sigma_1, \sigma_2) = (0.1, 10)$ and $(0.01, 10000)$. These simulated ROC curves match up extremely well with (19). Therefore, we may conclude that the approximate LR detector works even when the assumptions on σ_1 and σ_2 are strongly violated.

VI. CONCLUSION

In this paper, we have derived a detector function for QTMS and noise radar based on a second-order approximation of the likelihood ratio test statistic. This detector function is attractive because it has a simple mathematical form and does not require



(a)



(b)

Fig. 2. Comparison of simulated and theoretical ROC curves for the approximate LR detector when (σ_1, σ_2) equals (a) $(0.1, 10)$ and (b) $(0.01, 10000)$. In all cases, $N = 10000$.

large computational resources. We were able to find an explicit expression for the ROC curve in the case where the number of integrated samples are large. This expression was validated using simulations of radar detection data.

In deriving our approximate LR detector, we assumed $\sigma_1 = \sigma_2 = 1$ and $\phi = 0$. We showed via simulations that the detector is still viable when the assumption $\sigma_1 = \sigma_2 = 1$ is violated. In future work, we will generalize our LR detector to the case where ϕ is unknown.

REFERENCES

- [1] E. J. Kelly, I. S. Reed, and W. L. Root, "The detection of radar echoes in noise. I," *Journal of the Society for Industrial and Applied Mathematics*, vol. 8, no. 2, pp. 309–341, Jun. 1960.
- [2] O. Besson, A. Coluccia, E. Chaumette, G. Ricci, and F. Vincent, "Generalized likelihood ratio test for detection of Gaussian rank-one signals in Gaussian noise with unknown statistics," *IEEE Transactions on Signal Processing*, vol. 65, no. 4, pp. 1082–1092, Feb. 2017.
- [3] C. W. S. Chang, A. M. Vadiraj, J. Bourassa, B. Balaji, and C. M. Wilson, "Quantum-enhanced noise radar," *Applied Physics Letters*, vol. 114, no. 11, p. 112601, Mar. 2019.

- [4] D. Luong, C. W. S. Chang, A. M. Vadiraj, A. Damini, C. M. Wilson, and B. Balaji, "Receiver operating characteristics for a prototype quantum two-mode squeezing radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 56, no. 3, pp. 2041–2060, Jun. 2020.
- [5] S. Barzanjeh, S. Pirandola, D. Vitali, and J. M. Fink, "Microwave quantum illumination using a digital receiver," *Science Advances*, vol. 6, no. 19, p. eabb0451, May 2020.
- [6] M. Dawood and R. M. Narayanan, "Receiver operating characteristics for the coherent UWB random noise radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 37, no. 2, pp. 586–594, Apr. 2001.
- [7] Y. Zhang and R. M. Narayanan, "Design considerations for a real-time random-noise tracking radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 40, no. 2, pp. 434–445, Apr. 2004.
- [8] K. Kulpa, *Signal Processing in Noise Waveform Radar*, ser. Radar. Artech House, 2013.
- [9] K. Savci, A. G. Stove, A. Y. Erdogan, G. Galati, K. A. Lukin, G. Pavan, and C. Wasserzler, "Trials of a noise-modulated radar demonstrator—first results in a marine environment," in *Proceedings of the 20th International Radar Symposium (IRS)*. IEEE, Jun. 2019.
- [10] K. Savci, A. G. Stove, F. D. Palo, A. Y. Erdogan, G. Galati, K. A. Lukin, S. Lukin, P. Marques, G. Pavan, and C. Wasserzler, "Noise radar—overview and recent developments," *IEEE Aerospace and Electronic Systems Magazine*, vol. 35, no. 9, pp. 8–20, Sep. 2020.
- [11] D. Luong, S. Rajan, and B. Balaji, "Entanglement-based quantum radar: From myth to reality," *IEEE Aerospace and Electronics Systems Magazine*, vol. 35, no. 4, pp. 22–35, Apr. 2020.
- [12] D. Luong and B. Balaji, "Quantum two-mode squeezing radar and noise radar: covariance matrices for signal processing," *IET Radar, Sonar & Navigation*, vol. 14, no. 1, pp. 97–104, Jan. 2020.
- [13] D. Luong, B. Balaji, and S. Rajan, "Simulation study of a detector function for QTMS radar and noise radar," in *Proceedings of the 2020 IEEE Radar Conference (RadarConf)*, Sep. 2020, pp. 1–5.
- [14] D. Luong, S. Rajan, and B. Balaji, "Quantum two-mode squeezing radar and noise radar: Correlation coefficients for target detection," *IEEE Sensors Journal*, vol. 20, no. 10, pp. 5221–5228, May 2020.
- [15] D. Luong, B. Balaji, and S. Rajan, "Performance prediction for coherent noise radars using the correlation coefficient," *arXiv:2009.10271; submitted to IEEE Transactions on Aerospace and Electronic Systems*, Sep. 2020.
- [16] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *The Annals of Mathematical Statistics*, vol. 9, no. 1, pp. 60–62, Mar. 1938.
- [17] G. W. Bohmstedt and A. S. Goldberger, "On the exact covariance of products of random variables," *Journal of the American Statistical Association*, vol. 64, no. 328, pp. 1439–1442, Dec. 1969.

Detection of Weak Transient Signals Using a Broadband Subspace Approach

Stephan Weiss¹, Connor Delaosa¹, James Matthews², Ian K. Proudler¹, Ben A. Jackson³

¹ Department of Electronic & Electrical Engineering, University of Strathclyde, Glasgow G1 1XW, Scotland

² PA Consulting, Global Innovation and Technology Centre Cambridge, Melbourn SG8 6DP, UK

³ UK Defence Science & Technology Laboratory (Dstl), Porton Down, Salisbury SP4 0JQ, UK

{stephan.weiss,connor.delaosa,ian.proudler}@strath.ac.uk, james.matthews@paconsulting.com, bajackson@dstl.gov.uk

Abstract—We investigate the detection of broadband weak transient signals by monitoring a projection of the measurement data onto the noise-only subspace derived from the stationary sources. This projection utilises a broadband subspace decomposition of the data’s space-time covariance matrix. The energy in this projected ‘syndrome’ vector is more discriminative towards the presence or absence of a transient signal than the original data, and can be enhanced by temporal averaging. We investigate the statistics, and indicate in simulations how discrimination can be traded off with the time to reach a decision, as well as with the sample size over which the space-time covariance is estimated.

I. INTRODUCTION

In a number of problems, it is paramount to detect the emergence of a signal whose power may be significantly lower than other sources that are already active in the environment. This is the case e.g. in a cognitive radio scenario [1], where secondary users may be utilising a specific frequency band, but where the arrival of a distant and therefore quiet primary user must be detected in order to instigate the secondary users to vacate this part of the spectrum.

The detection of transient signals is generally based on energy criteria, and can involve a fixed transform such as a wavelet or short-time Fourier transform-type operation to reveal particular patterns of the transient source [2]–[4]. Particularly when multiple measurements are available, data-dependent transforms exploiting the eigenvalue decomposition (EVD) of the covariance matrix of the data can generally attain an optimum compaction of energy into a lower-dimensional subspace in the sense of the Karhunen-Loeve transform [5]; therefore subspace-based methods have emerged that exploit particular partitioning of the space spanned by eigenvectors of the EVD [6]–[10]. While this work has peaked two decades earlier, developments of energy-based subspace detectors are still afoot [11].

The above methods [2]–[4], [6]–[11] operate on narrowband data and calculate an instantaneous covariance matrix that will only capture phase shifts between elements of the data vector. To address the detection of broadband transient signals, it is possible to operate with tapped delay lines or in frequency bins created by a discrete Fourier transform (DFT), where problems

can be treated as narrowband ones. However, when addressing the problems in different bins independently, e.g. for the purpose of parallelisation, generally the coherence between bins is lost, leading to generally suboptimal solutions [12].

This paper investigates the detection of broadband weak transient sources via broadband subspace-based methods [13], which are afforded via a space-time covariance matrix that takes both spatial and temporal correlations in the data into account. Similar to narrowband subspace methods, a diagonalisation of this space-time covariance is required. For the broadband problem, we are looking towards polynomial EVD methods that can decouple the space-time covariance for every lag value [14] — such decompositions have been shown to exist in most case [15], [16] and a number of algorithms have been developed to solve this diagonalisation often with guaranteed convergence [14], [17]–[21].

In the following, we review the broadband signal scenario and the description of the data’s second order statistics by a space-time covariance in Sec. II. The broadband subspace decomposition based on the space-time covariance is outlined in Sec. III, leading to the proposed broadband subspace detector in Sec. IV, including an investigation of the subspace energy. The temporal correlation of the latter prohibits the direct description of the subspace energy, but a decorrelation via a decimated processor allows us to exploit some known statistical results for generalised χ^2 distributions [22]–[26], and to subsequently define metrics such as the discrimination and decision time in Sec. V. Following some numerical simulations, Sec. VI draws conclusions.

II. WIDEBAND SIGNAL MODEL

A. Source Model

We assume that M sensors acquire a measurement vector $\mathbf{x}[n] \in \mathbb{C}^M$ over discrete time $n \in \mathbb{Z}$, consisting of time series $\mathbf{x}[n] = [x_1[n], \dots, x_M[n]]^T \in \mathbb{C}^M$. This sensor array is illuminated by $L < M$ sources $s_\ell[n]$ $\ell = 1, \dots, L$ via transfer paths with impulse responses $a_{m,\ell}[n]$, $m = 1, \dots, M$, $\ell = 1, \dots, L$, whereby $a_{m,\ell}[n]$ describes the impulse response of the path from the ℓ th source to the m th sensor. This path could be a simple delay in the case of a free-space propagation environment, but can equally describe dispersive, i.e. multipath, channels. For the contribution of the ℓ th source, this scenario is illustrated in Fig. 1; the source signal $s_\ell[n]$,

This work was supported by Dstl via the DASA Invisible Shield programme and by the Engineering and Physical Sciences Research Council (EPSRC) Grant number EP/S000631/1 and the MOD University Defence Research Collaboration in Signal Processing.

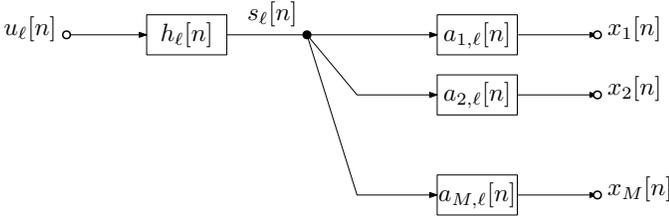


Fig. 1. Source model for a signal $s_\ell[n]$, $\ell = 1, \dots, L$, contributing to the sensor data. The source is characterised by an innovation filter $h_\ell[n]$ excited by zero-mean unit-variance uncorrelated Gaussian noise $u_\ell[n]$ that generates its desired PSD [27], and a broadband steering vector consisting of transfer functions.

$\ell = 1, \dots, L$, is generated by an innovation filter $h_\ell[n]$ excited by zero-mean unit-variance uncorrelated Gaussian noise $u_\ell[n]$ [27].

If we place the M transfer paths for the ℓ th source into a vector

$$\mathbf{a}_\ell[n] = \begin{bmatrix} a_{1,\ell}[n] \\ \vdots \\ a_{M,\ell}[n] \end{bmatrix}, \quad (1)$$

then the data can be modelled as

$$\mathbf{x}[n] = \sum_{\ell=1}^L \mathbf{a}_\ell[n] * s_\ell[n] + \mathbf{v}[n], \quad (2)$$

with $\mathbf{v}[n]$ being additive white zero mean uncorrelated Gaussian noise, and $*$ denoting the convolution operator. In particular, we assume that all source signals $s_\ell[n]$ as well as the noise in $\mathbf{v}[n]$ are mutually independent. This scenario is depicted in Fig. 1 for the case of a single source in a noise-free environment. For L sources, the data vector $\mathbf{x}[n]$ is obtained by superposition of L models as in Fig. 1.

B. Space-Time Covariance Matrix

To define problems based on cost functions such as the mean square error, we require access to the second order statistics of the data in $\mathbf{x}[n]$. This is captured by the space-time covariance matrix $\mathbf{R}[\tau] = \mathcal{E}\{\mathbf{x}[n]\mathbf{x}^H[n-\tau]\}$, where $\mathcal{E}\{\cdot\}$ is the expectation operator, $\{\cdot\}^H$ a Hermitian transposition, and $\tau \in \mathbb{Z}$ a lag parameter. This matrix contains auto- and cross-correlation sequences, and satisfies the symmetry $\mathbf{R}[\tau] = \mathbf{R}^H[-\tau]$. The z -transform $\mathbf{R}(z) = \sum_{\tau} \mathbf{R}[\tau]z^{-\tau}$, or in short $\mathbf{R}(z) \bullet \circ \mathbf{R}[\tau]$, is known as the cross-spectral density (CSD) matrix and satisfies the parahermitian property, s.t. $\mathbf{R}^P(z) = \mathbf{R}^H[1/z^*] = \mathbf{R}(z)$, whereby $\mathbf{R}(z)$ is identical to its parahermitian transpose $\mathbf{R}^P(z)$, with $\{\cdot\}^P$ the parahermitian transpose operator [28].

From (2) and $H_\ell(z) \bullet \circ h_\ell[n]$, and with $\gamma_\ell(z) = H_\ell(z)H_\ell^P(z)$, we can also express the CSD matrix $\mathbf{R}(z)$ as the expansion

$$\mathbf{R}(z) = \sum_{\ell=1}^L \mathbf{a}_\ell(z)\mathbf{a}_\ell^P(z)\gamma_\ell(z) + \sigma_v^2\mathbf{I}_M. \quad (3)$$

In (3), a vector $\mathbf{a}_\ell(z)$ is referred to as the broadband steering vector of the ℓ th source. A steering vector is used in the beamforming terminology to describe the signature of the source arriving from a specific direction. In the simplest case, $\mathbf{a}_\ell(z)$ can be a vector of fractional delays [29], [30], but can also be a vector of general, rational transfer functions.

III. BROADBAND SUBSPACE DECOMPOSITION

A. Parahermitian Matrix Eigenvalue Decomposition

If $\mathbf{R}(z)$ arises from a source model such as in Fig. 1 with stable and causal filters $h_\ell[n]$, and transfer paths $a_{m,\ell}[n]$, $\ell = 1, \dots, L$, $m = 1, \dots, M$, then $\mathbf{R}(z)$ is analytic. As a result, the parahermitian matrix $\mathbf{R}(z)$ admits a parahermitian matrix EVD (PhEVD) [15]

$$\mathbf{R}(z) = \mathbf{Q}(z)\mathbf{\Lambda}(z)\mathbf{Q}^P(z) \quad (4)$$

$$= \sum_{m=1}^M \mathbf{q}_m(z)\mathbf{q}_m^P(z)\lambda_m(z), \quad (5)$$

provided that the data vector $\mathbf{x}[n]$ is unmultiplexed [16]. In (4), $\mathbf{Q}(z)$ contains in its columns the eigenvectors $\mathbf{q}_m(z)$, $m = 1, \dots, M$ of (5). It is a paraunitary matrix such that

$$\mathbf{Q}(z)\mathbf{Q}^P(z) = \mathbf{Q}^P(z)\mathbf{Q}(z) = \mathbf{I}_M, \quad (6)$$

representing a lossless filter bank [28]. The matrix $\mathbf{\Lambda}(z)$ is diagonal and parahermitian, containing the eigenvalues

$$\mathbf{\Lambda}(z) = \text{diag}\{\lambda_1(z), \lambda_2(z), \dots, \lambda_M(z)\}. \quad (7)$$

Both $\mathbf{Q}(z)$ and $\mathbf{\Lambda}(z)$ are potentially transcendental but analytic functions, i.e. their time domain equivalents converge at least exponentially and can be well-approximated by Laurent polynomials.

B. Uniqueness and Ambiguity

Assuming that the M eigenvalues of $\mathbf{R}(z)$ are distinct and only intersect in a finite number of points, then there is only one solution for the functions $\lambda_m(z)$, $m = 1, \dots, M$ apart from a permutation. The latter can be addressed by ordering the eigenvalues according to their power, similar how in an ordered EVD eigenvalues are arranged in descending order [31].

For distinct eigenvalues, the associated eigenvectors each exist in uniquely defined 1-d subspaces, but can be multiplied by arbitrary allpass functions. Therefore, w.r.t. the decomposition in (4), the factor $\mathbf{\Lambda}(z)$ is unique, but there is an ambiguity w.r.t. the paraunitary matrix holding the eigenvectors in its columns: if $\mathbf{Q}(z)$ contains valid eigenvectors, then so does $\mathbf{Q}(z)\mathbf{\Phi}(z)$, where $\mathbf{\Phi}(z)$ is a diagonal matrix containing arbitrary allpass filters.

C. Broadband Subspace Partitioning

The PhEVD in (4) admits a subspace decomposition,

$$\mathbf{R}(z) = [\mathbf{U}(z) \mathbf{U}_\perp(z)] \begin{bmatrix} \mathbf{\Lambda}_s(z) & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_s(z) \end{bmatrix} \begin{bmatrix} \mathbf{U}^P(z) \\ \mathbf{U}_\perp^P(z) \end{bmatrix}, \quad (8)$$

where the eigenvalues are grouped into two submatrices $\Lambda_s(z) : \mathbb{C} \rightarrow \mathbb{C}^{R \times R}$ and $\Lambda_{\bar{s}}(z) : \mathbb{C} \rightarrow \mathbb{C}^{(M-R) \times (M-R)}$. The eigenvectors associated with the eigenvalues in $\Lambda_s(z)$ are the columns of the matrix $\mathbf{U}(z) : \mathbb{C} \rightarrow \mathbb{C}^{M \times R}$, which therefore span the subspace within which any components related to these eigenvalues reside. This subspace is orthogonal to its complement, spanned by the columns of $\mathbf{U}_{\perp}(z) : \mathbb{C} \rightarrow \mathbb{C}^{M \times (M-R)}$.

A common partitioning of $\mathbf{Q}(z)$ in (5) in both the narrow-band [31] or the broadband cases [13] is a subspace decomposition that defines signal-plus-noise and noise-only subspaces, where $\Lambda_s(z)$ contains the R signal-related, principal eigenvalues, and $\mathbf{U}(z)$ all associated eigenvector components. In contrast, the eigenvalues $\Lambda_{\bar{s}}(z)$ define the noise floor, and the columns of $\mathbf{U}_{\perp}(z)$ span the associated noise-only subspace.

IV. COMPLEMENTARY BROADBAND MATCHED SUBSPACE DETECTOR

A. Approach

We assume that a number of L sources have been stationary for a period of time, over which a space-time covariance matrix $\hat{\mathbf{R}}[\tau]$ has been estimated, using e.g. the procedures outlined in [32], [33] for the estimation and the optimum support length of this estimate. Using an approximation of the PhEVD in (4) by algorithms of the second order sequential best rotation (SBR2) [14], [17] or sequential matrix diagonalisation (SMD) [18], [34] families to factorise $\hat{\mathbf{R}}(z) \rightsquigarrow \mathbf{R}[\tau]$, we establish the broadband signal-plus-noise subspace spanned by the columns of $\mathbf{U}(z)$ and its complement, the noise-only subspace, spanned by the columns of $\mathbf{U}_{\perp}(z)$, as defined in (8) with $R = L$.

If a new source enters the scene, then some of its components may protrude into the noise-only subspace, where a step change in energy can be detected more easily than directly from the measurement $\mathbf{x}[n]$, since the energy contribution of the stationary sources will be removed. We therefore calculate a type of syndrome vector $\mathbf{y}[n] \in \mathbb{C}^{M-L}$ based on $\mathbf{U}_{\perp}[n] \circ \bullet \mathbf{U}_{\perp}(z)$,

$$\mathbf{y}[n] = \sum_{\nu} \mathbf{U}_{\perp}^H[-\nu] \mathbf{x}[n - \nu]. \quad (9)$$

This is a type of projection onto a reduced $(M-L)$ -dimensional space, and yields the same energy in $\mathcal{E}\{\|\mathbf{y}[n]\|^2\}$ as a projection using the projection operator $\mathbf{P}(z) = \mathbf{U}_{\perp}(z)\mathbf{U}_{\perp}^P(z)$, which can be shown via a polynomial singular value decomposition of $\mathbf{U}_{\perp}^P(z)$ [14]. Therefore, the energy of the syndrome $\mathbf{y}[n]$ can be used in a hypothesis test to detect the absence or presence of a transient signal.

B. Signal Statistics

In the absence of a transient signal, let the CSD matrix of $\mathbf{x}[n]$ be $\mathbf{R}(z)$ as in (3). Therefore the CSD of the syndrome vector is

$$\mathbf{R}_y(z) = \sigma_v^2 \mathbf{I}_{M-L}. \quad (10)$$

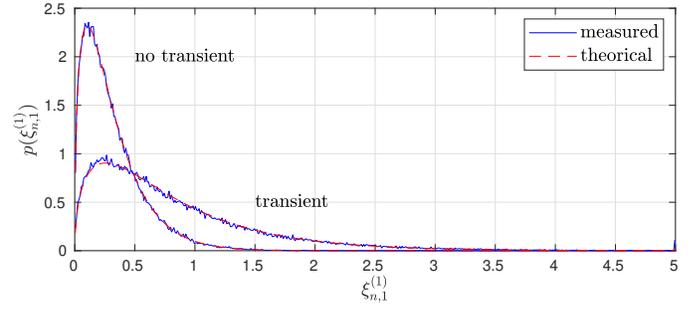


Fig. 2. Example of measured and theoretical generalised χ^2 distributions of the energy $\|\mathbf{y}[n]\|_2^2$ in the absence and presence of a transient signal.

If $\mathbf{x}'[n]$ is a modified data vector that in addition to the stationary sources captured in $\mathbf{x}[n]$ includes a transient signal with PSD $\gamma'(z)$ and steering vector $\mathbf{a}'(z)$, then a modified CSD matrix $\mathbf{R}'(z)$ arises

$$\mathbf{R}'(z) = \mathbf{R}(z) + \mathbf{a}'(z)\gamma'(z)\mathbf{a}'^P(z), \quad (11)$$

and the modified syndrome has the CSD matrix

$$\mathbf{R}'_y(z) = \mathbf{U}_{\perp}(z)\mathbf{a}'(z)\gamma'(z)\mathbf{a}'^P(z)\mathbf{U}_{\perp}^P(z) + \sigma_v^2 \mathbf{I}_{M-L}. \quad (12)$$

The first term on the r.h.s. of (12) creates the offset in power that allows us to potentially detect the presence of a transient signal.

To examine the statistics of $\|\mathbf{y}[n]\|_2^2$, we assume that the noise components of $\mathbf{x}[n]$ are zero-mean and identically distributed Gaussian random variables. However, in the presence of transient components, or subspace leakage of the stationary sources, the elements of the syndrome vector will have Gaussian distributions with different variances. Therefore, $\|\mathbf{y}[n]\|_2^2$ will adhere to a generalised χ^2 distribution [24], [25]. For an example with $M = 6$, $L = 3$, and broadband steering vectors of order 10 drawn from uncorrelated complex Gaussian distributions, Fig. 2 shows the measured distributions in the presence and absence of a transient component in comparison to the theoretical values of a generalised χ^2 distribution based on the implementation in [26].

The discrimination between the cases of a transient source being present or absent can be increased through temporal averaging of energy terms $\|\mathbf{y}[n]\|_2^2$. However, the generalised χ^2 distribution assumes a summation over squared Gaussian distributed, uncorrelated variables. From (12), it is clear that the terms contributed by the transient source will be correlated. Therefore, direct temporal averaging leads to a distribution that will no longer be covered by a generalised χ^2 distribution.

C. Decimated Subspace Detector

In order to deal with large data volumes acquired at a high sampling rate, and to concurrently temporally decorrelate successive syndrome vectors in case of temporal averaging, it is possible to sum over decimated syndrome vectors, such that

$$\xi_{n,D}^{(K)} = \frac{1}{K} \sum_{\mu=0}^{K-1} \|\mathbf{y}[n - D\mu]\|_2^2 \quad (13)$$

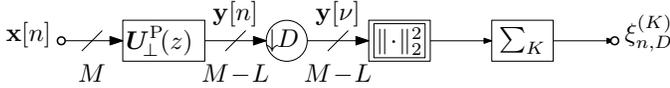


Fig. 3. Flow graph of decimated averaging of syndrome energy.

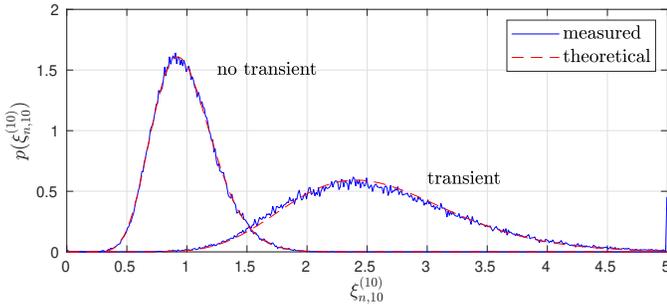


Fig. 4. Example of measured and theoretical generalised χ^2 distributions of the energy $\|y[nu]\|_2^2$ in the absence and presence of a transient signal.

can be used as a criterion for a hypothesis test. The processor is outlined by the flow graph in Fig. 3, where the syndrome energy is decimated by a factor D prior to averaging by a sliding window of length K . It is possible to change the order in which the decimator and the processor matrix $U_{\perp}(z)$ appear in Fig. 3 by means of noble identities [28], thus achieving a computationally efficient polyphase implementation of the arrangement in Fig. 3.

The decorrelation, which may be achieved through a decimation by D of the polynomial order of the processor matrix $U_{\perp}(z)$, ensures that $\xi_{n,D}^{(K)}$ follows a generalised χ^2 distribution, which permits theoretical predictions for a hypothesis test on $\xi_{n,D}^{(K)}$. As an example, to highlight both the fit of the distribution but also the increase in discrimination by averaging over a number of syndrome energies, Fig. 4 shows the scenario of Fig. 2 for the case of averaging over $K = 10$ with a decimation by $D = 10$.

V. SIMULATION AND RESULTS

A. Performance Metrics

We define two metrics for the detection of a transient signal. The first is the discrimination of the variable $\xi_{n,D}^{(K)}$ in (13), which can be assessed independently of a specific threshold e.g. by measuring the area under a receiver operating characteristic (ROC) derived from the type of distribution plots in Figs. 2 and 4 [35]. If A is the area under the ROC curve, then its extreme values are $A = \frac{1}{2}$ if the two distributions coincide, and $A = 1$ in case the distributions completely separate. Therefore, we have $\frac{1}{2} \leq A \leq 1$ and measure the discrimination $\mathcal{D} = 2(1 - A)$, which tends to zero as the distributions increasingly separate. Summing over K terms, and potentially including a decimation by D takes time to evaluate; therefore, a second metric is the decision time $\mathcal{T} = (K - 1)D + 1$, which is required to reach the above discrimination \mathcal{D} . We therefore assess discrimination $\mathcal{D}(\mathcal{T})$ as a function of the decision time \mathcal{T} below.

TABLE I

POWER OF CONTRIBUTIONS FOR REALISTIC CHANNEL SCENARIO.

signal	power
source 1	0.0000 dB
source 2	-4.3494 dB
source 3	-13.2865 dB
noise	-16.2865 dB

B. Test Scenario

Based on multipath propagation responses determine by a radio planning tool in the ISM band with a bandwidth of 20MHz, we have generated transfer functions of order 40 for the case of three sources and $M = 6$ receivers in a dense urban environment. The transfer functions are such that the total power at the receivers for the various sources is as given in Table I. Spatially and temporally uncorrelated additive Gaussian noise corrupts each receiver at 3dB below the weakest source.

We will work with two stationary sources, with the remaining third source — either the medium-powered source 2 or the weakest source 3 — contributing a transient signal. The CSD matrix is either estimated from a finite amount of N samples [32], [33], or calculated based on the ground truth steering vectors and power spectral densities of the stationary sources, i.e. for $N \rightarrow \infty$.

C. Discrimination vs Decision Time

Fig. 5 shows two groups of curves. For a first group, sources 1 and 3 from Tab. I make up the stationary signals, and source 2 — of medium power — mimics the transient signal. In this case, the distributions with the transitory source 2 being absent or present separate well, and drop to close to machine accuracy for a only a few number of summation terms K . The estimation of the CSD matrix leads to some subspace leakage [36], but with almost no performance degradation for sample sizes of $N = 10000$ and $N = 1000$, where working with an estimate is almost as good as operating with the ground truth CSD matrix. Still very good discrimination is achieved even when estimating $\hat{R}(z)$ over a very small sample size of $N = 100$. The discrimination time \mathcal{T} here refers to a sampling rate of 20MHz, with a value of $5\mu s$ being equivalent to averaging over $K = 10$ samples at a decimation ratio $D = 10$.

A second group of curves in Fig. 5 refers to sources 1 and 2 forming the group of stationary signals, and the weakest source 3 is taken as the transient signal. In this case, a larger value K for temporal averaging is required to achieve better discrimination, i.e. small values of \mathcal{D} . The subspace leakage experienced in the estimation process is now more pronounced for low sample sizes N , such that the approach still works for larger samples sizes but breaks down for $N = 100$, where the low power of the source and the comparatively high noise level lead to no discernible differences in the distributions.

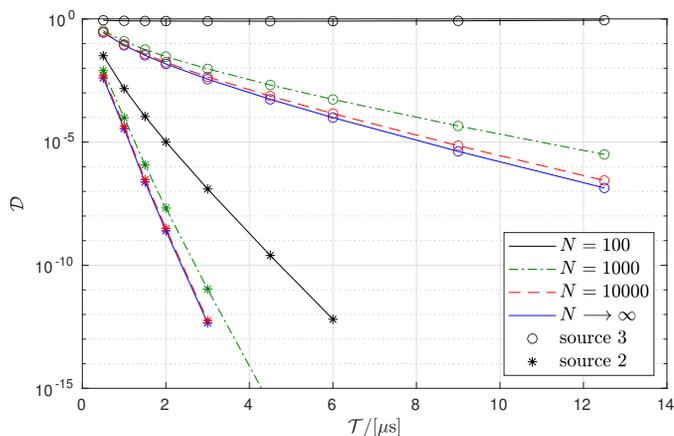


Fig. 5. Plots showing discrimination \mathcal{D} vs decision time \mathcal{T} for detecting sources of different strength, and based on estimates of the space-time covariance relying on various sample sizes N .

VI. CONCLUSION

In this paper we have proposed a broadband subspace-based processor for the detection of weak transient signals. By identifying the subspace of stationary sources, we monitor a 'syndrome vector' that gives access to the energy in the noise-only subspace, and can be indicative of emerging source signals. The discrimination of the processor can be enhanced by temporal averaging, whereby a decimation stage reduces the computation complexity as well the correlation in the data, permitting to use generalised χ^2 statistics in assessing the distribution of the data. A simulation has demonstrated that good discrimination is possible, whereby weaker sources can be reliably detected as long as the space-time covariance is estimated with sufficient accuracy, i.e. over a sufficiently long time window.

The evaluation of the discrimination is threshold-independent, and future work will focus on setting a suitable detection threshold for a hypothesis test on the absence or presence of a transient signal.

REFERENCES

- [1] Z. Quan, S. Cui, H.V. Poor, and A.H. Sayed, "Collaborative wideband sensing for cognitive radios," *IEEE SP Mag.*, **25**(6):60–73, Nov. 2008.
- [2] B. Friedlander and B. Porat, "Detection of transient signals by the Gabor representation," *IEEE Trans. ASSP*, **37**(2):169–180, Feb. 1989.
- [3] —, "Performance analysis of transient detectors based on a class of linear data transforms," *IEEE Trans. IT*, **38**(2):665–673, Mar. 1992.
- [4] B. Porat and B. Friedlander, "Performance analysis of a class of transient detection algorithms—a unified framework," *IEEE Trans. SP*, **40**(10):2536–2546, Oct. 1992.
- [5] S. Haykin, *Adaptive Filter Theory*, 2nd ed. Englewood Cliffs: Prentice Hall, 1991.
- [6] L.L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans. SP*, **42**(8):2146–2157, Aug. 1994.
- [7] P. Strobach, "Low rank detection of multichannel gaussian signals using a constrained inverse," in *IEEE ICASSP*, pp. IV/245–IV/248, Apr. 1994.
- [8] D. Lundstrom, M. Viberg, and A.M. Zoubir, "Multiple transient estimation using bootstrap and subspace methods," in *9th IEEE SSP*, pp. 184–187, Sep. 1998.
- [9] Z. Wang and P. Willett, "A performance study of some transient detectors," *IEEE Trans. SP*, **48**(9):2682–2685, Sep. 2000.
- [10] Z. Wang and P. K. Willett, "All-purpose and plug-in power-law detectors for transient signals," *IEEE Trans. SP*, **49**(11):2454–2466, Nov. 2001.

- [11] A. Marchioni, M. Mangia, F. Pareschi, R. Rovatti, and G. Setti, "Subspace energy monitoring for anomaly detection @sensor or @edge," *IEEE Internet of Things Journal*, **7**(8):7575–7589, Aug. 2020.
- [12] S. Weiss and I.K. Proudler, "Comparing Efficient Broadband Beamforming Architectures and Their Performance Trade-Offs," in *14th Int. Conf. DSP*, Santorini, Greece, pp. 417–422, July 2002.
- [13] M. Alrmah, J. Corr, A. Alzin, K. Thompson, and S. Weiss, "Polynomial subspace decomposition for broadband angle of arrival estimation," in *SSPD*, Edinburgh, Scotland, Sept 2014, pp. 1–5.
- [14] J.G. McWhirter, P.D. Baxter, T. Cooper, S. Redif, and J. Foster, "An EVD Algorithm for Para-Hermitian Polynomial Matrices," *IEEE Trans. SP*, **55**(5):2158–2169, May 2007.
- [15] S. Weiss, J. Pestana, and I.K. Proudler, "On the existence and uniqueness of the eigenvalue decomposition of a parahermitian matrix," *IEEE Trans. SP*, **66**(10):2659–2672, May 2018.
- [16] S. Weiss, J. Pestana, I. Proudler, and F. Coutts, "Corrections to "on the existence and uniqueness of the eigenvalue decomposition of a parahermitian matrix"," *IEEE Trans. SP*, **66**(23):6325–6327, Dec. 2018.
- [17] S. Redif, J. McWhirter, and S. Weiss, "Design of FIR paraunitary filter banks for subband coding using a polynomial eigenvalue decomposition," *IEEE Trans. SP*, **59**(11):5253–5264, Nov. 2011.
- [18] S. Redif, S. Weiss, and J. McWhirter, "Sequential matrix diagonalization algorithms for polynomial EVD of parahermitian matrices," *IEEE Trans. SP*, **63**(1):81–89, Jan. 2015.
- [19] S. Weiss, I. K. Proudler, F. K. Coutts, and J. Pestana, "Iterative approximation of analytic eigenvalues of a parahermitian matrix EVD," in *IEEE ICASSP*, Brighton, UK, May 2019.
- [20] S. Weiss, I.K. Proudler, F. Coutts, and J. Deeks, "Extraction of analytic eigenvalues from a parahermitian matrix," in *SSPD*, 2020.
- [21] S. Weiss, I.K. Proudler, and F.K. Coutts, "Eigenvalue decomposition of a parahermitian matrix: Extraction of analytic eigenvalues," *IEEE Trans. SP*, **69**:722–737, Jan. 2021.
- [22] H. Solomon and M.A. Stephens, "Distribution of a sum of weighted chi-square variables," *Journal of the American Statistical Association*, **72**(360):881–885, 1977.
- [23] R.B. Davies, "The distribution of a linear combination of χ^2 random variables," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **29**(3):323–333, 1980.
- [24] D. Hammarwall, M. Bengtsson, and B. Ottersten, "Acquiring partial CSI for spatially selective transmission by instantaneous channel norm feedback," *IEEE Trans. SP*, **56**(3):1188–1204, Mar. 2008.
- [25] E. Bjornson, D. Hammarwall, and B. Ottersten, "Exploiting quantized channel norm feedback through conditional statistics in arbitrarily correlated mimo systems," *IEEE Trans. SP*, **57**(10):4027–4041, Oct. 2009.
- [26] A. Das and W.S. Geisler, "A method to integrate and classify normal distributions," *arXiv e-prints*, p. arXiv:2012.14331, Dec. 2020.
- [27] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, 3rd ed. New York: McGraw-Hill, 1991.
- [28] P.P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs: Prentice Hall, 1993.
- [29] T.I. Laakso, V. Välimäki, M. Karjalainen, and U.K. Laine, "Splitting the Unit Delay," *IEEE Signal Processing Mag.*, **13**(1):30–60, Jan. 1996.
- [30] M. Alrmah, S. Weiss, and J. McWhirter, "Implementation of accurate broadband steering vectors for broadband angle of arrival estimation," in *IET Intelligent Signal Processing*, London, UK, Dec. 2013.
- [31] G.H. Golub and C.F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, Maryland: John Hopkins University Press, 1996.
- [32] C. Delaosa, J. Pestana, N.J. Goddard, S. Somasundaram, and S. Weiss, "Sample space-time covariance matrix estimation," in *IEEE ICASSP*, May 2019, pp. 8033–8037.
- [33] C. Delaosa, J. Pestana, N.J. Goddard, S. D. Somasundaram, and S. Weiss, "Support estimation of a sample space-time covariance matrix," in *SSPD*, Brighton, UK, Mar. 2019.
- [34] F. K. Coutts, I. K. Proudler, and S. Weiss, "Efficient implementation of iterative polynomial matrix evd algorithms exploiting structural redundancy and parallelisation," *IEEE Trans. CAS I: Regular Papers*, **66**(12):4753–4766, Dec. 2019.
- [35] J.A. Hanley and B.J. McNeil, "The Meaning and Use of the Area under a Receiver Operating Characteristic (ROC) Curve," *Radiology*, vol. 143, pp. 26–36, 1982.
- [36] C. Delaosa, F.K. Coutts, J. Pestana, and S. Weiss, "Impact of space-time covariance estimation errors on a parahermitian matrix EVD," in *10th IEEE SAM*, pp. 1–5, July 2018.

Rate Splitting Multiple Access for Multi-Antenna Multi-Carrier Joint Communications and Jamming

Onur Dizdar and Bruno Clerckx

Department of Electrical and Electronic Engineering, Imperial College London

Email: {o.dizdar, b.clerckx}@imperial.ac.uk

Abstract—Rate-Splitting Multiple Access (RSMA) is a robust multiple access scheme for downlink multi-antenna wireless networks. In this work, we investigate a novel application of RSMA for joint communications and jamming with a Multi-Carrier (MC) waveform in Multiple Input Single Output (MISO) Broadcast Channel (BC). Our aim is to simultaneously communicate with Information Users (IU) and jam Adversarial Users (AU) to disrupt their communications in a setting where all users perform broadband communications by MC waveforms in their respective networks. We consider the practical setting of imperfect CSI at transmitter (CSIT) for the IUs and statistical CSIT for AUs. The optimal information and jamming precoders are designed to maximize the sum-rate under jamming power constraints on the pilot subcarriers of AUs, a jamming method considered to be among the most destructive methods for MC waveforms under the considered system model. We compare the sum-rate performance of RSMA and Space Division Multiple Access (SDMA) schemes by numerical results to demonstrate that RSMA achieves a significant sum-rate gain compared to SDMA.

I. INTRODUCTION

Rate-Splitting Multiple Access (RSMA) is a multiple access scheme based on the concept of Rate-Splitting (RS) and linear precoding for multi-antenna multi-user communications. RSMA splits user messages into common and private parts, and encodes the common parts into one or several common streams while encoding the private parts into separate streams. The streams are precoded using the available (perfect or imperfect) Channel State Information at the Transmitter (CSIT), superposed and transmitted via the Multi-Input Multi-Output (MIMO) or Multi-Input Single-Output (MISO) channel [1]. All the receivers then decode the common stream(s), perform Successive Interference Cancellation (SIC) and then decode their respective private streams. Each receiver reconstructs its original message from the part of its message embedded in the common stream(s) and its intended private stream.

RSMA manages multi-user interference by allowing the interference to be partially decoded and partially treated as noise. RSMA has been shown to embrace and outperform existing multiple access schemes, i.e., Space Division Multiple Access (SDMA), Non-Orthogonal Multiple

Access (NOMA), Orthogonal Multiple Access (OMA) and multicasting. The sum-rate performance of RSMA has been demonstrated to be robust and to surpass the performance of SDMA and NOMA under perfect and imperfect CSIT in numerous works [1]–[5].

With increasing number of systems and applications, the Radio-Frequency (RF) spectrum has become a congested and contested environment. Both commercial and military systems require broadband communications to meet the data requirements for the advancing applications. In such a congested and contested environment, efficient use and sharing of spectrum is of high importance, especially for military communications with strict reliability and robustness requirements. An equally critical target in military communications is denying the Adversarial Users (AUs) of service. In this work, we are interested in jamming methods targeting Multi-Carrier (MC) waveforms used for broadband communications. Among the jamming methods for MC waveforms, pilot subcarrier jamming is accepted to be one of the most destructive ones which consist of Artificial Noise (AN) (*i.e.*, excluding the methods which use valid data signals to cause misconception in the network or attack the synchronisation of the waveform). Pilot jamming aims to disrupt the channel estimation procedure of the affected user to prevent error-free detection and decoding of its intended messages [7]–[11].

In this work, we consider a scenario where a multi-antenna military system aims to communicate in an RF-congested environment while simultaneously performing jamming to AUs in the same spectrum. We use RSMA for multi-carrier multiple-access communications. Our aim is to identify the performance benefits of RSMA for joint communications and jamming and obtain optimal precoders to maximize the mutual information for communications with Information Users (IUs) while performing jamming on pilot subcarriers of AUs efficiently. We consider the practical and realistic scenario of imperfect Channel State Information at Transmitter (CSIT) for the IUs and statistical CSIT for AUs, since obtaining an accurate channel estimate for the AUs is generally not feasible [13]. We give a formulation for MC waveforms to solve the abovementioned problem, which turns out to be non-convex. We propose a Successive Convex Approximation (SCA) based algorithm, similar to the algorithm for single carrier waveforms in [14], combined with the Sample

This work was supported by the Engineering and Physical Sciences Research Council of the UK (EPSRC) Grant number EP/S026657/1, and the UK MOD University Defence Research Collaboration (UDRC) in Signal Processing.

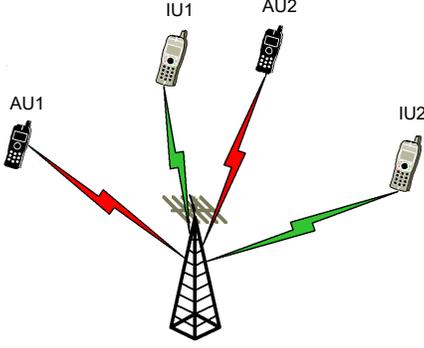


Fig. 1: System model.

Average Approximated Weighted Minimum Mean Square (SAA-WMMSE) algorithm in [3] to solve the problem. We demonstrate by simulations that RSMA achieves a significant sum-rate performance over SDMA.

Notations: Vectors are denoted by bold lowercase letters and matrices are denoted by bold uppercase letters. The operations $|\cdot|$ and $\|\cdot\|$ denote the absolute value of a scalar and l_2 -norm of a vector, respectively, unless stated otherwise. \mathbf{a}^H denotes the Hermitian transpose of a vector \mathbf{a} . $\mathcal{CN}(0, \sigma^2)$ denotes the Circularly Symmetric Complex Gaussian distribution with zero mean and variance σ^2 . \mathbf{I}_n denotes the n -by- n identity matrix. The operator $\text{tr}(\cdot)$ denotes the trace operation. The operator $\text{Diag}(X_1, \dots, X_K)$ builds a matrix \mathbf{X} by placing the matrices X_1, \dots, X_K diagonally and setting all other elements to zero.

II. SYSTEM MODEL

We consider a Multi-Input Single-Output (MISO) Broadcast Channel (BC) setting consisting of one transmitter with n_t transmit antennas, which aims to perform communications and jamming simultaneously in an RF-congested environment, as illustrated in Fig. 1. The transmitter serves K single-antenna IUs indexed by $\mathcal{K} = \{1, 2, \dots, K\}$, while simultaneously performing jamming on L single-antenna AUs indexed by $\mathcal{L} = \{1, 2, \dots, L\}$. We assume that the IUs in the network are operating in the same frequency band as the AUs. The transmitter employs an MC waveform to communicate with the IUs, while the AUs also use an MC waveform to communicate in their corresponding separate networks. We define the set of subcarrier indexes in the signal band as $\mathcal{S} = \{1, 2, \dots, N\}$. Also, we define the set of pilot subcarriers of AUs in the same signal band as $\mathcal{S}_p \subset \mathcal{S}$.

We employ 1-layer RSMA [4] to perform multiple-access communications in the considered setting. RSMA relies on splitting the user messages at the transmitter side. The messages intended for the users, $W_{k,n}$, are split into common and private parts, i.e., $W_{c,k,n}$ and $W_{p,k,n}$, $\forall k \in \mathcal{K}, n \in \mathcal{S}$. The common parts of the messages of all users are combined into the common message $W_{c,n}$. The common message $W_{c,n}$ and the private messages are independently encoded into streams $s_{c,n}$ and $s_{k,n}$, respectively. Jamming is performed on subcarrier- n of AU- l using the AN signal $s_{l,n}^e$, $\forall l \in \mathcal{L}$ and $\forall n \in \mathcal{S}_p$. We assume that each subcarrier is assigned a separate precoder. The MC transmit signal for RSMA is

$$x_n = \mathbf{p}_{c,n} s_{c,n}^f + \sum_{k=1}^K \mathbf{p}_{k,n} s_{k,n}^f + \sum_{l=1}^L \mathbf{f}_{l,n} s_{l,n}^e, \quad n \in \mathcal{S}.$$

The vectors $\mathbf{p}_{c,n} \in \mathbb{C}^{n_t}$ and $\mathbf{p}_{k,n} \in \mathbb{C}^{n_t}$ are the linear precoders applied to the common stream and the private stream of IU- k on subcarrier n , $\forall k \in \mathcal{K}$ and $\forall n \in \mathcal{S}$. The precoder $\mathbf{f}_{l,n}$ is used to transmit AN to AU- l , $\forall l \in \mathcal{L}$. The communications signals $s_{c,n}^f$ and $s_{k,n}^f$ and jamming signals $s_{l,n}^e$ are chosen independently from a Gaussian alphabet for theoretical analysis. We also assume that the streams have unit power, so that $\mathbb{E}\{\mathbf{s}_n(\mathbf{s}_n)^H\} = \mathbf{I}_{K+L+1}$, where $\mathbf{s}_n = [s_{c,n}^f, s_{1,n}^f, \dots, s_{K,n}^f, s_{1,n}^e, \dots, s_{L,n}^e]$. An average transmit power constraint is set as $\sum_{n=1}^N \text{tr}(\mathbf{P}_n \mathbf{P}_n^H) + \text{tr}(\mathbf{F}_n \mathbf{F}_n^H) \leq \bar{P}_t$, where $\mathbf{P}_n = [\mathbf{p}_{c,n} \mathbf{p}_{1,n}, \dots, \mathbf{p}_{K,n}]$ and $\mathbf{F}_n = [\mathbf{f}_{1,n}, \dots, \mathbf{f}_{L,n}]$. The signal received by IU- k on subcarrier- n is

$$y_{k,n} = \mathbf{h}_{k,n}^H \mathbf{x}_n + z_{k,n}, \quad k \in \mathcal{K}, n \in \mathcal{S}, \quad (1)$$

where $\mathbf{h}_{k,n} \in \mathbb{C}^{n_t}$ is the channel vector of IU- k on subcarrier- n and $z_{k,n} \sim \mathcal{CN}(0, 1)$ is the Additive White Gaussian Noise (AWGN). Similarly, the signal received by AU- l on subcarrier- n is written as

$$r_{l,n} = \mathbf{g}_{l,n}^H \mathbf{x}_n + \nu_{l,n}, \quad l \in \mathcal{L}, n \in \mathcal{S}, \quad (2)$$

where $\mathbf{g}_{l,n} \in \mathbb{C}^{n_t}$ is the channel vector of AU- l on subcarrier- n and $\nu_{l,n} \sim \mathcal{CN}(0, 1)$ is the AWGN.

At the receiver side, detection of the messages is carried out using Successive Interference Cancellation (SIC). The common stream is detected first to obtain the common message estimate $\hat{W}_{c,n}$ by treating the private streams as noise. The common stream is then reconstructed using $\hat{W}_{c,n}$ and subtracted from the received signal. The remaining signal is used to detect the private messages $\hat{W}_{p,k,n}$. Finally, the estimated message for IU- k , $\hat{W}_{k,n}$, is obtained by combining $\hat{W}_{c,k,n}$ and $\hat{W}_{p,k,n}$. We write the Signal-to-Interference-plus-Noise Ratio (SINR) expressions for the common and private streams at IU- k as

$$\gamma_{c,k,n} = \frac{|\mathbf{h}_{k,n}^H \mathbf{p}_{c,n}|^2}{1 + Z_{c,k,n} + J_{k,n}}, \quad \gamma_{k,n} = \frac{|\mathbf{h}_{k,n}^H \mathbf{p}_{k,n}|^2}{1 + Z_{k,n} + J_{k,n}},$$

with $Z_{c,k,n} = \sum_{i \in \mathcal{K}} |\mathbf{h}_{k,n}^H \mathbf{p}_{i,n}|^2$, $J_k = \sum_{j \in \mathcal{L}} |\mathbf{h}_{k,n}^H \mathbf{f}_{j,n}|^2$ and $Z_{k,n} = \sum_{i \in \mathcal{K}, i \neq k} |\mathbf{h}_{k,n}^H \mathbf{p}_{i,n}|^2$.

In this work, we consider the notion of jamming in the context of denial of service for the AUs. Our aim is to efficiently focus power on the AUs to disrupt the correct detection and decoding of their intended data transmissions from other users in their corresponding network. Our performance criterion is the focused power on pilot subcarrier- n of an AU- l , $n \in \mathcal{S}_p$ and $l \in \mathcal{L}$, expressed as

$$\Lambda_{l,n} = |\mathbf{g}_{l,n}^H \mathbf{p}_{c,n}|^2 + \sum_{k \in \mathcal{K}} |\mathbf{g}_{l,n}^H \mathbf{p}_{k,n}|^2 + \sum_{l' \in \mathcal{L}} |\mathbf{g}_{l,n}^H \mathbf{f}_{l',n}|^2.$$

We assume that the transmitter has synchronisation with the AU transmissions [9], [12], [13] and a perfect knowledge of \mathcal{S}_p [7]–[11].

We consider the practical case where the transmitter does not have access to perfect Channel State Information (CSI). The channel model of IU- k is expressed as

$$\mathbf{h}_{k,n} = \sqrt{1 - \sigma_{ie}^2} \hat{\mathbf{h}}_{k,n} + \sigma_{ie} \tilde{\mathbf{h}}_{k,n}, \quad (3)$$

where $\hat{\mathbf{h}}_{k,n}$ is the estimate of the channel on subcarrier- n at the transmitter and $\tilde{\mathbf{h}}_{k,n}$ is the channel estimation error with i.i.d. complex Gaussian elements of unit variance. The covariance matrix of the channel of AU- l on subcarrier- n is expressed as $\mathbf{R}_{g_{l,n}} = \mathbb{E} \left\{ \mathbf{g}_{l,n} \mathbf{g}_{l,n}^H \right\}$. We assume that the channel is fixed during the transmission of an MC waveform block. We also assume perfect CSI at the receivers.

III. RSMA FOR JOINT COMMUNICATIONS AND JAMMING WITH OFDM WAVEFORM

In this section, we give the problem formulation to obtain the optimal precoders for the system model in Section II. Our objective is to maximize the ergodic mutual information under imperfect CSIT for IUs while focusing a certain amount of jamming power on the pilot subcarriers of the AUs. The receiver employs carrier non-cooperative processing of the MC waveform. Such approach considers an independent processing of each subcarrier at the receiver, which is suitable for practical scenarios due to its low complexity [16].

We define the matrices $\mathbf{H}_k = \text{Diag}(\mathbf{h}_{k,1}, \dots, \mathbf{h}_{k,N})$, $\mathbf{P} = \text{Diag}(\mathbf{P}_1, \dots, \mathbf{P}_N)$ and $\mathbf{Z}_k = \text{Diag}((Z_{k,1} + J_{k,1} + N_0), \dots, (Z_{k,N} + J_{k,N} + N_0))$. Under the assumption of carrier non-cooperative processing, the mutual information at IU- k is expressed as $I_k = \log |\mathbf{I} + \mathbf{Z}_k^{-1} \mathbf{H}_k^H \mathbf{P} \mathbf{P}^H \mathbf{H}_k|$ [15], [16].

In order to obtain the optimal precoders that maximize the mutual information, we make use of the mutual information - Mean Square Error (MSE) relations. We note that in addition to the numerous works, such approach is taken for designing the optimal precoders for MC multi-antenna systems in [16] and RSMA in MISO BC in [3]. We first obtain the optimal receive filter, $g_{k,n}$, that minimizes the Mean Square Error (MSE) $\mathbb{E} \{ |\epsilon_{k,n}|^2 \} = \mathbb{E} \{ |g_{k,n} y_{k,n} - x_{k,n}|^2 \}$, $\forall k \in \mathcal{K}, \forall n \in \mathcal{S}$. It is well known that the solution is given by a Minimum MSE (MMSE) filter

$$g_{k,n}^{\text{opt}} = \mathbf{p}_{k,n}^H \mathbf{h}_{k,n} (|\mathbf{h}_{k,n}^H \mathbf{p}_{k,n}|^2 + Z_{k,n} + J_{k,n} + N_0)^{-1}.$$

The resulting MSE is expressed as

$$\epsilon_{k,n}^{\text{opt}} = (|\mathbf{h}_{k,n}^H \mathbf{p}_{k,n}|^2 + Z_{k,n} + J_{k,n} + N_0)^{-1} (Z_{k,n} + J_{k,n} + N_0). \quad (4)$$

The mutual information-MSE relation is given by $I_k = -\log |\mathbf{E}_k|$, where $\mathbf{E}_k = \text{Diag}(\epsilon_{k,1}^{\text{opt}}, \dots, \epsilon_{k,N}^{\text{opt}})$ [16], which can be expanded as

$$I_k = -\log \left(\prod_{n=1}^N \epsilon_{k,n}^{\text{opt}} \right) = -\sum_{n=1}^N \log(\epsilon_{k,n}^{\text{opt}}) = \sum_{n=1}^N I_{k,n}. \quad (5)$$

In the context of RSMA, (5) represents the mutual information for the private stream of IU- k and has the requirement of being decodable by the corresponding user. On the other hand, the common stream has a stricter requirement of being decodable by all IUs in the system. In order to satisfy such a requirement, we consider the mutual information per subcarrier for the common stream, so that the decodability of the message on each subcarrier is

guaranteed by all IUs¹. We define $I_{c,k,n} = -\log(\epsilon_{c,k,n}^{\text{opt}})$, where $\epsilon_{c,k,n}^{\text{opt}}$ is obtained by replacing $\mathbf{p}_{k,n}$ and $Z_{k,n}$ in (4) by $\mathbf{p}_{c,n}$ and $Z_{c,k,n}$, respectively.

Next, we determine the jamming power constraint for statistical CSIT. We consider the average power focused on subcarrier- n of AU- l [17]. Accordingly,

$$\begin{aligned} & \mathbb{E} \left\{ |\mathbf{g}_{l,n}^H \mathbf{p}_{c,n}|^2 + \sum_{k \in \mathcal{K}} |\mathbf{g}_{l,n}^H \mathbf{p}_{k,n}|^2 + \sum_{l' \in \mathcal{L}} |\mathbf{g}_{l,n}^H \mathbf{f}_{l',n}|^2 \right\} \\ &= \mathbf{p}_{c,n}^H \mathbf{R}_{g_{l,n}} \mathbf{p}_{c,n} + \sum_{k \in \mathcal{K}} \mathbf{p}_{k,n}^H \mathbf{R}_{g_{l,n}} \mathbf{p}_{k,n} + \sum_{l' \in \mathcal{L}} \mathbf{f}_{l',n}^H \mathbf{R}_{g_{l,n}} \mathbf{f}_{l',n} \\ &\triangleq \bar{\Lambda}_{l,n}(\mathbf{P}_n, \mathbf{F}_n). \end{aligned} \quad (6)$$

We formulate the optimization problem as

$$\max_{\mathbf{P}, \mathbf{F}, \mathbf{c}} \sum_{n \in \mathcal{S}} \sum_{k \in \mathcal{K}} (C_{k,n} + I_{k,n}(\mathbf{P}_n, \mathbf{F}_n)) \quad (7a)$$

$$\text{s.t.} \quad \sum_{k \in \mathcal{K}} C_{k,n} \leq I_{c,k,n}(\mathbf{P}_n, \mathbf{F}_n), \forall n \in \mathcal{S}, \forall k \in \mathcal{K} \quad (7b)$$

$$\bar{\Lambda}_{l,n}(\mathbf{P}_n, \mathbf{F}_n) \geq J_{l,n}^{\text{thr}}, \quad \forall n \in \mathcal{S}_p, \forall l \in \mathcal{L} \quad (7c)$$

$$\sum_{n \in \mathcal{S}} \text{tr}(\mathbf{P}_n \mathbf{P}_n^H) + \text{tr}(\mathbf{F}_n \mathbf{F}_n^H) \leq \bar{P}_t, \quad (7d)$$

$$\mathbf{c} \geq \mathbf{0}. \quad (7e)$$

where $\mathbf{c} = [C_{1,1}, \dots, C_{K,1}, \dots, C_{1,N}, \dots, C_{K,N}]^T$ with $C_{k,n}$ being the portion of the common mutual information for IU- k on subcarrier- n . Rate-MSE transformations similar to the one in (5) have been used in [3] without an MC waveform to transform the non-convex sum-rate maximization problem for RSMA into a convex one. Therefore, we follow the approach in [3] and define the augmented MSEs $\xi_{c,k,n} = u_{c,k,n} \epsilon_{c,k,n} - \log_2(u_{c,k,n})$ and $\xi_{k,n} = u_{k,n} \epsilon_{k,n} - \log_2(u_{k,n})$ to obtain $\xi_{k,n}^{\text{opt}} = 1 - I_{k,n}$ and $\xi_{c,k,n}^{\text{opt}} = 1 - I_{c,k,n}$, $\forall n \in \mathcal{S}$. The resulting formulation is

$$\min_{\mathbf{P}, \mathbf{F}, \mathbf{x}, \mathbf{u}, \mathbf{g}} \sum_{n \in \mathcal{S}} \sum_{k \in \mathcal{K}} (X_{k,n} + \xi_{k,n}(\mathbf{P}_n, \mathbf{F}_n)) \quad (8a)$$

$$\text{s.t.} \quad \sum_{k \in \mathcal{K}} X_{k,n} + 1 \geq \xi_{c,k,n}(\mathbf{P}_n, \mathbf{F}_n), \quad \forall n \in \mathcal{S}, \forall k \in \mathcal{K} \quad (8b)$$

$$\bar{\Lambda}_{l,n}(\mathbf{P}_n, \mathbf{F}_n) \geq J_{l,n}^{\text{thr}}, \quad \forall n \in \mathcal{S}_p, \forall l \in \mathcal{L} \quad (8c)$$

$$\sum_{n \in \mathcal{S}} \text{tr}(\mathbf{P}_n \mathbf{P}_n^H) + \text{tr}(\mathbf{F}_n \mathbf{F}_n^H) \leq \bar{P}_t, \quad (8d)$$

$$\mathbf{x} \leq \mathbf{0}, \quad (8e)$$

where $\mathbf{x} = [X_{1,1}, \dots, X_{K,1}, \dots, X_{1,N}, \dots, X_{K,N}]^T$ and $X_{k,n} = -C_{k,n}$. For the sake of brevity, we skip detailed derivations to extend the problem formulation for the imperfect CSIT case by the Sample Average Approximation (SAA) and refer the interested reader to [3].

We note that the constraint (8c) is not convex due to the convex function $\bar{\Lambda}_{l,n}(\mathbf{P}_n)$ constrained to a lower bound. We follow the approach followed in [14] and obtain a convex constraint by using first-order Taylor expansion for the quadratic function in (8c). Specifically, one can lower bound the term $\mathbf{p}_{k,n}^H \mathbf{R}_{g_{l,n}} \mathbf{p}_{k,n}$ at a point $\mathbf{p}_{k,n}^t$ as

¹By carrier cooperative processing, the decodability can be guaranteed over $I_{c,k}$ instead of $I_{c,k,n}$ with proper coding methods [18].

Algorithm 1: SCA-based algorithm

```

 $t \leftarrow 0, i \leftarrow 0, \text{WSR}^0 \leftarrow 0, \hat{\mathbf{P}}, \hat{\mathbf{F}}$ 
while  $|\text{WSR}^i - \text{WSR}^{i-1}| > \epsilon_r$  do
     $\mathbf{u}^i \leftarrow \text{updateWeights}(\hat{\mathbf{P}}, \hat{\mathbf{F}})$ 
     $\mathbf{g}^i \leftarrow \text{updateFilters}(\hat{\mathbf{P}}, \hat{\mathbf{F}})$ 
    while  $|\text{WMMSE}^t - \text{WMMSE}^{t-1}| > \epsilon_m$  do
         $(\mathbf{P}^{t+1}, \mathbf{F}^{t+1}, \mathbf{x}^{t+1}) \leftarrow (10)$  for given  $\mathbf{u}^i$ ,
         $\mathbf{g}^i$  and  $(\mathbf{P}^t, \mathbf{F}^t)$ 
         $\text{WMMSE}^{t+1} \leftarrow \text{updateWMMSE}(\mathbf{P}^{t+1}, \mathbf{F}^{t+1})$ 
         $t \leftarrow t + 1$ 
    end
     $\hat{\mathbf{P}} \leftarrow \mathbf{P}^t, \hat{\mathbf{F}} \leftarrow \mathbf{F}^t$ 
     $\text{WSR}^{i+1} \leftarrow \text{updateWSR}(\hat{\mathbf{P}}, \hat{\mathbf{F}})$ 
     $t \leftarrow 0, i \leftarrow i + 1$ 
end
return  $(\hat{\mathbf{P}}, \hat{\mathbf{F}})$ 

```

$$\begin{aligned}
 & \mathbf{p}_{k,n}^H \mathbf{R}_{g_{l,n}} \mathbf{p}_{k,n} \\
 & \geq 2\text{Re} \{ (\mathbf{p}_{k,n}^t)^H \mathbf{R}_{g_{l,n}} \mathbf{p}_{k,n} \} - (\mathbf{p}_{k,n}^t)^H \mathbf{R}_{g_{l,n}} \mathbf{p}_{k,n}^t \\
 & \triangleq \bar{\phi}^t(\mathbf{p}_{k,n}, \mathbf{R}_{l,n}) \quad (9)
 \end{aligned}$$

Using the function in (9), we write

$$\begin{aligned}
 \bar{\Lambda}_{l,n}(\mathbf{P}_n, \mathbf{F}_n) & \geq \\
 & \bar{\phi}^t(\mathbf{p}_{c,n}, \mathbf{R}_{l,n}) + \sum_{k \in \mathcal{K}} \bar{\phi}^t(\mathbf{p}_{k,n}, \mathbf{R}_{l,n}) + \sum_{l \in \mathcal{L}} \bar{\phi}^t(\mathbf{f}_{l,n}, \mathbf{R}_{l,n}) \\
 & \triangleq \bar{\Lambda}_{l,n}^t(\mathbf{P}_n, \mathbf{F}_n)
 \end{aligned}$$

The final problem formulation is written as

$$\begin{aligned}
 & \min_{\mathbf{P}, \mathbf{F}, \mathbf{x}, \mathbf{u}, \mathbf{g}} \sum_{n \in \mathcal{S}} \sum_{k \in \mathcal{K}} (X_{k,n} + \xi_{k,n}(\mathbf{P}_n, \mathbf{F}_n)) \quad (10a) \\
 & \text{s.t. } \bar{\Lambda}_{l,n}^t(\mathbf{P}_n, \mathbf{F}_n) \geq J_{l,n}^{thr}, n \in \mathcal{S}_p, l \in \mathcal{L} \quad (10b) \\
 & \quad (8b), (8d) (8e).
 \end{aligned}$$

The SCA-based algorithm to solve (10) is given in Alg. 1.

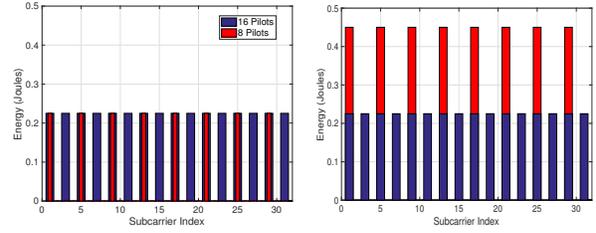
We set the power threshold on subcarrier- n of AU- l as

$$J_{l,n}^{thr} = \rho \frac{\bar{P}_t}{|S_p|L} \tau_{l,n}, \quad \forall n \in \mathcal{S}_p L, \forall l \in \mathcal{L}, \quad (11)$$

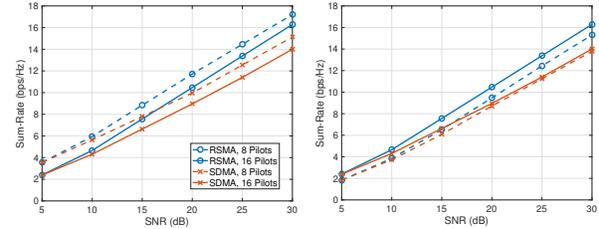
where $\rho \in [0, 1]$ determines the strictness of the jamming power threshold by adjusting it to a percentage of the maximum value, $|S_p|$ denotes the cardinality of the set S_p and $\tau_{n,l}$ is the largest eigenvalue of the matrix $\mathbf{R}_{g_{l,n}}$.

IV. SIMULATION RESULTS

We perform simulations to demonstrate the sum-rate performance achieved by RSMA and SDMA using the optimized precoders. Note that the optimal precoders for SDMA can be obtained by turning off the common stream in the optimization problem formulation. We set $n_t = 4$, $K = 2$ and $L = 1$ for the system parameters. We use Cyclic-Prefix (CP)-OFDM waveform with 32 subcarriers and a CP length of $10\mu\text{s}$. The error variance of the channels of the IUs is modeled as $\sigma_{ie}^2 = (\bar{P}_t/N)^{-\alpha}$, where α is the CSIT quality scaling factor [3], [6], and is set as $\alpha = 0.6$ in the simulations. We define the private rate of IU- k for an MC waveform as $R_k = I_k/N$, $\forall k \in \mathcal{K}$, following the formulation in [18] for the carrier



(a) Strategy 1 ($\rho = 0.9 \frac{|S_p|}{N/2}$). (b) Strategy 2 ($\rho = 0.9$).
Fig. 2: Power thresholds with different strategies.



(a) Sum-rate vs. SNR, strategy 1. (b) Sum-rate vs. SNR, strategy 2.
Fig. 3: Sum-rate vs. SNR with different strategies.

cooperative case. This serves as an upper bound for the non-cooperative case since carrier cooperative processing is a more general model [16]. Accordingly, the common rate is defined as $R_{c,k} = \frac{1}{N} \sum_{k \in \mathcal{K}} C_k$. The sum-rate for RSMA is calculated as $R_{\text{sum}} = \sum_{k \in \mathcal{K}} (\frac{C_k}{N} + R_k)$.

We consider two different strategies to set the threshold for the jamming power. Fig. 2 illustrates the jamming threshold values $J_{l,n}^{thr}$ obtained by the proposed strategies. In Strategy 1, the threshold coefficient ρ is varied proportional to the number of pilots, such that $\rho \propto |S_p|$. Such method ensures that the total jamming power focused on an AU varies with the number of pilots, while the focused jamming power threshold per subcarrier stays the same. In Strategy 2, the threshold coefficient ρ is kept constant irrespective of the number of pilots. In this case, the focused jamming power threshold varies with varying number of pilots, which leads to increased focused jamming power per subcarrier when the number of pilots decreases (or vice versa).

We start our analysis with basic multi-antenna channel models in the form of $\mathbf{h}_{1,n} = [1, 1, 1, 1]^H$, $\mathbf{h}_{2,n} = [1, e^{j\theta}, e^{j2\theta}, e^{j3\theta}]^H$ with $\theta = 4\pi/9$ and $\mathbf{g}_{1,n} = [1, e^{j\beta}, e^{j2\beta}, e^{j3\beta}]^H$ with $\beta = 2\pi/9$, $\forall n \in \mathcal{S}$. Note that the channel is not frequency-selective in the considered model. i.e., the channel frequency response is identical on each subcarrier. The second-order statistics of the channel $\mathbf{g}_{1,n}$ is obtained by averaging the correlation matrix over $\beta \in [0, +4\pi/9]$. Fig. 2 shows the jamming power thresholds for the two strategies described above.

Fig. 3a shows the sum-rate performance of RSMA and SDMA for jamming power threshold according to Strategy 1 as depicted in Fig. 2a. It is observed from the figure that the sum-rate decreases with increasing number of pilot subcarriers to jam as expected. The decrease in the sum-rate performance of SDMA is higher than that of RSMA, implying that RSMA achieves higher sum-rate on the subcarriers used for joint communications and jamming.

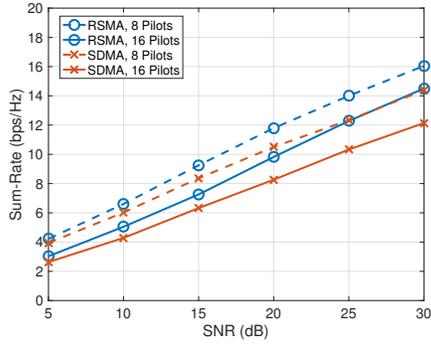


Fig. 4: Sum-rate vs. SNR, strategy 1, urban macro-cell.

Fig. 3b shows the sum-rate performance of RSMA and SDMA with respect to SNR for jamming power threshold set according to Strategy 2 as depicted in Fig. 2b. In contrast to the case in Fig. 3a the sum-rate increases as the number of pilots increases for both RSMA and SDMA. Such behaviour is explained as follows. Recall that the sum-rate is defined as the average of mutual information over all subcarriers. In jamming strategy 2, the portion of the total transmit power allocated to the pilot subcarriers does not change significantly with varying number of pilots. Similarly, the remaining transmit power distributed among the data subcarriers is approximately constant with varying number of data subcarriers. Consequently, the mutual information over data subcarriers decrease with increasing number of data subcarriers. Therefore, the use case with 8 pilot and 24 data subcarriers achieve less sum-rate than the one with 16 pilot and 16 data subcarriers, when the sum-rate is calculated as the average mutual information over all subcarriers. RSMA achieves a significantly higher sum-rate on the subcarriers with joint communications and jamming, thus resulting with an increase in the rate averaged over all subcarriers with increasing number of pilot subcarriers. On the other hand, SDMA achieves a lower rate on those subcarriers, thus the increase being much less significant compared to RSMA.

Next, we investigate the performance under a frequency-selective channel model. We use the Quadriga Channel Generator [19] to generate channels according to the 3GPP Urban Macro-Cell channel model. The channel each of IU have a delay spread of 1200ns and 23 clusters, with each cluster consisting of 20 rays. The channel of the AU have a delay spread of 1200ns with identical numbers of clusters and rays. The OFDM subcarrier spacing is set as 60kHz to observe the frequency-selectivity in the considered channel model. Fig. 4 shows the sum-rate performance of RSMA and SDMA under such channel model. The results show that RSMA has improved sum-rate performance with respect to SDMA under realistic frequency-selective channels.

V. CONCLUSION

We design optimal precoders for joint communications and jamming using RSMA with an MC waveform. We formulate the optimization problem as the maximization of the mutual information with a minimum jamming

power constraint on pilot subcarriers of the AUs while performing data transmission to IUs. We consider the practical case of imperfect CSIT for the IUs and statistical CSIT for the AUs. By simulation results, we show that RSMA can achieve significantly higher sum-rate than SDMA while focusing the same amount of jamming power on AU pilot subcarriers.

REFERENCES

- [1] B. Clerckx, H. Joudeh, C. Hao, M. Dai, and B. Rassouli, "Rate splitting for MIMO wireless networks: A promising PHY-layer strategy for LTE evolution," *IEEE Commun. Mag.*, vol. 54, no. 5, pp. 98–105, May 2016.
- [2] B. Clerckx, Y. Mao, R. Schober and H. V. Poor, "Rate-splitting unifying SDMA, OMA, NOMA, and multicasting in MISO broadcast channel: a simple two-user rate analysis," *IEEE Wireless Commun. Lett.*, vol. 9, no. 3, pp. 349–353, Mar. 2020.
- [3] H. Joudeh and B. Clerckx, "Sum-rate maximization for linearly precoded downlink multiuser MISO systems with partial CSIT: a rate-splitting approach," *IEEE Trans. Commun.*, vol. 64, no. 11, pp. 4847–4861, Nov. 2016.
- [4] Y. Mao, B. Clerckx and V. O. K. Li, "Rate-splitting multiple access for downlink communication systems: bridging, generalizing, and outperforming SDMA and NOMA," *EURASIP J. Wireless Commun. Netw.*, 2018.
- [5] Y. Mao, B. Clerckx, and V. O. K. Li, "Rate-splitting for multi-antenna non-orthogonal unicast and multicast transmission: spectral and energy efficiency analysis," *IEEE Trans. Commun.*, vol. 67, no. 12, pp. 8754–8770, Dec 2019.
- [6] S. Yang, M. Kobayashi, D. Gesbert, and X. Yi, "Degrees of freedom of time correlated MISO broadcast channel with delayed CSIT," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 315–328, Jan. 2013.
- [7] C. Shahriar et al., "PHY-Layer resiliency in OFDM communications: a tutorial," *IEEE Commun. Surveys & Tutorials*, vol. 17, no. 1, pp. 292–314, First quarter 2015.
- [8] T. C. Clancy, "Efficient OFDM denial: pilot jamming and pilot nulling," *Proc. 2011 IEEE Int. Conf. Commun. (ICC)*, Kyoto, Japan, 2011, pp. 1–5.
- [9] R. Miller and W. Trappe, "On the vulnerabilities of CSI in MIMO wireless communication systems," *IEEE Trans. Mobile Comp.*, vol. 11, no. 8, pp. 1386–1398, Aug. 2012.
- [10] C. S. Patel, G. L. Stuber and T. G. Pratt, "Analysis of OFDM/MC-CDMA under channel estimation and jamming," *2004 IEEE Wireless Commun. Net. Conf.*, Atlanta, GA, USA, 2004, pp. 954–958, Vol.2.
- [11] M. Han et al., "OFDM channel estimation with jammed pilot detector under narrow-band jamming," *IEEE Trans. Vehic. Tech.*, vol. 57, no. 3, pp. 1934–1939, May 2008.
- [12] M. Karlsson and E. G. Larsson, "Massive MIMO as a cyber-weapon," *2014 48th Asilomar Conf. Signals, Syst. Comp.*, Pacific Grove, CA, USA, 2014, pp. 661–665.
- [13] M. Karlsson, E. Björnson and E. G. Larsson, "Jamming a TDD point-to-point link using reciprocity-based MIMO," *Trans. Inf. Forensics Security*, vol. 12, no. 12, pp. 2957–2970, Dec. 2017.
- [14] Y. Mao, B. Clerckx and V. O. K. Li, "Rate-splitting for multi-user multi-antenna wireless information and power transfer," *Proc. IEEE 20th Intern. Workshop Sign. Process. Adv. Wireless Commun. (SPAWC)*, Cannes, France, 2019, pp. 1–5.
- [15] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [16] D. P. Palomar, J. M. Cioffi and M. A. Lagunas, "Joint tx-rx beamforming design for multicarrier MIMO channels: a unified framework for convex optimization," *IEEE Trans. Signal Process.*, vol. 51, no. 9, pp. 2381–2401, Sept. 2003.
- [17] C. Xing, N. Wang, J. Ni, Z. Fei and J. Kuang, "MIMO beamforming designs with partial CSI under energy harvesting constraints," *IEEE Signal Process. Lett.*, vol. 20, no. 4, pp. 363–366, April 2013.
- [18] G. G. Raleigh and J. M. Cioffi, "Spatio-temporal coding for wireless communication," *IEEE Trans. Commun.*, vol. 46, no. 3, pp. 357–366, March 1998.
- [19] S. Jaeckel, L. Raschkowski, K. Börner, L. Thiele, F. Burkhardt and E. Eberlein, "QuaDRiGa - quasi deterministic radio channel generator, user manual and documentation," Fraunhofer Heinrich Hertz Institute, Tech. Rep. v2.2.0. 2019.

Adaptive Kernel Kalman Filter

Mengwei Sun, Mike E. Davies, Ian Proudler, James R. Hopgood

Abstract—This paper presents a novel model-based Bayesian filter called the adaptive kernel Kalman filter (AKKF). The proposed filter approximates the arbitrary probability distribution functions (PDFs) of hidden states as empirical kernel mean embeddings (KMEs) in reproducing kernel Hilbert spaces (RKHSs). Specifically, particles are generated and updated in the data space to capture the properties of the dynamical system model, while the corresponding kernel weight vector and matrix associated with the particles' feature mappings are predicted and updated in the RKHS based on the kernel Kalman rule (KKR). We illustrate and confirm the advantages of our approach through simulation, offering detailed comparison with the unscented Kalman filter (UKF), particle filter (PF) and Gaussian particle filter (GPF) algorithms.

Index Terms—Adaptive kernel Kalman filter, kernel Kalman rule, kernel mean embedding, non-linear Bayesian filter

I. INTRODUCTION

Non-linear/non-Gaussian estimation problems in dynamic systems arise in many fields including statistical signal processing, target tracking, satellite navigation, and so on. In order to make inference about a dynamic system, dynamical state-space models (DSMs) are required, including a process model describing the evolution of the hidden states with time, and a measurement model relating the observations to the states. From a Bayesian perspective, the filter for dynamical system inference is designed to estimate the hidden states by recursively computing the posterior probability density function (PDF). Historically, the main focus of sequential Bayesian filters has been on model-based systems where there exists an explicit formulation of the DSM [1]. However, more recently data-driven Bayesian filters have been proposed where the DSM is unknown or partially known but data examples of state-observation pairs are provided [2]. In both scenarios the filters can be broken down into prediction and update stages.

The Kalman filter (KF) provides the optimal Bayesian solution for linear DSMs when both the prediction and posterior distributions are Gaussian. The extended Kalman filter (EKF) is a common form for the application of the KF to nonlinear systems [3], by using the first derivatives to approximate the observation function by a linear system of equations which can cause poor approximation performance when the model is highly non-linear or when the posterior distributions are multimodal. The unscented Kalman filter (UKF), an alternative to

the EKF, was proposed in [4] and uses a weighted set of deterministic particles (so called sigma points) in the state space to approximate the state distribution. Compared with the EKF, the UKF can significantly improve the accuracy of the approximations, but divergence can still occur as in both filters the state distributions are essentially approximated as Gaussian. A more general solution to the non-linear Bayesian filter can be found in the bootstrap particle filter (PF) proposed in [5], in which the hidden state distributions are represented through a weighted set of random particles. Resampling is a necessary step in the bootstrap PF which induces an increase in complexity and is hard to parallelize [6]. To avoid the need for resampling, some specific implementations of the bootstrap PF have been proposed that further approximate the hidden state distribution at each time with a Gaussian, such as the Gaussian particle filter (GPF) [6], and the Gauss–Hermite filter [7].

Different from the approaches above, a number of works have used the recently formulated kernel Bayes rule (KBR) to develop data-driven Bayesian filters based on kernel mean embeddings (KMEs) [2], [8]. Here the unknown measurement model was inferred from prior training data. Owing to the virtue of KMEs, these methods can effectively deal with problems that involve unknown models or strong non-linear structures [9]. However, the feature space for the kernel embeddings remains restricted to the feature space defined by the training data set. Therefore, the performance of these filters relies heavily on there being sufficient similarity between the training data and the test data [10].

Inspired by the KBR [8] and kernel Kalman rule (KKR) [11], we explore the potential of KMEs within full model based filters and introduce a new hybrid filter called the adaptive kernel Kalman filter (AKKF). The main contributions of this paper can be summarized as:

- We derive a new model based Bayesian filter that is a hybrid of kernel based methods and PFs, in which both the prediction and posterior distributions are embedded into a kernel feature space but the known measurement and transition operators are used to calculate the update rules. This is in contrast to the PF where the prediction and posterior distributions are calculated through empirical PDF estimates in the data space.
- The proposed filter can avoid the problematic resampling in most PFs. In passing, we also highlight a missing link between the UKF sigma point method and the kernel conditional embedding operator.

The rest of the paper is set out as follows. The KME and KKR are reviewed in Section II. Section III presents the proposed AKKF. Simulation results for bearing-only tracking (BOT) problem are presented in Section IV and finally conclusions

M. W. Sun, M. E. Davies and J. R. Hopgood are with Institute of Digital Communications, University of Edinburgh, Edinburgh, EH9 3FG, U.K. E-mail: (msun; mike.davies; james.hopgood)@ed.ac.uk.

I. Proudler is with the Centre for Signal & Image Processing (CeSIP), Department of Electronic & Electrical Engineering, University of Strathclyde, Glasgow, G1 1XW, U.K. E-mail: ian.proudlr@strath.ac.uk.

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant number EP/S000631/1; and the MOD University Defence Research Collaboration (UDRC) in Signal Processing.

are drawn in Section V.

II. PRELIMINARIES

In this section, we briefly review the frameworks of the KME and data-driven KKR, see [8] and [11] for details.

A. Kernel Mean Embedding

A reproducing kernel Hilbert space (RKHS) denoted as \mathcal{H}_x on the data space \mathcal{X} with a Kernel function $k_x(x, x')$ is defined as a Hilbert space of functions with the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_x}$ that has some additional properties [8]. The KME approach represents a conditional distribution $P(X|y)$ by an element in the RKHS as,

$$\mu_{X|y} := \mathbb{E}_{X|y} [\phi_x(X)] = \int_{\mathcal{X}} \phi_x(x) dP(x|y). \quad (1)$$

where $\phi_x(x) \in \mathcal{H}_x$ represents the feature mapping of x in RKHS \mathcal{H}_x for all $x \in \mathcal{X}$. $\mu_{X|y}$ is a family of points, each indexed by fixing Y to a particular value y . By defining the conditional operator $C_{X|Y}$ as the linear operator which takes the feature mapping of a fixed value y as the input and outputs the corresponding conditional KME [10], the KME of a conditional distribution defined in (1), under certain conditions, is calculated as,

$$\mu_{X|y} = C_{X|Y} \phi_y(y) = C_{XY} (C_{YY} + \lambda I)^{-1} \phi_y(y). \quad (2)$$

Here, C_{XY} and C_{YY} represent the covariance operators in the tensor product feature spaces $\mathcal{H}_x \otimes \mathcal{H}_y$ and $\mathcal{H}_y \otimes \mathcal{H}_y$, respectively. The term λ is a regularization parameter to ensure that the inverse is well defined.

If instead of access to the true underlying distribution, as required by (1), an empirical estimate of the PDF is available through a particle representation, the KMEs can be estimated directly from these particles. Hence, given the sample set $\mathcal{D}_{XY} = \{(x^{(1)}, y^{(1)}), \dots, (x^{(M)}, y^{(M)})\}$ which are drawn i.i.d. from $P(X, Y)$ with the feature mappings $\Phi := [\phi_x(x^{(1)}), \dots, \phi_x(x^{(M)})]$ and $\Upsilon := [\phi_y(y^{(1)}), \dots, \phi_y(y^{(M)})]$, the estimate of the conditional embedding operator $\hat{C}_{X|Y}$ is obtained as a linear regression in the RKHS [12], as shown in the illustration in Fig. 1. Then, the empirical KME of the conditional distribution is calculated by a linear algebra operation as,

$$\hat{\mu}_{X|y} = \hat{C}_{X|Y} \phi_y(y) = \Phi (G_{YY} + \lambda I)^{-1} \Upsilon^T \phi_y(y) = \Phi \mathbf{w}, \quad (3)$$

$$\mathbf{w} = (G_{YY} + \lambda I)^{-1} G_{:,y}. \quad (4)$$

Here, $G_{YY} = \Upsilon^T \Upsilon$ is the Gram matrix for the samples from the observation variable Y . The input test variable is $y \in \mathcal{Y}$. The kernel weight vector $\mathbf{w} = [w^{(1)}, \dots, w^{(M)}]^T$ includes M non-uniform weights and is calculated based on the vector of kernel functions $G_{:,y} = [k_y(y^{(1)}, y), \dots, k_y(y^{(M)}, y)]^T$. In summary, an empirical KME can represent a PDF over a basis at RKHS with the corresponding weight vector, which has the advantages of low computational cost and low sample complexity.

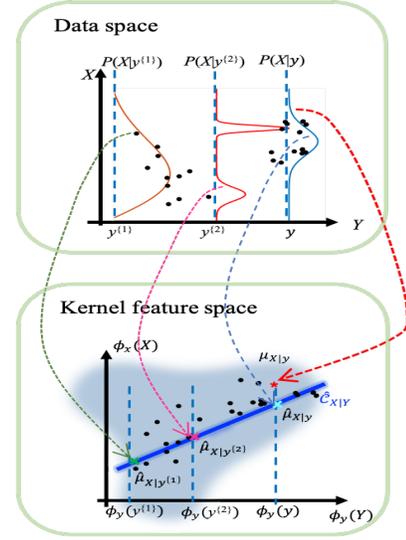


Fig. 1: KME of the conditional distribution $P(X|y)$ is embedded as a point in kernel feature space as $\mu_{X|y} = \int_{\mathcal{X}} \phi_x(x) dP(x|y)$. Given the training data sampled from $P(X, Y)$, the empirical KME of $P(X|y)$ is approximated as a linear operation in RKHS, i.e., $\hat{\mu}_{X|y} = \hat{C}_{X|Y} \phi_y(y) = \Phi \mathbf{w}$. Legend: \cdot samples, \times : empirical KME, $*$: KME.

B. Kernel Kalman Rule

The KKR was proposed in [11] as a recursive least squares estimator for KMEs of posterior distributions. In the proposed empirical KKR [11], the mean embedding and covariance operator are predicted and updated similar to the way a conventional KF does but relying on the training data set $\mathcal{D}_{\tilde{X}XY} = \{(\tilde{x}^{(1)}, x^{(1)}, y^{(1)}), \dots, (\tilde{x}^{(M)}, x^{(M)}, y^{(M)})\}$. Here, $\tilde{x}^{(i)}$ denotes the preceding state of $x^{(i)}$, $i = 1, \dots, M$, and $y^{(i)}$ is the corresponding observation. The feature mappings of training data are represented as $\tilde{\Phi} := [\phi_x(\tilde{x}^{(1)}), \dots, \phi_x(\tilde{x}^{(M)})]$, $\Phi := [\phi_x(x^{(1)}), \dots, \phi_x(x^{(M)})]$ and $\Upsilon := [\phi_y(y^{(1)}), \dots, \phi_y(y^{(M)})]$, respectively. The estimate of the preceding state is given by the KME $\hat{\mu}_{x_{n-1}}^+$ and the covariance operator $\hat{C}_{x_{n-1}, x_{n-1}}^+$. Based on the derivations in [11], the kernel Kalman filter prediction and update steps consist of the following:

$$\hat{\mu}_{x_n}^- = \hat{C}_{X|\tilde{X}} \hat{\mu}_{x_{n-1}}^+, \quad (5)$$

$$\hat{C}_{x_n, x_n}^- = \hat{C}_{X|\tilde{X}} \hat{C}_{x_{n-1}, x_{n-1}}^+ \hat{C}_{X|\tilde{X}}^T + \mathcal{V}. \quad (6)$$

$$\hat{\mu}_{x_n}^+ = \hat{\mu}_{x_n}^- + \mathbf{Q}_n (\phi_y(y_n) - \hat{C}_{Y|X} \hat{\mu}_{x_n}^-), \quad (7)$$

$$\hat{C}_{x_n, x_n}^+ = \hat{C}_{x_n, x_n}^- - \mathbf{Q}_n \hat{C}_{Y|X} \hat{C}_{x_n, x_n}^-. \quad (8)$$

Here, the conditional embedding operators for the distributions $P(X|\tilde{X})$ and $P(Y|X)$, represented by $\hat{C}_{X|\tilde{X}}$ and $\hat{C}_{Y|X}$, are calculated based on training data as $\hat{C}_{X|\tilde{X}} = \Phi (K_{\tilde{x}\tilde{x}} + \lambda_{\tilde{K}})^{-1} \tilde{\Phi}$ and $\hat{C}_{Y|X} = \Upsilon (K_{xx} + \lambda_K)^{-1} \Phi$, respectively. The Gram matrices are $K_{\tilde{x}\tilde{x}} = \tilde{\Phi}^T \tilde{\Phi}$ and $K_{xx} = \Phi^T \Phi$. The covariance of the transition residual matrix is represented as \mathcal{V} , and the kernel Kalman gain operator \mathbf{Q}_n is given by [11],

$$\mathbf{Q}_n = \hat{C}_{x_n, x_n}^- \hat{C}_{Y|X}^T (\hat{C}_{Y|X} \hat{C}_{x_n, x_n}^- \hat{C}_{Y|X}^T + \mathcal{R})^{-1}. \quad (9)$$

where \mathcal{R} is the covariance of the residual of the observation operator.

It should be noted that the existing filters based on the KKR are fully data driven and therefore of use when the DSM is not available and the test data has high similarities to the training data. Data-driven based KKR filters have been used for tracking problems that include the position estimate of a target which follows rotation in a circle or oscillatory rotation [11]. However, the data-driven based filters are only effective when training data provides a good description for the current state, and will fail when the target moves out of the training space. To mitigate this shortcoming, we present a new type of kernel Kalman filter defined for model based scenarios in Section III.

III. ADAPTIVE KERNEL KALMAN FILTER

Inspired by the data-driven based KKR [11] and PF, the proposed adaptive kernel Kalman filter aims to take all the benefits of the KKR and PF. The proposed AKKF is executed in both data space and kernel feature space. In the kernel feature space, the kernel weight vector and positive definite weight matrix are estimated using the KKR, which requires an embedding of the state update function to update the estimate from time $n - 1$ to time n . Then an embedding of the measurement function is used to update the prior estimate at time n to the posterior estimate at time n . In data space, the embeddings for the state process and measurement functions are obtained as follows: a proposal distribution is generated using information from the kernel space at time $n - 1$, which is then propagated through the non-linear DSM. The following subsections will derive the proposed AKKF, with the implementation is summarized in Algorithm 1.

A. Embedding the Posterior Distribution at Time $n - 1$

Let the particles and corresponding kernel feature mappings at time $n - 1$ be represented by $x_{n-1}^{(i=1:M)}$ and $\phi_x(x_{n-1}^{(i=1:M)})$, respectively. Given also the previous weight vector \mathbf{w}_{n-1}^+ and positive definite weight matrix S_{n-1}^+ , the empirical KME and covariance operator for the posterior $p(x_{n-1}|y_{1:n-1})$ are:

$$\hat{\mu}_{x_{n-1}}^+ = \Phi_{n-1} \mathbf{w}_{n-1}^+, \quad (10)$$

$$\hat{C}_{x_{n-1}x_{n-1}}^+ = \Phi_{n-1} S_{n-1}^+ \Phi_{n-1}^T. \quad (11)$$

where the feature mappings is calculated as $\Phi_{n-1} = [\phi_x(x_{n-1}^{(1)}), \dots, \phi_x(x_{n-1}^{(M)})]$ [11]. Specifically, suppose $x_{n-1} = [x_{n-1,1}, \dots, x_{n-1,d}]^T$ is a d -dimension vector and the quadratic kernels utilized. Then, its feature mapping is [13],

$$\phi_x(x_{n-1}) = [x_{n-1,d}^2, \dots, x_{n-1,1}^2, \sqrt{2}x_{n-1,d}x_{n-1,d-1}, \dots, \sqrt{2}x_{n-1,2}x_{n-1,1}, \sqrt{2}x_{n-1,d}, \dots, \sqrt{2}x_{n-1,1}]^T. \quad (12)$$

Therefore, from (10), the empirical KME $\hat{\mu}_{x_{n-1}}^+$ is represented in terms of the expectations of X_{n-1}^2 and X_{n-1} , as $\hat{\mu}_{x_{n-1}}^+ = [\mathbb{E}(X_{n-1}^2), \sqrt{2}\mathbb{E}(X_{n-1}X'_{n-1}), \sqrt{2}\mathbb{E}(X_{n-1}), 1]^T$. Then, the $\mathbb{E}(X_{n-1})$ and $\mathbb{E}(X_{n-1}^2)$ are extracted from $\hat{\mu}_{x_{n-1}}^+$ and passed to the data space. As pointed out in [14], the approximation of a Gaussian distribution is easier to realize than the approximation of an arbitrary non-linear function. Hence, the proposed AKKF uses a new weighted sample representation called

Algorithm 1 Adaptive kernel Kalman filter

Require: DSM: process model and measurement model.

- 1: **Initialization:** Set the initial particles in the data space $\tilde{x}_0^{(i=1:M)} \sim P_{\text{init}}$, $\mathbf{w}_0 = 1/M [1, \dots, 1]^T$.
- 2: **for** $n = 1 : N$ **do**
- 3: **Prediction:**
 - First, in the data space: $x_n^{(i)} = f(\tilde{x}_{n-1}^{(i)}, u_n^{(i)})$,
 - \Rightarrow Second, in the kernel feature space with basis Φ_n : $\mathbf{w}_n^- = \Gamma_{n-1} \mathbf{w}_{n-1}^+$, $S_n^- = \tilde{S}_{n-1}^+ + V_n$.
- 4: **Update:**
 - First, in the data space: $y_n^{(i)} = h(x_n^{(i)}, v_n^{(i)})$,
 - \Rightarrow Second, in the kernel feature space with basis Φ_n : $\mathbf{w}_n^+ = \mathbf{w}_n^- + Q_n (G_{:,y_n} - G_{yy} \mathbf{w}_n^-)$, $S_n^+ = S_n^- - Q_n G_{yy} S_n^-$.
 - $\hat{\mu}_{x_n} = \Phi_n \mathbf{w}_n^+$.
- 5: **Proposal particles draw:**
 - First, in the data space: $\tilde{x}_n^{(i=1:M)} \sim \mathcal{N}(\mathbb{E}(X_n), \mathbb{E}(X_n^2) - \mathbb{E}(X_n) \mathbb{E}(X_n)^T)$.
 - \Rightarrow Second, in the kernel feature space with basis Ψ_n : $\tilde{\mathbf{w}}_n^+ = \Gamma_n \mathbf{w}_n^+$, $\tilde{S}_n^+ = \Gamma_n S_n^+ \Gamma_n^T$.
- 6: **end for**

proposal particles to approximate the KME that can be exactly propagated through the non-linearity. The proposal particles are generated according to the importance distribution as,

$$\tilde{x}_{n-1}^{(i=1:M)} \sim \mathcal{N}(\mathbb{E}(X_{n-1}), \text{Cov}(X_{n-1})), \quad (13)$$

$$\text{Cov}(X_{n-1}) = \mathbb{E}(X_{n-1}^2) - \mathbb{E}(X_{n-1}) \mathbb{E}(X_{n-1})^T. \quad (14)$$

The feature mappings of the proposal particles are represented as $\Psi_{n-1} = [\phi_x(\tilde{x}_{n-1}^{(1)}), \dots, \phi_x(\tilde{x}_{n-1}^{(M)})]$. Then, the posterior distribution $p(x_{n-1}|y_{1:n-1})$ can also be embedded using the new basis Ψ_{n-1} and therefore the weight vector and covariance operator are transformed into Ψ_{n-1} as,

$$\hat{\mu}_{x_{n-1}}^+ = \Psi_{n-1} \tilde{\mathbf{w}}_{n-1}^+, \quad (15)$$

$$\hat{C}_{x_{n-1}x_{n-1}}^+ = \Psi_{n-1} \tilde{S}_{n-1}^+ \Psi_{n-1}^T. \quad (16)$$

Substituting (15) into (10), and (16) into (11), respectively, the proposal kernel weight vector $\tilde{\mathbf{w}}_{n-1}^+$ and matrix \tilde{S}_{n-1}^+ are calculated as,

$$\tilde{\mathbf{w}}_{n-1}^+ = (K_{\tilde{x}\tilde{x}} + \lambda_{\tilde{K}} I)^{-1} K_{\tilde{x}x} \mathbf{w}_{n-1}^+ = \Gamma_{n-1} \mathbf{w}_{n-1}^+, \quad (17)$$

$$\tilde{S}_{n-1}^+ = \Gamma_{n-1} S_{n-1}^+ \Gamma_{n-1}^T, \quad (18)$$

where Γ_{n-1} defined in (17) represents the change of basis from Φ_{n-1} to Ψ_{n-1} , $K_{\tilde{x}\tilde{x}} = \Psi_{n-1}^T \Psi_{n-1}$ represents the Gram matrix of the proposal particles at time $n - 1$, $K_{\tilde{x}x} = \Psi_{n-1}^T \Phi_{n-1}$ is the matrix between the particles and proposal particles at time $n - 1$, and $\lambda_{\tilde{K}}$ is the regularization parameter to modify $K_{\tilde{x}\tilde{x}}$.

B. Prediction from Time $n - 1$ to Time n

The proposal particles at time $n - 1$ are propagated through the process function to achieve the prediction particles, i.e.,

$$x_n^{(i)} = f(\tilde{x}_{n-1}^{(i)}, u_n^{(i)}), \quad i = 1 \dots M. \quad (19)$$

where $u_n^{(i)}$ represents a process noise sample drawn from the process noise distribution. Then, the transitional probability $p(x_n|x_{n-1})$ is embedded using the new basis defined by the feature mappings of the prediction particles $\Phi_n = [\phi_x(x_n^{(1)}), \dots, \phi_x(x_n^{(M)})]$, and is approximated as:

$$\begin{aligned} p(x_n|x_{n-1}) &\mapsto \hat{\mu}_{x_n}^- = \Phi_n \mathbf{w}_n^- = \hat{C}_{x_n|x_{n-1}} \hat{\mu}_{x_{n-1}}^+, \\ &= \Phi_n (K_{\tilde{x}\tilde{x}} + \lambda_{\tilde{K}} I)^{-1} K_{\tilde{x}\tilde{x}} \mathbf{w}_{n-1}^+, \end{aligned} \quad (20)$$

where \mathbf{w}_n^- is the prior kernel weight vector and $\hat{C}_{x_n|x_{n-1}}$ represents the empirical transition operator. Next, the empirical predictive covariance operator $\hat{C}_{x_n x_n}^-$ with the corresponding prior kernel weight matrix S_n^- is computed as,

$$\begin{aligned} \hat{C}_{x_n x_n}^- &= \Phi_n S_n^- \Phi_n^T = \hat{C}_{x_n|\tilde{x}_{n-1}} \hat{C}_{x_{n-1} x_{n-1}}^+ \hat{C}_{x_n|\tilde{x}_{n-1}}^T + \mathcal{V}_n, \\ &= \hat{C}_{x_n|\tilde{x}_{n-1}} \Psi_{n-1} \tilde{S}_{n-1}^+ \Psi_{n-1}^T \hat{C}_{x_n|\tilde{x}_{n-1}}^T + \mathcal{V}_n = \Phi_n \tilde{S}_{n-1}^+ \Phi_n^T + \mathcal{V}_n. \end{aligned} \quad (21)$$

Here, \mathcal{V}_n represents the transition residual matrix,

$$\begin{aligned} \mathcal{V}_n &= \frac{1}{M} (\hat{C}_{x_n|\tilde{x}_{n-1}} \Psi_{n-1} - \Phi_n) (\hat{C}_{x_n|\tilde{x}_{n-1}} \Psi_{n-1} - \Phi_n)^T, \\ &= \Phi_n \underbrace{\left[\frac{1}{M} ((K_{\tilde{x}\tilde{x}} + \lambda_{\tilde{K}} I)^{-1} K_{\tilde{x}\tilde{x}} - I) ((K_{\tilde{x}\tilde{x}} + \lambda_{\tilde{K}} I)^{-1} K_{\tilde{x}\tilde{x}} - I)^T \right]}_{V_n} \Phi_n^T, \end{aligned} \quad (22)$$

where V_n is the finite matrix representation of \mathcal{V}_n . Based on (20)-(22), the prior \mathbf{w}_n^- and S_n^- are calculated as,

$$\mathbf{w}_n^- = (K_{\tilde{x}\tilde{x}} + \lambda_{\tilde{K}} I)^{-1} K_{\tilde{x}\tilde{x}} \mathbf{w}_{n-1}^+ = \Gamma_{n-1} \mathbf{w}_{n-1}^+, \quad (23)$$

$$S_n^- = \tilde{S}_{n-1}^+ + V_n. \quad (24)$$

C. Update at Time n

The observation particles are updated based on the observation model as,

$$y_n^{(i)} = g(\tilde{x}_n^{(i)}, v_n^{(i)}), \quad i = 1 \dots M. \quad (25)$$

where $v_n^{(i)}$ represents a measurement noise sample drawn from the measurement noise distribution. The kernel mappings of observation particles in the kernel feature space are $\Upsilon_n = [\phi_y(y_n^{(1)}), \dots, \phi_y(y_n^{(M)})]$. The posterior KME and the corresponding covariance operator are calculated as [11],

$$\hat{\mu}_{x_n}^+ = \Phi_n \mathbf{w}_n^+ = \hat{\mu}_{x_n}^- + Q_n [\phi_y(y_n) - \hat{C}_{y_n|x_n} \hat{\mu}_{x_n}^-], \quad (26)$$

$$\hat{C}_{x_n x_n}^+ = \Phi_n S_n^+ \Phi_n^T = \text{cov}(\phi_x(x_n) - \hat{\mu}_{x_n}^+). \quad (27)$$

where \mathbf{w}_n^+ and S_n^+ represent the posterior kernel weight and matrix, respectively. The kernel Kalman gain operator denoted as Q_n is derived by minimizing the residual error $\hat{C}_{x_n x_n}^+$. According to derivations in [11], Q_n is calculated as,

$$Q_n = \hat{C}_{x_n x_n}^- C_{y_n|x_n}^T (\hat{C}_{y_n|x_n} \hat{C}_{x_n x_n}^- \hat{C}_{y_n|x_n}^T + \mathcal{R})^{-1}. \quad (28)$$

where \mathcal{R} is the covariance matrix of the observation operator residual and is approximated as $\mathcal{R} = \kappa I$. The empirical likelihood operator is calculated as,

$$\begin{aligned} \hat{C}_{y_n|x_n} &= \hat{C}_{y_n x_n} \hat{C}_{x_n x_n}^{-1} = \Upsilon_n (\Phi_n^T \Phi_n + \lambda_K I)^{-1} \Phi_n^T, \\ &= \Upsilon_n (K_{xx} + \lambda_K I)^{-1} \Phi_n^T = \Upsilon_n K_{xx}^{-1} \Phi_n^T. \end{aligned} \quad (29)$$

Here, the Gram matrix of particles at time n is calculated as $K_{xx} = \Phi_n^T \Phi_n$, and λ_K is the regularization parameter to modify the covariance operator K_{xx} . In this paper, λ_K is set to be 0. Substituting (21) and (29) into (28), Q_n can be calculated as,

$$Q_n = \Phi_n S_n^- \Upsilon_n^T (\Upsilon_n S_n^- \Upsilon_n^T + \kappa I)^{-1} = \Phi_n S_n^- \underbrace{(G_{yy} S_n^- + \kappa I)^{-1}}_{Q_n} \Upsilon_n^T. \quad (30)$$

where Q_n is the finite matrix representation of Q_n in terms of the current basis Φ_n . The Gram matrix of the observation at time n is $G_{yy} = \Upsilon_n^T \Upsilon_n$. Then, the updated KME vector and matrix are given by,

$$\hat{\mu}_{x_n}^+ = \Phi_n \mathbf{w}_n^+ = \Phi_n [\mathbf{w}_n^- + Q_n (G_{:,y_n} - G_{yy} \mathbf{w}_n^-)], \quad (31)$$

$$\hat{C}_{x_n x_n}^+ = \Phi_n S_n^+ \Phi_n^T = \hat{C}_{x_n x_n}^- - \Phi_n Q_n G_{yy} S_n^- \Phi_n^T. \quad (32)$$

where the kernel vector of the measurement at time n is $G_{:,y_n} = \Upsilon_n^T \phi_y(y_n)$. Based on the derivations above, the weight mean vector and covariance matrix are finally updated as,

$$\mathbf{w}_n^+ = \mathbf{w}_n^- + Q_n (G_{:,y_n} - G_{yy} \mathbf{w}_n^-), \quad (33)$$

$$S_n^+ = S_n^- - Q_n G_{yy} S_n^-. \quad (34)$$

IV. SIMULATION RESULTS

Bearing-only tracking (BOT) is one of the fundamental problems in target tracking systems. In this section, we report the tracking performance of different filters applied to BOT problems of a single target using a single sensor in a 2-D space. The corresponding dynamical state-space model (DSM) is described by the equations:

$$\mathbf{x}_n = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}_{n-1} + \begin{bmatrix} 0.5 & 0 \\ 1 & 0 \\ 0 & 0.5 \\ 0 & 1 \end{bmatrix} \mathbf{u}_n, \quad (35)$$

$$y_n = \tan^{-1} \left(\frac{\eta_n}{\xi_n} \right) + v_n. \quad (36)$$

Here, n represents time index and $n = 1, \dots, N$. The hidden states are $\mathbf{x}_n = [\xi_n, \dot{\xi}_n, \eta_n, \dot{\eta}_n]^T$, where (ξ_n, η_n) and $(\dot{\xi}_n, \dot{\eta}_n)$ represent the target position and the corresponding velocity in X-axis and Y-axis, y_n is the corresponding observation. The process noise \mathbf{u}_n follows Gaussian distribution $\mathbf{u}_n \sim \mathcal{N}(\mathbf{0}, \sigma_u^2 \mathbf{I}_2)$ and $\sigma_u = 0.001$. Following [6], the prior distribution for the initial state is specified as $\mathbf{x}_0 \sim \mathcal{N}(\mu_0, \mathbf{P}_0)$ with $\mu_0 = [-0.05, 0.001, 0.7, -0.05]^T$ and,

$$\mathbf{P}_0 = \begin{bmatrix} 0.1 & 0 & 0 & 0 \\ 0 & 0.005 & 0 & 0 \\ 0 & 0 & 0.1 & 1 \\ 0 & 0 & 0 & 0.01 \end{bmatrix}.$$

Fig. 2 and Fig. 3 display two representative trajectories and the tracking performance obtained by four filters: UKF, GPF, PF, and the proposed AKKF using a quadratic kernel. The observer is located at $[0, 0]$. The numbers of particles used for

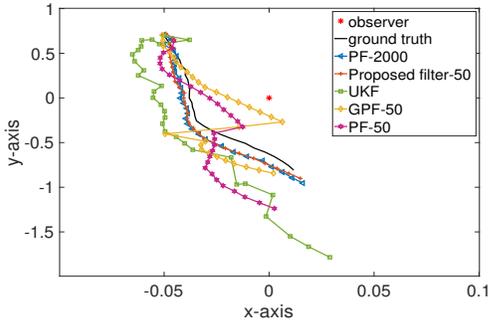


Fig. 2: Trajectory-1: Example of tracking a moving target in two dimensions with the PF, UKF, GPF, and AKKF filters.

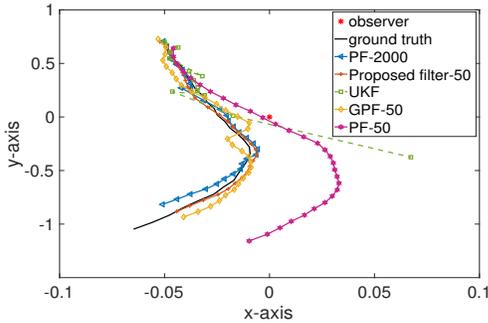


Fig. 3: Trajectory-2: Example of tracking a moving target in two dimensions with the PF, UKF, GPF, and AKKF filters, where UKF diverges.

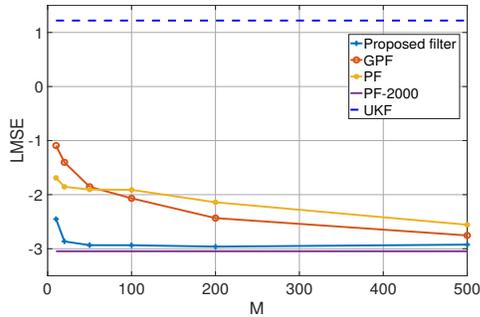


Fig. 4: LMSE performance comparison of the PF, UKF, GPF, and AKKF filters for Trajectory-1.

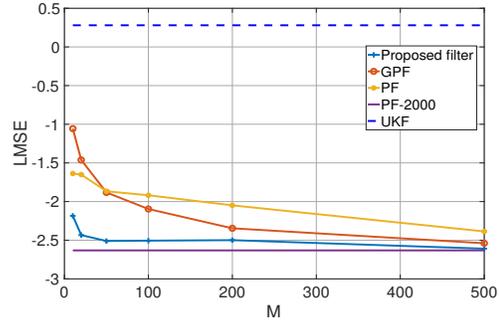


Fig. 5: LMSE performance comparison of the PF, UKF, GPF, and AKKF filters for Trajectory-2.

V. CONCLUSIONS

This paper provided a novel model based kernel Kalman filter. By embedding the probabilities into kernel spaces, more feature information of the hidden states and observations can be captured and recorded. Therefore, the proposed AKKF outperforms existing algorithms when applied to a BOT problem.

REFERENCES

- [1] M. S. Grewal, A. P. Andrews, and C. G. Bartone, *Kalman Filtering*, 2020, pp. 355–417.
- [2] M. Kanagawa, Y. Nishiyama, A. Gretton, and K. Fukumizu, “Filtering with state-observation examples via kernel monte carlo filter,” *Neural Computation*, vol. 28, no. 2, pp. 382–444, 2016.
- [3] S. Julier and J. Uhlmann, “Unscented filtering and nonlinear estimation,” *Proceedings of the IEEE*, vol. 92, no. 3, pp. 401–422, 2004.
- [4] S. J. Julier and J. K. Uhlmann, “New extension of the Kalman filter to nonlinear systems,” in *Signal Processing, Sensor Fusion, and Target Recognition VI*, I. Kadar, Ed., vol. 3068, International Society for Optics and Photonics. SPIE, 1997, pp. 182 – 193. [Online]. Available: <https://doi.org/10.1117/12.280797>
- [5] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, “A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking,” *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, 2002.
- [6] J. H. Kotecha and P. M. Djuric, “Gaussian particle filtering,” *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2592–2601, 2003.
- [7] I. Arasaratnam, S. Haykin, and R. J. Elliott, “Discrete-time nonlinear filtering algorithms using gauss–hermite quadrature,” *Proceedings of the IEEE*, vol. 95, no. 5, pp. 953–977, 2007.
- [8] L. Song, K. Fukumizu, and A. Gretton, “Kernel embeddings of conditional distributions: A unified kernel framework for nonparametric inference in graphical models,” *IEEE Signal Processing Mag.*, vol. 30, no. 4, pp. 98–111, Jun. 2013.
- [9] M. Kanagawa, Y. Nishiyama, A. Gretton, and K. Fukumizu, “Monte carlo filtering using kernel embedding of distributions,” in *Proc. Twenty-Eighth AAAI Conference on Artificial Intelligence (AAAI)*, 2014.
- [10] K. Fukumizu, L. Song, and A. Gretton, “Kernel bayes’ rule: Bayesian inference with positive definite kernels,” *J. Mach. Learn. Res.*, vol. 14, no. 1, pp. 3753– 3783, Dec. 2013.
- [11] G. Gebhardt, A. Kupcsik, and G. Neumann, “The kernel kalman rule,” *Machine Learning*, pp. 2113–2157, Jun. 2019.
- [12] S. Grünewälder, G. Lever, L. Baldassarre, S. Patterson, A. Gretton, and Massimiliano, “Conditional mean embeddings as regressors,” in *Proc. the 29th International Conference on International Conference on Machine*, vol. 1, 2012, pp. 1803–1810.
- [13] K. Muandet, K. Fukumizu, B. Sriperumbudur, and B. Schölkopf, “Kernel mean embedding of distributions: A review and beyond,” *Foundations and Trends in Machine Learning*, vol. 10, no. 1-2, pp. 1–141, 2017.
- [14] S. Julier and J. K. Uhlmann, “A general method for approximating nonlinear transformations of probability distributions,” Tech. Rep., 1996.

PF, GPF and AKKF are 50, while the benchmark performance is given by a PF with 2000 particles. Fig. 4 and Fig. 5 shows the average logarithmic mean square error (LMSE) obtained for 100 random realizations of trajectory-1 and trajectory-2 as a function of particle number denoted as M . From the simulation results, we can arrive at the following conclusions. First, for BOT problems, the tracking performance of PF, GPF and AKKF is obviously better than UKF which shows divergence for trajectory-2 as shown in Fig. 3. Second, the proposed AKKF shows significant improvement compared to the PF and GPF with small particle numbers, as shown in Fig. 4 and Fig. 5.

Detection of Malicious Intent in Non-cooperative Drone Surveillance

Jiaming Liang[†], Bashar I. Ahmad[‡], Mohammad Jahangir[§] and Simon Godsill[†]

[†]Engineering Department, University of Cambridge, UK

[‡]Aveillant, Thales Land and Air Systems, Cambridge, UK

[§]Department of Electronic, Electrical and Systems Engineering, University of Birmingham, UK

Emails: jl809@cam.ac.uk bashar.ahmad@aveillant.com m.jahangir@bham.ac.uk sjg30@cam.ac.uk

Abstract—In this paper, a Bayesian approach is proposed for the early detection of a drone threatening or anomalous behaviour in a surveyed region. This is in relation to revealing, as early as possible, the drone intent to either leave a geographical area where it is authorised to fly (e.g. to conduct inspection work) or reach a prohibited zone (e.g. runway protection zones at airports or a critical infrastructure site). The inference here is based on the noisy sensory observations of the target state from a non-cooperative surveillance system such as a radar. Data from Aveillant’s Gamekeeper radar from a live drone trial is used to illustrate the efficacy of the introduced approach.

Index Terms—Bayesian inference, drone, intent prediction, Kalman filtering, non-cooperative surveillance, radar

I. INTRODUCTION

Small unmanned air systems (sUASs) or drones are poised to play an increasingly important role in various fields given their high potential benefits, for example in agriculture, e-commerce, media, inspection and maintenance, to name a few. The trends in UAS technologies suggest that the low altitude airspace could be populated by different drone platforms in the near future. This motivated large initiatives such as the EU Single European Sky ATM Research (SESAR) Joint Undertaking programme [1] to enable complex drone operations with a high degree of automation. This encompasses ground-based technologies for real-time Unmanned Traffic Management System (UTMS). However, the threat drones can pose to the safety and security of the public (e.g. operating in the presence of manned aviation) is widely recognised and non-cooperative drone surveillance is hence key for UMTS.

Several civilian ground UAS surveillance technologies have emerged, where typically only radar sensors can detect sUASs at long distances and wide ranges [2]–[4], unlike for example electro-optical/infrared and acoustic ones. Nonetheless, they often lack the means to determine which of the detected drones require the operator’s (urgent) attention or further scrutiny or even necessitate triggering an alarm (e.g. closing the airspace). This is particularly critical in the future as scenes become more congested with the commercial use of sUAS and UTMS operators can be easily overwhelmed.

In this paper, we attempt to address from a Bayesian perspective the problem of predicting, as early as possible,

anomalous or threatening behaviour of a target based on noisy sensory observations, for instance from a radar. Anomaly or threat here is defined as the UAS intending to leave (maliciously or not) a geographic region which it is authorised to fly in (e.g. to carry out surveillance or inspection tasks) or reach prohibited area(s) such as no-fly zones near airports [5]. Real radar data is used to demonstrate the performance of the proposed approach. Such *predictive* capability can be crucial for smart UMTS solutions because it facilitates automated decision making and prioritisation of potential threats for effective resources allocation, e.g. operator’s attention or secondary systems. It can also circumvent the need for human-intensive inspection and provide sufficient time to avoid triggering drastic measures (e.g. warning a hobbyist drone operator instead of closing the airspace).

A. Problem Formulation

Let \mathcal{A} be the authorised flying zone for a sUAS and \mathbf{m}_n at time instant t_n be the drone state (e.g. its location in 3-D or any other spatio-temporal characteristics) measured by a non-cooperative surveillance sensor (e.g. radar). The objective in this paper is to *sequentially* estimate the probability, $\Pr(\text{Des} = \mathcal{A} | \mathbf{m}_{0:n})$, of \mathcal{A} being the intended destination of the UAS, i.e. the drone intends to remain in this area. The observation sequence $\mathbf{m}_{0:n}$ is made successively at ordered time instants $\{t_0, t_1, \dots, t_n\}$. Similarly, \mathcal{A} can be defined as a prohibited flying region for the UAS and the task becomes monitoring the probability of the drone aiming to reach it. With the proposed intent inference approach a number of areas, $\mathcal{A}_i, i = 1, 2, \dots, M$, can be simultaneously considered and no restriction on the region(s) shape is imposed in principle.

B. Related Work and Contributions

In recent years there has been a surging interest in predicting intent (such as destination or future actions) of tracked objects owing to its potential applications in smart navigation, robotics, etc. Some methods, dubbed data-driven for instance [6]–[8], capitalise on the availability of sufficiently large data to train a prediction model (e.g. a neural network). However, such training data may not be always available and its collection-labelling can be prohibitively expensive. For instance, data of drones flying within a surveillance area is often limited due to operational constraints. Therefore, the

This work is funded by the Defence Science and Technology Laboratory (DSTL) under the DASA contract DSTLX1000144447.

intent prediction problem in this paper is tackled within a Bayesian object tracking framework with minimal training data requirements, only to parameterise the underlying models. A model-based approach is subsequently designed which can effectively infer the un-observable UAS intent.

Existing model-based intent inference techniques in object tracking are largely focused on: 1) determining the target final destination out of a finite set of nominal endpoints at known locations, e.g. [9]–[13], see [14], [15] for an overview; 2) destination-aware tracking based on reciprocal processes [16] for pre-defined endpoints; 3) long-term trajectory forecasting using motion models learnt in advance, e.g. [17], [18]. However, they either cannot be directly applied to scenarios where the object destination is unknown *a priori*, time-varying and can be anywhere in the considered region (i.e. a continuous intent space), e.g. [9]–[15], or require an off-line training phase for the learning of intent parameters [17], [18].

Compared to the above related work, here we introduce a more natural solution to the considered intent prediction problem which relies on devising a dynamical model to the *latent* intent and connecting it to the target (e.g. UAS) kinematic states via some carefully designed stochastic process. The idea is close in spirit to the virtual leader model [19], [20] and the bulk model [21] for group tracking and has recently been used for destination inference in [22], [23]. More specifically, in this paper we

- introduce a probabilistic framework for the early detection of threatening/malicious drone intent (e.g. exiting or reaching \mathcal{A}) in non-cooperative surveillance contexts and demonstrate its performance using real radar data,
- adopt a new modelling approach based on the virtual leader setting in [20], however not for group object tracking, but for predicting the target intent. The latter can dynamically change over time here and the destination sought by the object/UAS can be an extended spatial region of any shape, located anywhere in the surveyed area. Thus, it does not require the prior knowledge of a set of possible endpoints as in [9]–[15],
- propose a simple, low-complexity, Kalman filter-based inference scheme based on continuous state-space models and asynchronous measurements can be treated.

II. MOTION AND INTENT MODEL

Since the movement pattern of a target is often related to its underlying intent (e.g. reaching or remaining in a given location or area), we consider a dynamical model that represents the object motion under the influence of a latent destination. Here, this hidden destination can dynamically change over time and it is continuous in space. Specifically, the applied model has the following state transition density

$$p(\mathbf{s}_{n+1}|\mathbf{s}_n) = p(\mathbf{x}_{n+1}, \mathbf{r}_{n+1}|\mathbf{x}_n, \mathbf{r}_n) \quad (1)$$

with $\mathbf{s}_n = [\mathbf{x}_n, \mathbf{r}_n]^T$ being the overall state variable at t_n . Here, \mathbf{x}_n consists of object kinematic states such as position, velocity and acceleration while \mathbf{r}_n is a latent variable for the target/drone intent. Models with the structure depicted by (1)

have been widely studied and appeared under various guises in the object tracking field. Examples include the virtual leader model, e.g. [19], [20], the bulk model [21], and the latent destination model [23]. In this paper, we adopt the following linear stochastic differential equations (SDEs) for the object motion $\mathbf{x}_t = [x_t, \dot{x}_t]^T$ and the intended destination location $\mathbf{r}_t = r_t^x$ in a 1-dimensional Cartesian coordinate system

$$d\dot{x}_t = \{\eta_x(r_t^x - x_t) - \rho_x \dot{x}_t\}dt + \sigma_x dB_t \quad (2)$$

$$dr_t^x = \sigma_r dB_t \quad (3)$$

with η_x and ρ_x being positive constants. B_t is a 1-dimensional standard Brownian motion and σ_x, σ_{r_x} are the diffusion constants. According to (2), the position and velocity of the object will revert towards the location of the latent destination at r_t^x and 0, respectively. The zero-reverting part $-\rho_x \dot{x}_t$ serves to prevent the object velocity from becoming excessively high. Consequently, (1) can be factorised as

$$p(\mathbf{s}_{n+1}|\mathbf{s}_n) = p(\mathbf{x}_{n+1}|\mathbf{x}_n, \mathbf{r}_n)p(\mathbf{r}_{n+1}|\mathbf{r}_n)$$

This model can be readily adapted to scenarios where the object has a fixed final destination by setting $\sigma_r = 0$. It is also possible to include higher-order kinematics, for instances, $\mathbf{s}_t = [x_t, \dot{x}_t, r_t^x, \dot{r}_t^x]^T$ with the velocity \dot{r}_t^x the object velocity reverts to being part of the state. In this case, the model bears a resemblance to the group virtual leader model [20] with the number of objects being one. Moreover, the presented model can be used for estimating the intended destination of a fast-maneuvring object if (2) is, for example, Lévy process-driven [23]. The above SDEs can be formalised in a vector-matrix form as below, with $\mathbf{s}_t = [x_t, \dot{x}_t, r_t^x]^T$,

$$d\mathbf{s}_t = \mathbf{A}\mathbf{s}_t dt + \mathbf{L}dB_t^s \quad (4)$$

$$\mathbf{A} = \begin{bmatrix} 0 & 1 & 0 \\ -\eta_x & -\rho_x & \eta_x \\ 0 & 0 & 0 \end{bmatrix}, \mathbf{L} = \begin{bmatrix} 0 & 0 \\ \sigma_x & 0 \\ 0 & \sigma_r \end{bmatrix}, B_t^s = [B_t, B_t]^T$$

A k -dimensional system model can be obtained readily by cascading several such 1-dimensional models. It can be shown that the solution to (4) is given by:

$$\mathbf{s}_n = F_\tau \mathbf{s}_{n-1} + \int_0^\tau e^{A(\tau-u)} \mathbf{L} dB_u^s$$

with $\tau = t_n - t_{n-1}$ and $F_\tau = e^{A\tau}$. Correspondingly, the state transition density is of a Gaussian form:

$$p(\mathbf{s}_n|\mathbf{s}_{n-1}) = \mathcal{N}(\mathbf{s}_n|F_\tau \mathbf{s}_{n-1}, Q_\tau) \quad (5)$$

where $Q_\tau = \int_0^\tau e^{A(\tau-u)} \mathbf{L} \mathbf{L}^T e^{A(\tau-u)^T} du$.

It is emphasised that with this modelling approach the target destination can: a) dynamically change over time subject to the underlying stochastic model; b) be anywhere within the surveyed region since \mathbf{r}_n is a continuous random variable with a support \mathbb{R}^k , where $k = 1, 2$ or 3 .

III. PROPOSED INTENT PREDICTION METHOD

The pseudo-code of the proposed, simple, intent inference routine at time t_n is provided in Algorithm 1. Next, we detail how this predictor is formulated.

Algorithm 1 Intent Detection Algorithm**Input:** $\mathbf{m}_n, p(\mathbf{s}_{n-1}|\mathbf{m}_{0:n-1})$ **Output:** $p(\mathbf{s}_n|\mathbf{m}_{0:n}), \Pr(\text{Des} = \mathcal{A}|\mathbf{m}_{0:n})$

1. Compute $p(\mathbf{s}_n|\mathbf{m}_{0:n-1})$ via standard Kalman prediction;
2. Compute $p(\mathbf{s}_n|\mathbf{m}_{0:n})$ via standard Kalman correction;
3. Obtain $p(\mathbf{r}_n|\mathbf{m}_{0:n})$ from $p(\mathbf{s}_n|\mathbf{m}_{0:n})$;
4. Compute the prob. of \mathcal{A} being the destination via (7);

A. State estimation

The aim here is to estimate the latent destination density $p(\mathbf{r}_n|\mathbf{m}_{0:n})$ given indirect sensor observations made up to the current time t_n , for instance noisy measurements of a target trajectory in 2-D or 3-D. This can be the drone location produced by the Gamekeeper radar as in Section IV. The observation density is given by

$$p(\mathbf{m}_n|\mathbf{s}_n) = \mathcal{N}(\mathbf{m}_n|G_n\mathbf{s}_n, \Sigma_n) \quad (6)$$

where the observation matrix G_n extracts the Cartesian coordinates from the state vector while Σ_n is the noise covariance matrix. This particular setting immediately permits us to use a Kalman filter for the online state estimation task. As a result, we can easily obtain the desired latent destination posterior density $p(\mathbf{r}_n|\mathbf{m}_{0:n})$ as a marginal of the overall Gaussian distribution $p(\mathbf{s}_n|\mathbf{m}_{0:n})$. Nonetheless, it is worth pointing out that the method proposed here is not limited to linear Gaussian systems. When non-linear and/or non-Gaussian observation model is treated, a particle filtering based scheme can be applied for state (including intent) estimation.

B. Spatial region as the intended destination

Since the posterior density of intent is estimated in this paper as described above, the probability that a target aims to leave (or reach) an extended spatial region (or any point within it) can be readily obtained by evaluating the integral

$$\Pr(\text{Des} = \mathcal{A}|\mathbf{m}_{0:n}) = \int_{\mathcal{A}} p(\mathbf{r}_n|\mathbf{m}_{0:n})d\mathbf{r}_n, \quad (7)$$

where \mathcal{A} is the authorised/prohibited region. The integral in (7) can be easily calculated for a Gaussian (or a mixture of Gaussians) and regions of rectangular (or cuboid) shapes can simplify the evaluation (e.g. using the multivariate cumulative distribution functions). For other distributions and geometric shapes, approximations can be applied and tractability of the integral estimation should be considered. Whilst this will be treated in more detail in future work, it is not expected to lead to drastically different outcomes to approximating the area of interest with L simpler shapes (e.g. rectangles, ellipsoids, etc.) via $\sum_l \int_{\mathcal{A}_l} p(\mathbf{r}_n|\mathbf{m}_{0:n})d\mathbf{r}_n$.

Compared to examining specific spatial points or locations, e.g. in [9], [15], calculating the probability that geographical area(s) or any point/region within it/them in this paper is a more generic and fitting formulation for the tackled problem of early prediction of a threatening or anomalous drone behaviour.

IV. EXPERIMENTAL RESULTS

To show the capability of the proposed intent inference approach real measurements, namely 3-D positional information of the tracked-recognised drone, from Aveillant's Gamekeeper radar is used below. They are for a drone flight from the SESAR SAFIR live demonstrations performed close to the Port of Antwerp, Belgium, see [24]. We specifically consider the task of predicting, as early as possible, when the drone intends to leave an authorised flying zone, if at all. The covariance of the noisy 3-D Cartesian observations from the radar is given by $\Sigma_n = 15^2 I_{3 \times 3}$.

A. Overview of Aveillant Gamekeeper radar

The Aveillant Gamekeeper 16U staring radar system in Figure 1 has been specifically designed for high performance detection-recognition of drones within the range of 7.5km in its current configuration. It differs fundamentally from existing mechanical or electronic-steering scanning radars by continuously illuminating the volume of interest on transmit and forms multiple simultaneous receive beams. This enables continuous dwelling on target and provides exceptionally detailed target specific measurements (e.g. high Doppler resolution) that are utilised by specialised processing techniques for better detection, tracking in 3-D and discrimination of slow, low altitude, and small UAS targets [2], [3].

B. Considered Scenarios

Here, a ‘‘permitted’’ geographical area \mathcal{A} is *artificially* chosen to enable assessing the following two settings:

- *Drone remains in the region:* in this example \mathcal{A} covers all of the six waypoints of the drone trajectory as depicted in Figure 2. It can be seen from the figure that the calculated probability of the drone intending to remain in the authorised area (or equivalently having a destination within it) at each time step is noticeably high. For instance, the introduced prediction algorithm reports probabilities consistently above 0.8 when $\sigma_r = 25$. Although $\sigma_r = 0$ consistently produces probabilities of one, the model in this case assumes a fixed latent destination which renders the proposed approach not capable of effectively handling change in intent (e.g. exist or re-entry as in the next



Fig. 1. Aveillant's Gamekeeper staring radar.

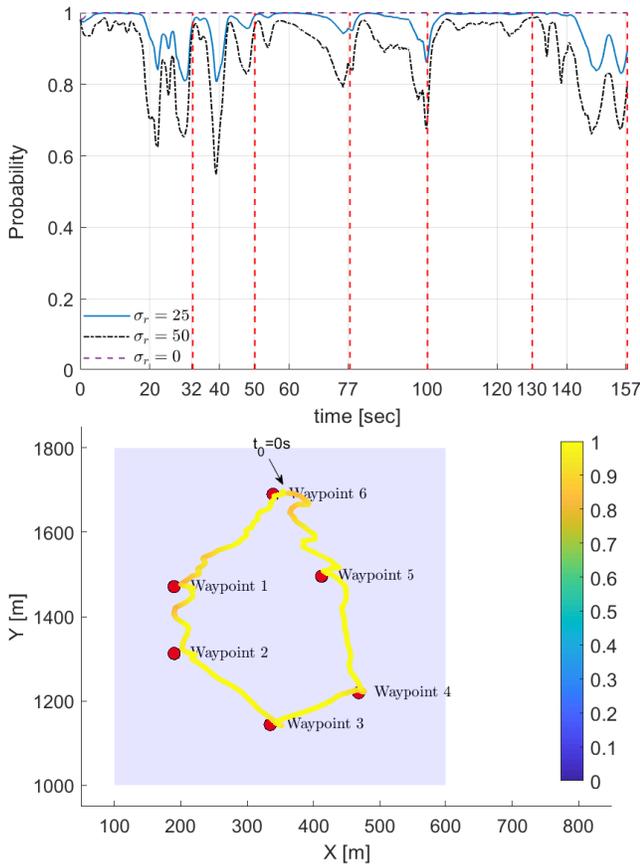


Fig. 2. A drone flying in a permitted zone \mathcal{A} (shaded region). Top: estimated probabilities of the drone intending to stay in \mathcal{A} for various parameters of the proposed approach; vertical lines indicate the timings when the UAS reaches each of its six waypoints. Bottom: filtered target track and its color signifies the probability of UAS remaining in \mathcal{A} for $\sigma_r = 25$; $t_0 = 0s$ is the flight start time.

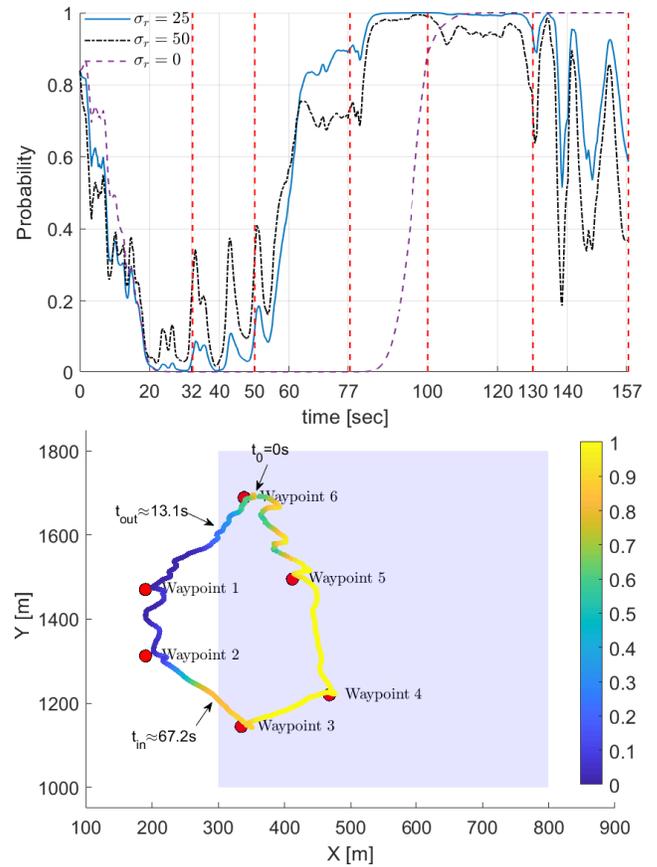


Fig. 3. An example with the drone exiting region \mathcal{A} (shaded region) at $t_{out} \approx 13.1s$ and then re-enters at $t_{in} \approx 67.2s$; $t_0 = 0s$ is the flight start time. Top: estimated probability of the target intent to remain in \mathcal{A} and the six vertical lines indicate when the drone reached each of the six waypoints in its trajectory. Bottom: filtered target track and its color signifies probability of the drone remaining in or entering \mathcal{A} for $\sigma_r = 25$.

example). Further, several fluctuations in the calculated probabilities can be observed. This is to be expected and reflects the fact that the algorithm does respond to a target behaviour that might suggest it is intending to exist \mathcal{A} such as approaching the region boundary (e.g. whilst heading to waypoint 6 towards the end of the flight) or seemingly heading there (e.g. when it nears waypoint 1 and travels between waypoints 1 and 2). At such time instants, the drone velocity can be key to determine the underlying intent; this is seamlessly considered by the proposed inference technique and it is incorporated in the adopted extended motion and intent model.

- *Drone exits and re-enters the region:* in Figure 3 the authorised zone \mathcal{A} is now shifted such that the drone waypoints 1 and 2 are outside \mathcal{A} . It is clear from the figure that the applied predictor can reveal, away from the boundary and with increased confidence, the drone intent to: a) leave \mathcal{A} shortly after take-off at t_0 (i.e. near waypoint 6) where the probability rapidly decreases during this flight section and then reduces to nearly zero as the target exits the permitted region, and b) enter again \mathcal{A} when travelling from waypoint 2 to 3 and quickly

becomes nearly 1 as the UAS re-enters the authorised area. Conversely, region \mathcal{A} can be treated as a prohibited region and the objective is then to detect if the drone aims to reach it. In this case, the predictor can infer the UAS intent to enter this region well before it does (approximately 17.2s in advance from waypoint 2). It can also be noticed that there are several drops (except when $\sigma_r \neq 0$) in probabilities when the drone travelled from waypoint 5 to 6. This is attributed to the drone undertaking manoeuvring turns during this leg of the journey and its velocity can support the possibility that it might leave the area. The algorithm responds accordingly, especially as the drone nears the boundary of \mathcal{A} at waypoint 6. It is noted that such unsmooth track sections can be induced by noisy sensor observations. Again, assuming a fixed latent destination makes the algorithm insensitive to the change of intent; this is obvious from Figure 3 top panel where the predictor with zero destination process noise takes substantially longer time to react to the re-entry intent in comparison to a non-zero σ_r accounting for a dynamically changing intent.

It must be noted that the available short real drone trajectory

(just over two minutes) is for a benign waypoint driven flight. This should be taken into account when assessing the advantages of the presented inference framework, namely how early it can reveal a malicious behaviour should be viewed relative to the corresponding time between a waypoint and the drone exist/entry location to \mathcal{A} . Whilst remarkably early intent predictions can be delivered by the introduced algorithms for longer UAS threatening tracks, this will require further evaluation from representative data for such scenarios; a topic that can be addressed in future work.

The above two examples clearly illustrate the efficacy of the proposed approach, which can give additional time for the non-cooperative surveillance system to take timely action to minimise the impact of a drone threatening or anomalous behaviour. This includes warning the drone operator (if possible) and employing secondary systems. On the other hand, alternative simple methods (e.g. those based on the observed drone's positions and proximity to a region) do not have a predictive capability and typically report a problem when the drone has already left (or reached) \mathcal{A} . They can also be prone to errors due to inaccurate sensor observations, unlike here where suitable models are applied and inference is done sequentially such that the intent prediction certainty is propagated throughout the target track. Additionally, late threat detection can in some cases lead to drastic measures such as triggering an alarm and closure of airspace.

V. CONCLUSION

A Bayesian inference approach for the detection of drone threatening behaviour in a surveyed region is presented. Whilst it is discussed in the context of radar-based non-cooperative drone surveillance, the introduced framework is agnostic to the employed sensor (i.e. not limited to radar) and/or the target type (i.e. not necessarily a UAS). The intent detection approach can facilitate automated decision making and resource allocation in UMTS. More importantly, it is software-based and can be employed by an extant systems, e.g. at the command and control level. In a future publication, we will report on additional work that we have carried out under this DSTL contract in which waypoints for drones are explicitly detected and learned from the data within a jump particle filter framework, reporting tracking as well as more intent prediction results.

ACKNOWLEDGMENT

Authors thank Aveillant for providing real radar data and supporting this work. We acknowledge DSTL's financial support via DASA (Countering Drones Phase 1) under contract DSTLX1000144447.

REFERENCES

- [1] SESAR JU, "Single European sky ATM research joint undertaking," <http://www.sesarju.eu/>, Accessed: 26/04/2021.
- [2] M. Jahangir and C. J. Baker, "L-band staring radar performance against micro-drones," in *2018 19th International Radar Symposium (IRS)*. IEEE, 2018, pp. 1–10.
- [3] M. Jahangir and J. Baker, "Characterisation of low observable targets with a multi-beam staring radar," in *International Conference on Radar Systems (Radar 2017)*. 2017, IET.
- [4] S. Harman, "Analysis of the radar return of micro-UAVs in flight," in *2017 IEEE Radar Conference (RadarConf)*. IEEE, 2017, pp. 1159–1164.
- [5] Civil Aviation Authority, "Airspace restrictions for unmanned aircraft and drones," <https://www.caa.co.uk/Consumers/Unmanned-aircraft/Our-role/Airspace-restrictions-for-unmanned-aircraft-and-drones>, Accessed: 26/04/2021.
- [6] K. Saleh, M. Hossny, and S. Nahavandi, "Intent prediction of pedestrians via motion trajectories using stacked recurrent neural networks," *IEEE Transactions on Intelligent Vehicles*, vol. 3, no. 4, pp. 414–424, 2018.
- [7] B. Völz, H. Mielenz, I. Gilitschenski, R. Siegwart, and J. Nieto, "Inferring pedestrian motions at urban crosswalks," *IEEE Transactions on Intelligent Transportation Systems*, 2018.
- [8] S. Gaurav and B. D. Ziebart, "Discriminatively learning inverse optimal control models for predicting human intentions," in *International Conference on Autonomous Agents and Multiagent Systems*, 2019.
- [9] B. I. Ahmad, J. K. Murphy, P. M. Langdon, S. J. Godsill, R. Hardy, and L. Skrypchuk, "Intent inference for hand pointing gesture-based interactions in vehicles," *IEEE Transactions on Cybernetics*, vol. 46, no. 4, pp. 878–889, 2016.
- [10] B. I. Ahmad, J. K. Murphy, P. M. Langdon, and S. J. Godsill, "Bayesian intent prediction in object tracking using bridging distributions," *IEEE Transactions on Cybernetics*, vol. 48, no. 1, pp. 215–227, 2018.
- [11] J. Liang, B. I. Ahmad, R. Gan, P. Langdon, R. Hardy, and S. Godsill, "On destination prediction based on Markov bridging distributions," *IEEE Signal Processing Letters*, vol. 26, no. 11, pp. 1663–1667, 2019.
- [12] J. Liang, B. I. Ahmad, and S. Godsill, "Simultaneous intent prediction and state estimation using an intent-driven intrinsic coordinate model," in *2020 IEEE 30th International Workshop on Machine Learning for Signal Processing (MLSP)*, 2020, pp. 1–6.
- [13] L. Vladimirov and S. Maskell, "A SMC sampler for joint tracking and destination estimation from noisy data," in *2020 IEEE 23rd International Conference on Information Fusion (FUSION)*, 2020, pp. 1–8.
- [14] B. I. Ahmad, J. K. Murphy, S. Godsill, P. Langdon, and R. Hardy, "Intelligent interactive displays in vehicles with intent prediction: A Bayesian framework," *IEEE Signal Processing Magazine*, vol. 34, no. 2, pp. 82–94, 2017.
- [15] R. Gan, J. Liang, B. I. Ahmad, and S. Godsill, "Modeling intent and destination prediction within a Bayesian framework: Predictive touch as a usecase," *Data-Centric Engineering*, vol. 1, pp. e12, 2020.
- [16] R. Rezaie and X. R. Li, "Destination-directed trajectory modeling, filtering, and prediction using conditionally markov sequences," *IEEE Transactions on Aerospace and Electronic Systems*, 2020.
- [17] L. M. Millefiori, P. Braca, K. Bryan, and P. Willett, "Modeling vessel kinematics using a stochastic mean-reverting process for long-term prediction," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 52, no. 5, pp. 2313–2330, 2016.
- [18] M. Üney, L. M. Millefiori, and P. Braca, "Data driven vessel trajectory forecasting using stochastic generative models," in *2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 8459–8463.
- [19] R. P. S. Mahler, *Statistical Multisource-Multitarget Information Fusion*. Artech House, Inc., USA, 2007.
- [20] S. K. Pang, J. Li, and S. J. Godsill, "Detection and tracking of coordinated groups," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 47, no. 1, pp. 472–502, January 2011.
- [21] D. J. Salmond and N. J. Gordon, "Group and extended object tracking," in *Signal and Data Processing of Small Targets 1999*, O. E. Drummond, Ed. International Society for Optics and Photonics, 1999, vol. 3809, pp. 284 – 296, SPIE.
- [22] G. Zhou, K. Li, and T. Kirubarajan, "Constrained state estimation using noisy destination information," *Signal Processing*, vol. 166, pp. 107226, 2020.
- [23] R. Gan, B. I. Ahmad, and S. Godsill, "Lévy state-space models for tracking and intent prediction of highly manoeuvrable objects," *IEEE Transactions on Aerospace and Electronic Systems*, 2021, Accepted.
- [24] Project SAFIR, "Safe and flexible integration of initial U-space services in a real environment," <https://www.sesarju.eu/index.php/projects/safir>, Accessed: 26/04/2021.

Modelling bi-static uncertainties in sequential Monte Carlo with the GLMB model

Murat Üney, Alexey Narykov, Jason Ralph, Simon Maskell
 School of Electrical Engineering, Electronics and Computer Science
 University of Liverpool, L69 3BX, Liverpool, UK
 Emails: {M.Uney, A.Narykov, jfralph, smaskell}@liverpool.ac.uk

Abstract—Bi-static sensing, where the transmitter and receiver of sensors are separately located, underlies a wide range of collaborative sensing systems. Bi-static detections generally feature a signal time-of-flight (ToF) and an angle-of-arrival (AoA). The current practice in multi-object tracking uses the bi-static geometry to map these pairs onto a selected coordinate frame and filter the mapped detections with a noisy range-bearing (i.e., a mono-static) sensor model. However, the uncertainties in ToF-AoA pairs are not equivalently captured by this model, and the sensing geometry may result in significant degradation of the modelling accuracy. We introduce bi-static likelihood and false alarm models together with Monte Carlo (MC) computational methods to accurately capture the uncertainties involved and use them within Bayesian filtering. We demonstrate the efficacy of our proposed model in simulations with multiple objects using a sequential MC version of the generalised labelled multi-Bernoulli (GLMB) track filter. We compare the filtering performance with the conventional approximation mentioned above.

I. INTRODUCTION

Situational awareness in both defence and non-defence contexts benefits from separately placed transmitter and receiver nodes. These systems provide spatial diversity that improve the sensing performance [1], [2], flexibility in resource use, and, robustness against failures. Example applications include underwater surveillance using multi-static sonar networks [3]–[5] (including ‘dipping’ sonars and sonobuoys deployed from helicopters), counter rocket-artillery-mortar (CRAM) applications [6], and urban air space control with multi-static radars [7].

Each separately located transmitter/receiver pair in these systems form a bi-static pair as the atomic active sensing unit to consider when processing the signals [8]. In this work, we consider Bayesian filtering of detections from bi-static pairs and estimating the number of objects and their trajectories, i.e., bi-static tracking, motivated by the above mentioned applications.

Coupled with a bi-static detection are estimates of the angle-of-arrival (AoA) and time-of-flight (ToF) of the signal that has led to the detection. The AoA is the angle with which the wavefront of the probing pulse sent by the transmitter has arrived at the receiver after getting reflected by the object of interest, or background reflectors (see, Fig.1). The ToF is the time it took for the probing signal to travel from the transmitter to the reflector and then to the receiver. The product of ToF with the propagation speed is often referred to as the bi-static range.

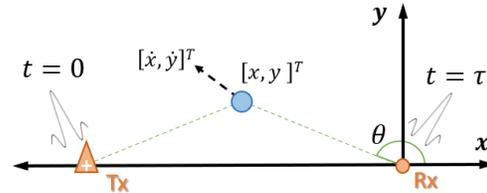


Fig. 1: Bi-static sensing: A transmitter (orange triangle) sends a probing waveform at time zero, a reflector (blue circle) is at $[x, y]^T$ moving with velocity $[\dot{x}, \dot{y}]^T$, and a receiver (orange circle) receives the reflected probe at time τ with wavefront approach angle θ .

Bi-static detections involve uncertainties in these estimates, thus, AoA and ToF are noisy. In addition, some of these detections may be false alarms. Bayesian filtering algorithms take into account these uncertainties through measurement likelihood and clutter models [9], [10]. As a result, the fidelity of these models directly affect the tracking performance.

A common practice in tracking with bi-static detections is to map pairs of AoA and ToF onto the Euclidean plane, i.e., calculate the coordinates of the reflector position in Fig.1, and use range-bearing uncertainty models (see, e.g., [11], [12]). However, conventional likelihood and clutter models in the range-bearing domain are valid for mono-static configurations in which the transmitter and the receiver are co-located. The degradation in modelling accuracy might be small when the reflector is located at a much larger distance compared to the bi-static baseline, i.e., the distance between the transmitter and the receiver. In some applications including underwater sensing, however, such advantageous geometries should not be expected, especially when mobile transmitter and receiver platforms are used [5].

We address modelling of bi-static measurements in Bayesian track filtering by directly relating (ToF, AoA) pairs to the reflector kinematics thereby resolving the discrepancy of the measurement noise model in the above mentioned standard approximation. This also allows us to specify false alarms distribution in the original measurement domain as per the modelling requirements in tracking [13] and avoid false track initiations that stem only from imprecise modelling of false alarms. The result is an endogenous generative model for bi-static detections collected in scans.

The structure of the paper is as follows: In Sec. II we introduce the endogenous bi-static model and compare it

with a standard approximation. We explain how this model is used in multi-object tracking in Sec. III. We demonstrate the efficacy of our approach through a simulation example in Sec. IV, and, finally we conclude.

II. BI-STATIC DETECTIONS AND MODELLING

In a bi-static system, a transmitter located at x_{tx} probes a region of interest with a waveform (Fig. 1). A receiver located at x_{rx} collects reflected signals from the region and processes the front-end signals to detect reflected replicas by matched filtering [14].

Each detection is associated with the total time for the probing waveform to propagate and reach the reflector and be received at the receiver, i.e., ToF. Let us denote this quantity by τ , the total length of the path by \tilde{R} , and the propagation speed by c (assumed constant for simplicity). For a reflector located at x , the ToF is given by

$$\tau = \frac{(\|x_{tx} - x\| + \|x_{rx} - x\|)}{c}, \tag{1}$$

where $\|\cdot\|$ denotes the l_2 vector norm and the numerator $\tilde{R} \triangleq \tau \times c$ is referred to as the bi-static range. Note that this value is lower bounded by the baseline distance between the transmitter and the receiver R_b . Iso-range curves are hence ellipsoids with x_{tx} and x_{rx} at the foci [15], [16].

In this work we assume that the receiver is capable of finding the AoA of the detected signal wavefronts given by

$$\theta = \angle(x - x_{rx}). \tag{2}$$

Some bi-static systems can measure the Doppler shift of the detected signals and distinguish between returns from moving targets of interest and stationary objects. However, the focus of this paper is on settings where such processing is not available.

Measurement likelihood models relate a state vector \mathbf{x} that is often selected as a concatenation of a location x and a velocity vector \dot{x} . The above ToF and AoA equations (1) and (2) specify a bi-static mapping $z = B(x; x_{tx}, x_{rx})$ where $z \triangleq [\tau, \theta]^T$ denotes a bi-static measurement and T is the vector transposition operation. The deviations that lead to errors in τ and θ exhibit the law of large numbers and the errors are normal. Therefore, the likelihood function is given by

$$l(z = (\tau, \theta) | \mathbf{x}; x_{tx}, x_{rx}) = \mathcal{N}(z; B(x; x_{tx}, x_{rx}), \Sigma_B), \tag{3}$$

where the right hand side of the above equation is a bi-variate Gaussian with a covariance of Σ_B . For example, invoking the assumption that the errors in τ and θ are independent leads to a diagonal covariance matrix $\Sigma_B = \text{diag}(\sigma_\tau^2, \sigma_\theta^2)$ where σ_τ^2 and σ_θ^2 are the noise variances in ToF and AoA values, respectively.

The false alarms in Bayesian track filtering algorithms are modelled using a population process on the measurement space which is $\mathcal{Z} = (\tau_{\min}, \tau_{\max}) \times [0, 2\pi)$ where $\tau_{\min} = R_b/c$ is the elapsed time for a transmitted waveform to propagate along the baseline distance from the transmitter directly to the receiver. Often a Poisson distribution with rate λ and uniform

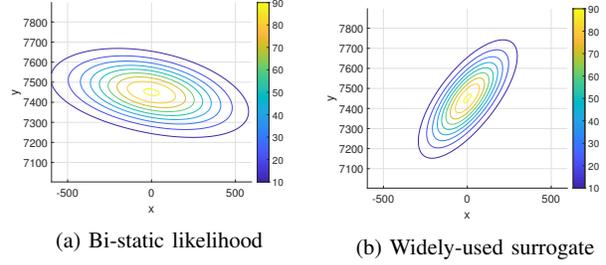


Fig. 2: Likelihoods evaluated for $(\tau = 10\sqrt{2}c, \theta = 3\pi/4)$ and $\sigma_\tau = 0.1s, \sigma_\theta = 1\text{rad}$ over a grid of reflector positions in the $x - y$ plane using (a) the endogenous bi-static model (3) for a transmitter and receiver located at $x = -5c$ and $x = 5c$, respectively, and, (b) the surrogate model (5) for a virtually co-located transmitter/receiver located at $x = 5c$.

spatial distribution over the measurement space \mathcal{Z} are used. Let us denote this distribution by

$$\kappa(z) = \text{Pois}(\lambda, \mathcal{U}_{\mathcal{Z}}), \tag{4}$$

where $\mathcal{U}_{\mathcal{Z}}$ denotes a uniform density over \mathcal{Z} .

The endogenous bi-static model in (1)–(4) directly fits into the dynamic multi-object models used in Bayesian track filtering [17] as the measurement model.

A. Widely used surrogate likelihood

Most work on bi-static tracking replaces the bi-static measurement pair with the corresponding point on the Euclidean plane. Then a virtual co-located transmitter/receiver pair located at x_{vs} measures the range and bearing of this point. However, the uncertainties in the bi-static domain are not transformed. Instead, the likelihood function used in Bayesian filtering implies that the errors of these virtual measurements are assumed to be normal, i.e.,

$$l_S(z | \mathbf{x}) \triangleq \mathcal{N}(B(y; x_{vs}, x_{vs}); B(x; x_{vs}, x_{vs}), \tilde{\Sigma}_S) \tag{5}$$

$$y \triangleq B^{-1}(z = [\tau, \theta]; x_{tx}, x_{rx}),$$

with the inverse transform above explicitly given in the Appendix.

It is instructive to compare the bi-static likelihood in (3) and the widely used surrogate in (5). Both likelihoods use a Gaussian template, however, the domain of the distributions are different because the surrogate model transforms the bi-static measurement to a virtual mono-static sensor’s measurement parameterised on x_{vs} . When x is much more distant than the bi-static baseline, i.e., $\|x - x_{tx}\| \approx \|x - x_{rx}\| \gg \|x_{rx} - x_{tx}\|$, the region of typical measurements in both likelihoods are similar. However, when this condition is not met, there might be large discrepancies. Moreover, the level of accuracy varies with x_{vs} given x_{rx} and x_{tx} making it impractical to optimise when transmitter/receiver platforms are mobile.

We illustrate a comparative example in Fig. 2. We consider a transmitter located at $x_{tx} = [-5c, 0]^T$ and a receiver at

$x_{rx} = [5c, 0]^T$ with c denoting the propagation speed (see, also Fig. 1). We evaluate the endogenous likelihood in (3) for $z = (10\sqrt{2}c, 3\pi/4)$, $\sigma_\tau = 0.1s$, $\sigma_\theta = 1rad$, and a grid of x over which the likelihood evaluates larger compared to other possible values of x . The resulting contour map is given in Fig. 2a. In order to visually reveal the effects of using the surrogate, we repeat the evaluation with (5) for a virtual co-located transmitter/receiver pair at $x_{vs} = [5c, 0]^T$. The resulting contour map is in Fig. 2b. The uncertainty regions in Fig. 2a and Fig. 2b appear similar to ellipses with almost perpendicular minor and major axes which results in significantly different posterior distributions when these likelihoods are used in a Bayes update.

The surrogate likelihood (5) imposes the use of a different clutter distribution than (4). This distribution is (equivalently) selected to have a uniform spatial distribution over $\mathcal{Z}_S = (0, \tau_{\max}) \times (0, 2\pi]$, instead of \mathcal{Z} defined above which leaves out $[0, \tau_{\min}]$ in its domain. In other words, instead of $\kappa(z)$ in (4),

$$\kappa_S(z) = Pois(\lambda, \mathcal{U}_{\mathcal{Z}_S}) \quad (6)$$

is used in filtering bi-static detection with the surrogate model.

III. BI-STATIC UPDATE IN BAYESIAN TRACKING

In this section, we address the utilisation of the proposed bi-static model in multi-object filtering and in particular in sequential Monte Carlo (SMC) generalised labelled multi-Bernoulli filtering (GLMB). The measurement likelihood function of the GLMB filter is introduced in Sec. III-A. Important aspects of the filter implementation that recursively computes the filtering posteriors using this likelihood are highlighted in Sec. III-B.

A. Generalised labelled multi-Bernoulli filter [9]

The GLMB model and recursive Bayesian filtering formulae were introduced in [9] to address estimation of target trajectories in a random finite set (RFS) framework. The computational details of the prediction and measurement update equations of the GLMB filter can be found in [18]. For the sake of simplicity, we shall limit the discussion to the multi-object measurement likelihood function which encapsulates all elements directly affected by the sensor modelling and evaluated in the update step.

A population of objects is described by a multi-object state X in an appropriate state space \mathcal{X} . At time k , each object with state $(\mathbf{x}, \ell) \in X$, comprising a kinematic state \mathbf{x} and a discrete label ℓ described by its state space \mathcal{L} , is detected with probability $p_D(\mathbf{x}, \ell)$, and if so, it produces a measurement whose state is distributed according to a likelihood $l(\cdot|\mathbf{x}, \ell)$. The multi-object observation $Z = \{z_1, \dots, z_{|Z|}\} \in \mathcal{Z}$ is then the union of measurements of detected objects and Poisson distributed false alarms arriving with intensity κ .

Let $h^X \triangleq \prod_{\mathbf{x} \in X} h(\mathbf{x})$ denote multi-object exponential of the real-valued function h . Under assumptions that the measurements originating from object detections are generated independently from each other, and the false alarms are

independent of the detections, the multi-object likelihood is given by [18]

$$g(Z|X) = e^{-\left(\int \kappa(z') dz'\right)} \kappa^Z \sum_{\theta \in \Theta(L(X))} [\psi_Z(\cdot; \theta)]^X, \quad (7)$$

$$\psi_Z(\mathbf{x}, \ell; \theta) = \begin{cases} \frac{p_D(\mathbf{x}, \ell) l(z_{\theta(\ell)}|\mathbf{x}, \ell)}{\kappa(z_{\theta(\ell)})}, & \text{if } \theta(\ell) > 0 \\ 1 - p_D(\mathbf{x}, \ell), & \text{if } \theta(\ell) = 0 \end{cases}. \quad (8)$$

Here, $\theta : \mathcal{L} \rightarrow \{0, 1, \dots, |Z|\}$ is an association map that hypothesises track l generated measurement $\theta(l)$ and imposes the constraint that a single track can generate at most one measurement. p_D is the probability of detection as a function of the target state. It can be selected as zero if \mathbf{x} is on the path between the transmitter and the receiver to account for the receiver being blinded by the direct path signal.

In (7), $\Theta(L(X))$ is the subset of association maps from those labels only in X , i.e., $L(X)$ where $L(\{\mathbf{x}, \ell\}) \triangleq \ell$. Thus, GLMB filtering marginalises out all such hypotheses unlike tracking algorithms that find the most likely global association hypothesis [9], in principle. The implementation approach we adopt from [18] uses Murty's algorithm [19] to approximate (weighted) summations of these terms when computing the filtering posterior with the above likelihood.

In summary, the multi-object measurement likelihood in (7), (8) captures a sensor-specific single-object measurement likelihood, detection and false alarm models. In Section IV, we compare the proposed endogenous model $l(z|\mathbf{x}, \kappa(z))$ with the widely used surrogate model $l_S(z|\mathbf{x}, \kappa_S(z))$ in the tracking performance they result when used in (7), (8).

B. Implementation details

In this paper we are using the SMC (or particle) implementation of the GLMB filter, with details available in [18], [20]. Thus, the filtering update boils down to evaluating the proposed likelihood at the particle points and the false alarm function for the bi-static measurements.

The standard GLMB formulation [18] assumes a priori information on object birth, which is rarely available in practice. Therefore, this paper adopts an adaptive birth model that initiates the birth components from measurement data [21] (for a similar approach, see also [22]).

IV. EXAMPLE

In this section, we provide an example in which there are four objects moving with almost constant speed, and, a mobile transmitter platform and a receiver both following a clock-wise arc as depicted in Fig. 3. The propagation speed is selected as $c = 1490$ m/s which is one of the typical configurations when using underwater sonar. The transmitter is omni-directional and sends the probing waveform to the environment every 20 seconds. The receiver detects reflections with probability $P_D(x) = 0.95$ (independent of the state, for simplicity) and finds ToF and AoA with standard deviations of $\sigma_\tau = 0.1s$ and $\sigma_\theta = 1rad$, respectively. The expected number of false alarms is selected as $\lambda = 5$. We consider 50 scans. The resulting bi-static ToF and AoA measurements are given in Fig. 4a. The

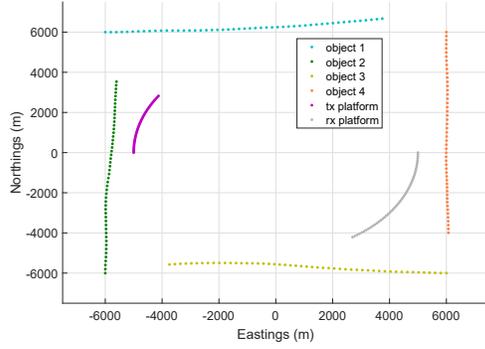


Fig. 3: Multi-object bi-static tracking scenario: Trajectories of the objects, transmitter platform and the receiver platform.

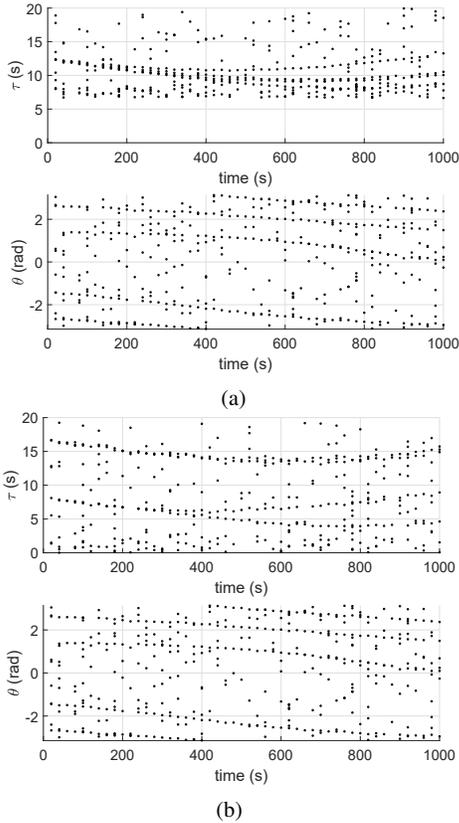


Fig. 4: (a) Bi-static ToF and AoA measurements over time. (b) Virtual mono-static ToF and AoA measurements.

ToF is lower bounded by the time it takes for one transmitted pulse to travel directly from the transmitter to the receiver.

First, we use the proposed endogenous likelihood and false alarm model (1)–(4) in GLMB filtering implemented using SMC as detailed in Section III-B. The trajectories output for the bi-static ToF and AoA measurements in Fig. 4a are given in Fig. 5. The results exhibit a reasonable level of accuracy in estimating the number of trajectories and localisation. It is noted that the track continuity for the horizontal northern trajectory is not as good as that for the other three trajectories.

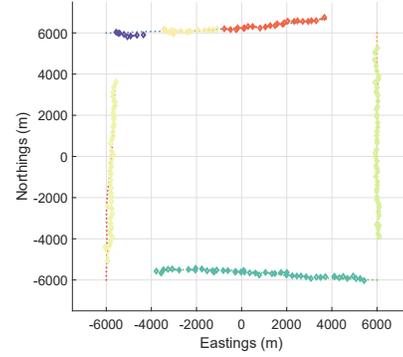


Fig. 5: Estimated trajectories using the endogenous bi-static model in GLMB filtering. Different track labels are depicted with different colours.

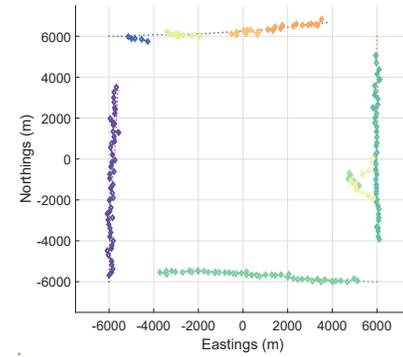


Fig. 6: Estimated trajectories using the surrogate observation model in GLMB filtering. Different track labels are depicted with different colours.

This might be related to the bi-static geometry of this track. Before continuing with the a quantitative assessment of the tracking performance, we present the results obtained using the surrogate model.

The widely-used surrogate model in bi-static processing uses mono-static ToF and AoA measurements as explained in Sec. II. These measurements for a virtual transmitter/receiver pair following the original bi-static receiver’s location are given in Fig. 4b. We filter these measurements using the surrogate model in the GLMB filter. The resulting trajectory estimates are given in Fig. 6. Note that there are false trajectories around the vertical track in the East. The track continuity and localisation accuracy for the horizontal track in the North is also worse compared to that obtained by using the proposed approach.

We make a quantitative comparison of the tracking performances by using the OSPA-on-OSPA, or, OSPA2, distance metric for trajectories [23]. OSPA2 satisfies the metric axioms between arbitrary sets of trajectories and penalises track switches along with location and cardinality errors.

First, we evaluate the OSPA2 between the ground truth in Fig. 3 and the estimates output by the proposed model in Fig. 5, and, depict the results in Fig. 7. Specifically, we perform the evaluation for 5 scan long time windows after

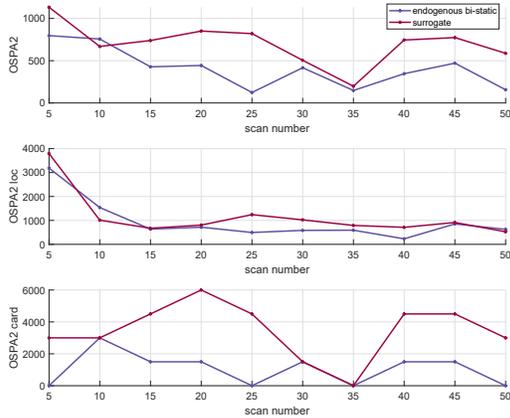


Fig. 7: OSPA2 error metric from the ground truth to the estimated trajectories using the proposed model (blue lines) and the widely-used surrogate (red lines): Total OSPA2 (top), localisation error (mid-pane) and cardinality error (bottom).

partitioning the trajectory sets over time to understand the dynamic behaviour of the errors. The OSPA order and the cut-off parameter used are $p = 1$ and $c = 1500m$, respectively, with the l_2 distance as the base distance. OSPA2 can be decomposed into a localisation error and a cardinality error (i.e., error in the number of trajectories). The total error, localisation error and the are depicted by blue lines in Fig. 7.

Second, we find the error figures for the surrogate model which are given in red lines. The proposed approach consistently performs better than the widely-used surrogate model in terms of localisation accuracy, avoidance to initiating false trajectories and total OSPA2.

V. CONCLUSIONS

In this work, we have considered Bayesian filtering of bi-static detections for multi-object tracking and proposed an endogenous likelihood and false alarm model. The modelling assertion is that the errors in the time-of-flight and angle-of-arrival values associated with bi-static detections are Gaussian, but distinct from the usual approximations used for mono-static detections. The model is amenable for Monte Carlo computational methods in Bayesian tracking. In a simulated example, we have demonstrated that the proposed model outperforms a widely-used standard approximation.

APPENDIX

A. Inverse bi-static mapping B^{-1}

Let us consider the inverse bi-static mapping given by $B^{-1} : (\tau_{min}, \tau_{max}] \times [0, 2\pi) \times \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}^2$ which maps bi-static measurements to points in the Cartesian space given the location of the transmitter and receiver, i.e., $y = B^{-1}(\tau, \theta; x_{tx}, x_{rx})$. Let us denote the bi-static range of the ToF by $R = c \times \tau$, the unit vector pointing at the AoA by $\mathbf{e} \triangleq [\cos(\theta), \sin(\theta)]^T$, and the bi-static baseline by $\mathbf{b} = x_{tx} - x_{rx}$. Then, the point y on the Euclidean plane that induces the bi-static pair (τ, θ) is found as

$$y = x_{rx} + \frac{R^2 - \mathbf{b}^T \mathbf{b}}{2R - 2\mathbf{b}^T \mathbf{e}} \mathbf{e}. \quad (9)$$

ACKNOWLEDGEMENT

The authors would like to thank the UK Defence Science and Technology Laboratories (Dstl, Grant no. 1000143726) for financial support as part of Project BLUE, which is part of the UK MoD University Defence Research Collaboration (UDRC) in Signal Processing.

REFERENCES

- [1] A. M. Haimovich, R. S. Blum, and L. J. Cimini, "MIMO Radar with Widely Separated Antennas," *IEEE Sig. Proc. Mag.*, vol. 25, no. 1, pp. 116–129, 2008.
- [2] Y. Pailhas, Y. Petillot, K. Brown, and B. Mulgrew, "Spatially distributed mimo sonar systems: Principles and capabilities," *IEEE Journal of Oceanic Engineering*, vol. 42, no. 3, pp. 738–751, 2017.
- [3] R. Tharmarasa, T. Kirubarajan, and T. Lang, "Joint path planning and sensor subset selection for multistatic sensor networks," in *2009 IEEE Symp. on Comp. Int. for Security and Defense App.*, 2009, pp. 51–8.
- [4] G. Yang, Y. Li, X. Xiang, and Z. Wang, "Review of development of multi-static sonar for underwater object detection," in *2012 International Conference on Computer Application and System Modeling*, 2012.
- [5] G. De Magistris, M. Uney, P. Stinco, G. Ferri, A. Tesei, K. Le Page, "Selective information transmission using convolutional neural networks for cooperative underwater surveillance," the *Proc. of FUSION'20*, 2020, pp. 1–8.
- [6] J. F. Ralph and J. M. Davies, "Semi-active guidance using event driven tracking," in the *Proc. of FUSION'11*, 2011, pp. 1–7.
- [7] B. Griffin, A. Balleri, C. Baker, and M. Jahangir, "Optimal receiver placement in staring cooperative radar networks for detection of drones," in *2020 IEEE Radar Conference (RadarConf20)*, 2020, pp. 1–6.
- [8] J. Li and P. Stoica, Eds., *MIMO Radar Signal Processing*. John Wiley & Sons, 2009.
- [9] B. -T. Vo and B. -N. Vo, "Labeled random finite sets and multi-object conjugate priors," *IEEE Trans. on Sig. Proc.*, vol. 61, no. 13, pp. 3460–3475, 2013.
- [10] M. Uney, "Type II approximate Bayes perspective to multiple hypothesis tracking," in the *Proc. of FUSION'19*, 2019, pp. 1–8.
- [11] F. Meyer, P. Braca, P. Willett, and F. Hlawatsch, "A scalable algorithm for tracking an unknown number of targets using multiple sensors," *IEEE Trans. on Sig. Proc.*, vol. 65, no. 13, pp. 3478–3493, 2017.
- [12] D. Gaglione, G. Soldi, F. Meyer, F. Hlawatsch, P. Braca, A. Farina, and M. Z. Win, "Bayesian information fusion and multitarget tracking for maritime situational awareness," *IET Radar, Sonar & Navigation*, vol. 14, pp. 1845–1857(12), December 2020.
- [13] B.-N. Vo, M. Mallick, Y. Bar-shalom, S. Coraluppi, R. Osborne, R. Mahler, B. Vo, and J. G. Webster, *Wiley Encyclopedia of Elec. and Electronics Eng.*. J Wiley & Sons, Inc., 2015, ch. Multitarget Tracking.
- [14] H. L. V. Trees, *Detection, Estimation, and Modulation Theory: Radar-Sonar Sig. Proc.*. Krieger Pub. Co., Inc., 1992.
- [15] N. J. Willis, *Bistatic Radar*, 2nd ed. Scitech Publishing, 2005.
- [16] H. Cox, *Underwater acoustic data processing*. Kluwer Academic Publishers, 1989, ch. Fundamentals of bistatic active sonar, pp. 3–24.
- [17] E. Delande, M. Uney, J. Houssineau, and D. E. Clark, "Regional variance for multi-object filtering," *IEEE Transactions on Signal Processing*, vol. 62, no. 13, pp. 3415–3428, 2014.
- [18] B. Vo, B. Vo, and D. Phung, "Labeled random finite sets and the bayes multi-target tracking filter," *IEEE Transactions on Signal Processing*, vol. 62, no. 24, pp. 6554–6567, 2014.
- [19] K. G. Murty, "An algorithm for ranking all the assignments in order of increasing cost," *Oper. Res.*, vol. 16, no. 3, pp. 682–687, 1968.
- [20] B. Ristic, M. Beard, and C. Fantacci, "An overview of particle methods for random finite set models," *Inf. Fusion*, vol. 31, pp. 110–126, 2016.
- [21] B. Ristic, D. Clark, B.-N. Vo, and B.-T. Vo, "Adaptive target birth intensity for PHD and CPHD filters," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 48, no. 2, pp. 1656–1668, 2012.
- [22] S. Lin, B. T. Vo, and S. E. Nordholm, "Measurement driven birth model for the generalized labeled multi-Bernoulli filter," in *2016 Int. Conf. on Cont., Automation and Inf. Sciences (ICCAIS)*. IEEE, 2016, pp. 94–99.
- [23] M. Beard, B. T. Vo, and B. Vo, "A solution for large-scale multi-object tracking," *IEEE Trans. on Sig. Proc.*, vol. 68, pp. 2754–2769, 2020.

Graph Filter Design for Distributed Network Processing: A Comparison between Adaptive Algorithms

Atiyeh Alinaghi, Stephan Weiss, Vladimir Stankovic, Ian Proudler
Dept. of Electronic & Electrical Engineering, University of Strathclyde, Glasgow, UK
 {atiyeh.alinaghi, stephan.weiss, vladimir.stankovic, ian.proudler}@strath.ac.uk

Abstract—Graph filters (GFs) have attracted great interest since they can be directly implemented in a diffused way. Thus it is interesting to investigate GFs to implement signal processing operations in a distributed manner. However, in most GF models, the input signals are assumed to be time-invariant, *static*, or change at a very low rate. In addition to that, the GF coefficients are usually set to be *node-invariant*, i.e. the same for all the nodes. Yet, in general, the input signals may evolve with time and the underlying GF may have parameters dependent on the nodes. Therefore, in this paper, we consider *dynamic* input signals and both types of GF coefficients, *node-variant*, i.e. vary on different nodes, and *node-invariant*. Then, we apply LMS and RLS algorithms for GF design, along with two others called adapt-then-combine (ATC) and combined RLS (CRLS) to estimate the GF coefficients. We study and compare the performance of the algorithms and show that in the case of node-invariant GF coefficients, CRLS gives the best performance with lowest mean-square-displacement (MSD), whereas, for node-variant case, RLS represents the best results. The effect of bias in the input signal has also been examined.

Index Terms—Graph Signal Processing, Graph Filtering, Distributed Processing, Adaptive algorithms

I. INTRODUCTION

In order to infer information from sensor networks, conventional *centralized processing* is often used but may require high transmission power, large communication bandwidth, and costly energy consumption in the central unit. Therefore, a *distributed* approach has been suggested where each node only communicates with its neighbouring nodes, exchanging the information locally, and has its own light processing unit [1]–[4].

A sensor network can be modelled as a graph where the sensors are represented as graph nodes and the inter-sensor communications links are denoted as graph edges. As such, we can apply graph signal processing (GSP) methods, which has emerged recently to extend the classical signal processing concepts to the signals on the vertices of a graph [5]. Spectra of graphs, graph filters (GFs) and other tools to process graph signals are defined accordingly [6], [7]. A practical benefit of GFs is that they may be directly implemented in a distributed manner [8]–[12].

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant number EP/S000631/1 and the MOD University Defence Research Collaboration in Signal Processing.

GSP has been applied to various applications ranging from social media webs, brain activity connections and wireless sensor networks, to power grids and transportation networks [6]. For defence applications, one can imagine situations where a number of sensors, such as sonobuoys or disposable RF connected sensors, are deployed to communicate and infer information based on the observations and measurements. These sensors can be considered as the nodes in a graph and then GSP methods, such as GF, can be applied to analyse the data more efficiently in a distributed fashion.

Graph filtering is one of the core tools in GSP, which can be exploited to represent a transformation of the graph input signals. In the general case, a GF is a transformation matrix mapping the input signal, \mathbf{x} , to the output. Therefore, any linear transformation, $\mathbf{T}(\mathbf{x}) = \mathbf{B}\mathbf{x}$, can be implemented by a GF. Since the structure of the network, the graph topology, is fundamental in distributed processing, it is required to incorporate the graph shift operator, \mathbf{S} (a matrix representing the graph structure) in the GF design. In this paper, we assume that \mathbf{B} is shift invariant with respect to \mathbf{S} and can be represented as a polynomial of \mathbf{S} , i.e. \mathbf{S} is shift enabled [11].

In most models such as [12]–[14], the input signals on the nodes are assumed to be static, i.e. time-invariant, which is not always true. Therefore, in [9], a GF model is introduced which incorporates the time-varying behaviour of the signals. Even though this model is more general, the authors still assume that the GF coefficients are node-invariant, i.e. are the same on different nodes. On the other hand, in [13], node-variant filters are proposed such that the coefficients vary from node to node. It gives more flexibility which enables the design of more general filters with lower orders. Nevertheless, the model is for time-invariant inputs. Therefore, to be more generic, we consider a node-variant GF model, with time-variant inputs.

Having decided on the GF model, we need an algorithm to estimate the GF coefficients. Given the reference system we then use adaptive GF algorithms to estimate the coefficients. First, we consider an adapt-then-combine (ATC) least mean squares (LMS) algorithm which combines the estimated coefficients on the neighbouring nodes [9]. We compare this to an LMS algorithm that operates independently on each node, which is equivalent to ATC without its combine step, and therefore permits a node-variant solution. In addition to

that, we apply the recursive least square (RLS) algorithm in [14] to use it for graph filter design by integrating the graph shift operator (GSO) in the model. As in [14], we also add a combine step to RLS, called combine RLS (CRLS), which makes it comparable to ATC. Ultimately, we implement these approaches and compare their performance with biased and unbiased time-variant input signals.

In the following, we start with a review of graph filter design in Section II. Then in Section III we explain adaptive filters on graphs. Section IV is dedicated to experimental results, while conclusions are drawn in Section V.

II. GRAPH FILTER DESIGN AND IMPLEMENTATION

Various approaches to designing a distributed system are possible. Here we envisage finding the coefficients off-line for the GF design. Then the GF and hence $\mathbf{T}(\mathbf{x})$ can be implemented in a distributed manner.

A. Graph Filter Models

Let us consider an undirected graph \mathcal{G} with a set of N nodes or vertices, \mathcal{N} , and a set of edges or links, \mathcal{E} , such that if node i is connected to node j , then $(i, j) \in \mathcal{E}$. The neighbourhood of node i is defined as the set of nodes $\mathcal{N}_i = \{j | (i, j) \in \mathcal{E}\}$ connected to i including node i itself. For any given graph, an adjacency matrix, \mathbf{A} , is defined as an $N \times N$ square matrix with non-zero elements $a_{ji} \in \mathbb{R}$ if and only if $(i, j) \in \mathcal{E}$. For each graph, there exists a GSO, \mathbf{S} , which can be chosen as \mathbf{A} , the adjacency matrix, or \mathbf{L} , the Laplacian matrix [6], [7]. In this paper we will be using the adjacency matrix as the shift operator.

In GSP applications, one is interested in the analysis of signals on the graph, defined by $\mathbf{x} = [x_1, x_2, \dots, x_N]^T \in \mathbb{R}^N$, where x_k represents the value of the signal at node k . The graph signal at discrete time t is denoted by $\mathbf{x}(t)$. In distributed processing problems, we have a transform, \mathbf{T} , which maps the input graph signals, $\mathbf{x}(t)$, to the desired outputs on the nodes, $\mathbf{y}(t) = \mathbf{T}(\mathbf{x}(t))$. Recall that $\mathbf{T}(\mathbf{x})$ can be represented as a polynomial in the shift matrix. Our goal is to model this transform as a GF.

In this section we start with node-variant GFs with time-varying inputs. We also assume that the shift operation takes a non-negligible amount of time with the same delay as the sampling period of the input signal. In this node-variant case, the GF coefficients are different on each node, k , and can be put in an $M \times N$ matrix \mathbf{D} with each element $\mathbf{D}(m, k) = h_{m,k}^o$, being the optimal m th coefficient on the k th node. Then, an N element vector $\mathbf{h}^{(m)}$ can be defined as the m th row of \mathbf{D} , while the k th column of \mathbf{D} can be symbolized by an $M \times 1$ vector \mathbf{h}_k^o . The filter model can be then described as:

$$\mathbf{y}(t) = \sum_{m=0}^{M-1} \text{diag}(\mathbf{h}^{(m)}) \mathbf{S}^m \mathbf{x}(t-m) + \mathbf{v}(t), \quad (1)$$

where M is the order of the graph filter, and $\mathbf{v}(t) = [v_1(t), v_2(t), \dots, v_N(t)]^T$ is multivariate Gaussian distributed noise which may not be always present, but we include that to be comparable to the model in [9].

For the simplest case where the coefficients are node-invariant, all the N elements of $\mathbf{h}^{(m)}$ will be the same on different nodes, and so $\text{diag}(\mathbf{h}^{(m)}) = \mathbf{I}_N h_m^o$, $h_m^o \in \mathbb{R}$. Thus for the time-invariant input and node-invariant filter, the equation can be written as:

$$\mathbf{y}(t) = \sum_{m=0}^{M-1} h_m^o \mathbf{S}^m \mathbf{x}(t) + \mathbf{v}(t). \quad (2)$$

In other words, the general model (1) becomes time-invariant just by setting $\mathbf{x}(t-m) = \mathbf{x}(t)$ for all m , and node-invariant by setting $h_{m,k}^o = h_m^o$ for all k , enforcing the coefficients to be the same on different nodes as in (2).

III. ADAPTIVE FILTERS ON GRAPHS

Based on (1) for the general node-variant case, the output signal on each node, say k th node, $y_k(t)$, can be modelled separately as:

$$y_k(t) = \sum_{m=0}^{M-1} h_{m,k}^o [\mathbf{S}^m \mathbf{x}(t-m)]_k + [\mathbf{v}(t)]_k, \quad (3)$$

where $[\cdot]_k$ represents the k th row of a vector.

Similar to [9], we define vector $\mathbf{z}(t-m) \triangleq \mathbf{S}^m \mathbf{x}(t-m)$, and then an $M \times 1$ vector $\mathbf{z}_k(t)$ as:

$$\mathbf{z}_k(t) \triangleq \text{col}\{[\mathbf{z}(t)]_k, [\mathbf{z}(t-1)]_k, \dots, [\mathbf{z}(t-M+1)]_k\}. \quad (4)$$

Then (3) can be written alternatively as:

$$y_k(t) = \mathbf{z}_k^T(t) \mathbf{h}_k^o + v_k(t), \quad t \geq M-1. \quad (5)$$

where the optimum \mathbf{h}_k^o is the k th column of \mathbf{D} , and the adaptive filter model is $\mathbf{z}_k^T(t) \mathbf{h}_k$. We then define a global cost function:

$$\mathbf{J}(\mathbf{D}) = \sum_{k=1}^N \mathbf{J}_k(\mathbf{h}_k), \quad (6)$$

where $\mathbf{J}_k(\mathbf{h}_k)$ is the local cost function at node k which can be mean square error for LMS based algorithms:

$$\mathbf{J}_k(\mathbf{h}_k) = \mathbb{E}\{|y_k(t) - \mathbf{z}_k^T(t) \mathbf{h}_k|^2\}, \quad (7)$$

or weighted least squares error function for RLS based algorithms:

$$\mathbf{J}_k(\mathbf{h}_k) = \sum_{i=1}^t \lambda^{t-i} |y_k(i) - \mathbf{z}_k^T(i) \mathbf{h}_k|^2, \quad (8)$$

where λ is a forgetting factor. This optimization problems are to be solved using adaptive methods as described in the following sections.

A. LMS-Based Algorithms

In [9] the ATC approach is proposed to estimate the filter coefficients iteratively:

$$\begin{cases} \boldsymbol{\psi}_k(t+1) = \mathbf{h}_k(t) + \mu_k \mathbf{z}_k(t)(y_k(t) - \mathbf{z}_k^T(t)\mathbf{h}_k(t)), \\ \mathbf{h}_k(t+1) = \sum_{l \in \mathcal{N}_k} c_{lk} \boldsymbol{\psi}_l(t+1), \end{cases} \quad (9)$$

where $\mathbf{h}_k(t)$ is the vector of filter coefficients updated at time t , μ_k the step-size, and c_{lk} are non-negative combination factors:

$$c_{lk} > 0, \quad \sum_{l=1}^N c_{lk} = 1 \quad \text{and} \quad c_{lk} = 0 \quad \text{if} \quad l \notin \mathcal{N}_k. \quad (10)$$

In other words, $\mathbf{C} \triangleq [c_{lk}]$ is a left-stochastic matrix and $\boldsymbol{\psi}_k(t+1)$ is an $M \times 1$ intermediate vector to represent the two step algorithm. As a result, the coefficients will be averaged over the nodes, converging to the same values on different nodes. Thus, the ATC is intended for a problem with node-invariant coefficients.

To modify this algorithm to be node-variant we effectively remove the combine step by setting the combination matrix \mathbf{C} to an identity matrix \mathbf{I} .

B. RLS-Based Algorithms

In this section, we apply the RLS algorithm for use with a graph. We incorporate the structure of the underlying graph of the network by including the GSO, \mathbf{S} . Moreover, we estimate distinctive parameters (GF coefficients) on each node. Even though the RLS algorithm has previously been applied to a sensor network problem in [14], the authors did not include the shift operator in the model.

To perform an RLS algorithm, the weighted sample covariance matrix at each time is required:

$$\hat{\mathbf{R}}_{\mathbf{z}_k}(t) = \sum_{i=1}^t \lambda^{t-i} \mathbf{z}_k(i) \mathbf{z}_k^T(i) + \lambda^t \delta \mathbf{I}_M, \quad (11)$$

where $\mathbf{z}_k(t)$ is the vector defined in (4) which can also be represented as:

$$\mathbf{z}_k(t) \triangleq [\mathbf{x}(t), \mathbf{S}\mathbf{x}(t-1), \dots, \mathbf{S}^{M-1}\mathbf{x}(t-M+1)]^T. \quad (12)$$

The parameter, λ , is the forgetting factor and δ is a small positive number that serves as a regularization parameter. We need the inverse matrix $\mathbf{P}_k(t) = \hat{\mathbf{R}}_{\mathbf{z}_k}^{-1}(t)$ which is not feasible to calculate in a compact form. Therefore, we update it iteratively along with the coefficients. The following algorithm is RLS with a combine step which we call CRLS. This is a modified version of Diffusion BC-RLS algorithm in [14], but with the GSO, \mathbf{S} , incorporated in the regressor, $\mathbf{z}_k(t)$, and no compensation step:

$$\begin{aligned} \mathbf{P}_k(t+1) &= \lambda^{-1} \left(\mathbf{P}_k(t) - \frac{\lambda^{-1} \mathbf{P}_k(t) \mathbf{z}_k(t) \mathbf{z}_k^T(t) \mathbf{P}_k(t)}{1 + \lambda^{-1} \mathbf{z}_k^T(t) \mathbf{P}_k(t) \mathbf{z}_k(t)} \right) \\ \phi_k(t+1) &= \mathbf{h}_k(t) + \mathbf{P}_k(t) \mathbf{z}_k(t) (y_k(t) - \mathbf{z}_k^T(t) \mathbf{h}_k(t)), \\ \mathbf{h}_k(t+1) &= \sum_{l \in \mathcal{N}_k} c_{kl} \phi_k(t+1). \end{aligned} \quad (13)$$

Similar to ATC, CRLS has a combine step with combination matrix \mathbf{C} , which averages the coefficients across the neighbouring nodes. Therefore, it is also a node-invariant algorithm. For the node-variant case we effectively remove the combine step by setting the \mathbf{C} to an identity matrix \mathbf{I} . This results in a node-variant RLS algorithm.

C. Effect of Biased Graph Signals

The data that has been addressed in various other publications [9], [13], [14], are often assumed to be zero-mean. It is possible that the graph signal, $\mathbf{x}(t)$, is not zero mean [9]. Therefore, we briefly investigate the effect that a bias term in the data has on the above algorithms. In this case there is a bias term, $\mathbf{b} \in \mathbb{R}^N$ added to the data, such that $\mathbf{x}_b(t) = \mathbf{x}(t) + \mathbf{b}$. We have $\mathbb{E}\{\mathbf{x}_b\} = \mathbf{b}$, so that, $\mathbf{R}_b = \mathbf{R} + \mathbf{b}\mathbf{b}^H$, where $\mathbf{b}\mathbf{b}^H$ is a rank-one matrix. As a result, the bias term changes the eigenvalues of the covariance matrix, and generally increases both the signal power and the condition number. Since every value in \mathbf{R}_b is larger than the equivalent values in \mathbf{R} , we find that the covariance matrix of non-zero-mean data will contain a larger estimation error [15]. As the bias increases the condition number, it will deform the mean-squared-error (MSE) cost function, and gradient-based methods such as the LMS may converge slower. For RLS-type algorithms, this difference in convergence will be less pronounced due to the gradient correction via the estimate of the inverse covariance matrix.

IV. EXPERIMENTAL RESULTS

Similar to [9], we generated random connected Erdős-Renyi graphs, with $N = 20$ nodes; and example of which is shown in Fig. 1. Using a similar construction as in [9], “this graph was obtained by generating an $N \times N$ symmetric shift matrix, \mathbf{S} , whose entries were governed by Gaussian distribution $\mathcal{N}(0, 1)$ and then threshold edges to be between 1.2 and 1.8 in absolute value to yield an effective probability of an edge $p \approx 0.07$. The edges were soft thresholded by 1.1 to be between 0.1 and 0.7 in magnitude.” The shift matrix, \mathbf{S} , was then normalized by 1.1 times its largest eigenvalue to prevent instability. It ensures that \mathbf{S}^m contracts for increasing m , and hence does not cause stability issues.

In [9], the authors assumed that the graph signal, $\mathbf{x}(t)$, is i.i.d Gaussian with zero-mean and covariance matrix \mathbf{R}_x , where \mathbf{R}_x was chosen as the solution of the Lyapunov equation $\mathbf{S}\mathbf{R}_x\mathbf{S}^T - \mathbf{R}_x + \mathbf{I} = 0$. This is equivalent to generating the signals through a Gauss-Markov model with the shift operator, \mathbf{S} , as its transfer function. The logic behind this choice is that usually in real situations the input signals on the neighbouring nodes are somehow related.

To investigate the effect of bias, we added a DC component to the input signal, which was chosen to be 2.5 times the input variance σ_x^2 . The factor 2.5 is selected based on the convergence behaviour of the algorithms. In order to be comparable to [9], “the noise $\mathbf{v}(t)$ was also set to zero-mean Gaussian with covariance $\mathbf{R}_v = \text{diag}\{\sigma_{v,k}^2\}_{k=1}^N$, where the variances $\sigma_{v,k}^2$ were randomly generated from uniform

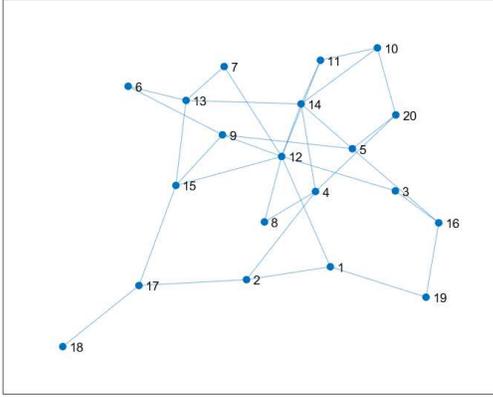


Fig. 1. A random Erdos-Renyi graph with $N = 20$ nodes.

distribution $\mathcal{U}(0.1, 0.15)$. The filter order M was set to 3 and the coefficients $h_{m,k}^o$ of the target filter, i.e $\mathbf{T}(\mathbf{x})$ were randomly drawn from a uniform distribution $\mathcal{U}(0, 1)$."

For LMS and ATC algorithms, the step-size, μ , was chosen as $1/\zeta_{max}$, where ζ_{max} is the maximum eigenvalue of the covariance matrix $\mathbf{R}_{\mathbf{z}_k}$ as in (11) with $i = t$, $\lambda = 1$ and $\delta = 0$. For RLS and CRLS algorithms, the forgetting factor, λ , was selected as 0.9999. The combine factors for ATC [9], and CRLS were set to $c_{lk} = 1/|\mathcal{N}_k|$ for $l \in \mathcal{N}_k$, where $|\cdot|$ denotes the cardinality of its entry.

We ran 200 Monte-Carlo simulations and for each simulation a new random Erdős-Renyi graph was generated. Since the explained procedure does not guarantee the connectivity of the generated random graphs, we discarded all the graphs that were not connected. Along with that, at each simulation, a new input signal was generated based on the new shift matrix. The ground truth filter coefficients were also randomly generated at each run. After each simulation the mean-squared-displacement (MSD), $\mathbb{E}\{\|\mathbf{h}_k^o - \mathbf{h}_k\|^2\}$, between the ground truth coefficients and the estimated parameters were calculated for each node, k , and then after 200 runs, averaged over all the simulations. This MSD metric calculates the deviation of each estimated parameter from the ground truth coefficients and so demonstrates how well a system has been identified.

We set the ground truth coefficients, \mathbf{h}_k^o , to be either node-invariant or node-variant. Then we ran the algorithms for both biased and unbiased input signals. Fig. 2 shows an example MSD of the estimated coefficients on all the nodes, $k = 1, \dots, N$, for one of the coefficients, m , where the ground truth was node-invariant and input was biased. We can see that CRLS and ATC show the best performance with the lowest MSD, while RLS has moderate performance, and LMS results are not very appealing. It is not surprising since ATC and CRLS combine the coefficients at each iteration steps which results in an average value on different nodes, or in other words, it results in node-invariant estimations. The algorithms typically suffer from gradient noise, in particular the LMS type

ones. Additional averaging over the graph dimension helps to reduce its effect.

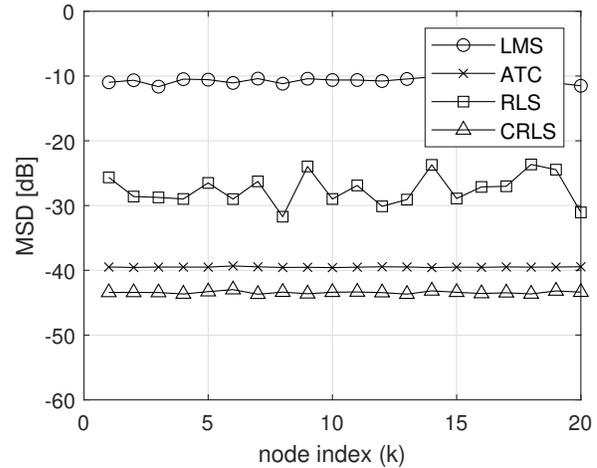


Fig. 2. MSD between the node-invariant ground truth GF coefficients and the estimated values applying, LMS, ATC, RLS, and CRLS algorithms with biased input on all the nodes $k = 1, \dots, N$, averaged over 200 Monte-Carlo simulations in dB.

Table I represents the averaged MSD over all the nodes for the four different scenarios, node-invariant and node-variant ground truths with unbiased and biased inputs. We can see that for node-invariant ground truth, ATC and CRLS show better performance with lower MSD, while LMS shows the worst results. This is reasonable as in the ATC and CRLS there is a combine step which gets the average of the coefficients to be the same on all the nodes. On the contrary, for node-variant case, ATC and CRLS display higher, worse, MSDs, while RLS gives the lowest MSD. It is also noticeable that the performance of LMS for both node-variant and invariant cases are more or less the same. Similar behaviour can also be observed for RLS algorithm. Therefore, for real data and applications, where the nature of underlying GF model is not necessarily known, RLS, can be a better choice since it shows robustness under four different conditions.

By looking at Table I, we can also see that the bias has considerable effect on LMS, increasing its MSD. However, the main impact of the bias is the convergence speed of the error shown in Fig. 3, which is calculated on each node as:

$$e_k(t) = \mathbb{E}\{|y_k(t) - \sum_{m=0}^{M-1} h_{m,k}[\mathbf{S}^m \mathbf{x}(t-m)]_k|^2\}, \quad (14)$$

where $\mathbb{E}\{\cdot\}$ here represents an ensemble average over 200 Monte-Carlo simulations. As seen in Fig. 3, the LMS and ATC algorithms, converge considerably faster when the input signal is unbiased. Bias also increases the convergence time for RLS and CRLS, but it is not that pronounced.

V. CONCLUSION

In this paper a general graph filter (GF) model has been considered, which can be applied for both time-invariant and

TABLE I
MSD VALUES BETWEEN THE GROUND TRUTH AND THE ESTIMATIONS
AVERAGED OVER ALL THE NODES, k , AND COEFFICIENTS, m , AFTER 200
MONTE-CARLO SIMULATIONS IN [dB].

	node-invariant		node-variant	
	unbiased	biased	unbiased	biased
LMS	-22.72	-15.81	-22.26	-15.76
ATC	-39.39	-42.33	(-10.74)	(-8.78)
RLS	-31.38	-32.60	-30.82	-32.79
CRLS	-46.37	-48.03	(-10.91)	(-10.89)

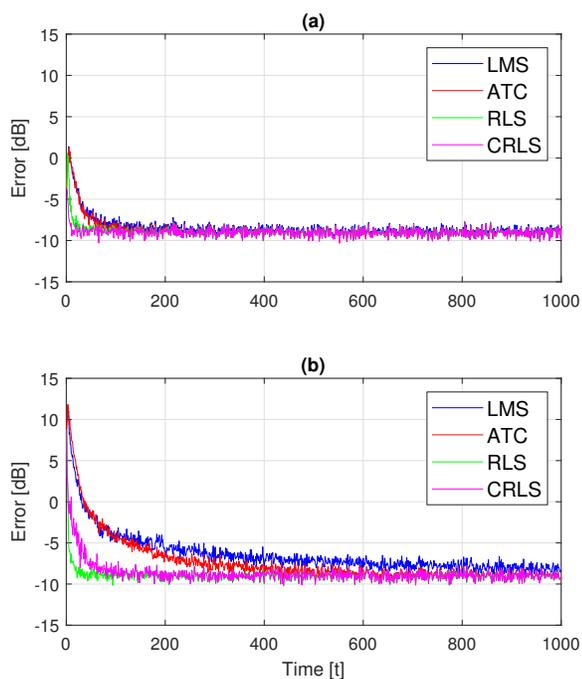


Fig. 3. Error between the desired output and the estimated output after each iteration on a random node, with node-invariant coefficients for (a) unbiased and (b) biased inputs.

time-varying inputs. In other words, it integrates the dynamic behaviour of the input signal. In addition to that, the model is modified to be compatible for both node-invariant, where the parameters are the same on different nodes, and node-variant, where the parameters are node dependent. Then, four different adaptive algorithms, namely LMS, ATC, RLS and, CRLS, have been applied to estimate the model parameters, GF coefficients, by minimizing a cost function. By comparing the algorithms, CRLS has shown the best performance for node-invariant case, while the RLS has given the lowest MSD for node-variant coefficients. The impact of bias in the input signal has also been examined, showing the most deterioration for LMS algorithms. The bias has also shown a considerable

effect on the convergence of LMS and ATC algorithms by decreasing their convergence speed. Having considered all the factors, RLS seems to be a better choice as it shows robust performance under different conditions.

REFERENCES

- [1] C. G. Lopes and A. H. Sayed, "Diffusion least-mean squares over adaptive networks: Formulation and performance analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 7, pp. 3122–3136, 2008.
- [2] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks—part i: Sequential node updating," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5277–5291, 2010.
- [3] —, "Distributed adaptive node-specific signal estimation in fully connected sensor networks—part ii: Simultaneous and asynchronous node updating," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5292–5306, 2010.
- [4] F. de la Hucha Arce, M. Moonen, M. Verhelst, and A. Bertrand, "Distributed adaptive node-specific signal estimation in a wireless sensor network with noisy links," *Signal Processing*, vol. 166, p. 107220, 2020.
- [5] E. Isufi, G. Leus, and P. Banelli, "2-dimensional finite impulse response graph-temporal filters," in *2016 IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2016, pp. 405–409.
- [6] A. Sandryhaila and J. M. F. Moura, "Discrete signal processing on graphs," *IEEE Transactions on Signal Processing*, vol. 61, no. 7, pp. 1644–1656, 2013.
- [7] L. Stankovic, D. P. Mandic, M. Dakovic, I. Kisil, E. Sejdic, and A. G. Constantinides, "Understanding the basis of graph signal processing via an intuitive example-driven approach [lecture notes]," *IEEE Signal Processing Magazine*, vol. 36, no. 6, pp. 133–145, 2019.
- [8] R. Nassif, C. Richard, J. Chen, and A. H. Sayed, "A graph diffusion lms strategy for adaptive graph signal processing," in *2017 51st Asilomar Conference on Signals, Systems, and Computers*, 2017, pp. 1973–1976.
- [9] —, "Distributed diffusion adaptation over graph signals," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018, pp. 4129–4133.
- [10] D. G. Tiglea, R. Candido, and M. T. M. Silva, "An adaptive sampling technique for graph diffusion lms algorithm," in *2019 27th European Signal Processing Conference (EUSIPCO)*, 2019, pp. 1–5.
- [11] L. Chen, S. Cheng, V. Stankovic, and L. Stankovic, "Shift-enabled graphs: Graphs where shift-invariant filters are representable as polynomials of shift operations," *IEEE Signal Processing Letters*, vol. 25, no. 9, pp. 1305–1309, 2018.
- [12] M. Coutino, E. Isufi, and G. Leus, "Advances in distributed graph filtering," *IEEE Transactions on Signal Processing*, vol. 67, no. 9, pp. 2320–2333, 2019.
- [13] S. Segarra, A. G. Marques, and A. Ribeiro, "Optimal graph-filter design and applications to distributed linear network operators," *IEEE Transactions on Signal Processing*, vol. 65, no. 15, pp. 4117–4131, 2017.
- [14] A. Bertrand and M. Moonen, "Distributed adaptive estimation of node-specific signals in wireless sensor networks with a tree topology," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2196–2210, 2011.
- [15] C. Delaosa, J. Pestana, N. J. Goddard, S. Somasundaram, and S. Weiss, "Sample space-time covariance matrix estimation," in *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2019, pp. 8033–8037.
- [16] Yinman Lee and Wen-Rong Wu, "An lms-based adaptive generalized sidelobe canceller with decision feedback," in *IEEE International Conference on Communications, 2005. ICC 2005. 2005*, vol. 3, 2005, pp. 2047–2051 Vol. 3.
- [17] F. Hua, R. Nassif, C. Richard, H. Wang, and A. H. Sayed, "A preconditioned graph diffusion lms for adaptive graph signal processing," in *2018 26th European Signal Processing Conference (EUSIPCO)*, 2018, pp. 111–115.
- [18] E. Isufi, A. Loukas, A. Simonetto, and G. Leus, "Filtering random graph processes over random time-varying graphs," *IEEE Transactions on Signal Processing*, vol. 65, no. 16, pp. 4406–4421, 2017.
- [19] S. Haykin, *Adaptive Filters*. NJ, Englewood Cliffs: Prentice-Hall, 1995.

Approximate Proximal-Gradient Methods

Anis Hamadouche, Yun Wu, Andrew M. Wallace, and João F. C. Mota

Abstract—We study the convergence of the *Proximal-Gradient* algorithm for convex composite problems when both the gradient and the proximal mapping are computed approximately. This scenario occurs when the gradient is computationally expensive and the proximal operator is not available in closed form and may be computed only up to a certain fixed precision. We establish tight deterministic bounds and propose new probabilistic upper bounds on the suboptimality of the function values along the iterations under some statistical assumptions on the perturbed iterates. We use the *Proximal-Gradient* algorithm to solve randomly generated LASSO problems while varying the fixed-point machine representation and the proximal computation precision.

Index Terms—Convex Optimization, Proximal Gradient, Approximate Algorithms.

I. INTRODUCTION

Many problems in statistics, machine learning, and engineering can be posed as *composite optimization problems*:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) := g(x) + h(x), \quad (1)$$

where $x \in \mathbb{R}^n$ is the optimization variable, $g : \mathbb{R}^n \rightarrow \mathbb{R}$ a differentiable convex function, and $h : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is a closed, proper, and convex function, which is not necessarily differentiable but which enables the inclusion of constraints into (1).

An important example is empirical risk minimization (ERM), the foundational framework in machine learning. There, $g(x) = (1/m) \sum_{i=1}^m \ell(w(z_i; x), y_i)$, where $\{(z_i, y_i)\}_{i=1}^m$ is a collection of training feature vectors z_i and associated labels y_i that we wish to fit with a parametric function $w(\cdot, x)$, and $h(x)$ is a regularizer on the parameters x , e.g., a norm of x . A concrete example of this framework is logistic regression [1].

Another example is compressed sensing [2], in which one attempts to reconstruct a sparse vector $x^* \in \mathbb{R}^n$ from linear measurements $y = Ax^*$, where $A \in \mathbb{R}^{m \times n}$ has more columns than rows, i.e., $m < n$. One way to achieve this is by solving (1) with $g(x) = \|Ax - y\|_2^2$ and $h(x) = \|x\|_1$.

Finally, composite problems like (1) arise in control applications, for example in the control of the trajectory of a drone, in which x encodes both a state-vector (e.g., position and velocity of a drone) and the input (e.g., the acceleration in a given direction and steering). In this case, g often encodes a final goal for the state-vector as well as energy penalties,

Work supported by UK's EPSRC (EP/T026111/1, EP/S000631/1), and the MOD University Defence Research Collaboration.

Anis Hamadouche, Yun Wu, Andrew M. Wallace, and João F. C. Mota are with the School of Engineering & Physical Sciences, Heriot-Watt University, Edinburgh EH14 4AS, UK. (e-mail: {ah225,y.wu,a.m.wallace,j.mota}@hw.ac.uk).

while h encodes state-space dynamics and control constraints [3].

Resource-constrained platforms. Most algorithms that solve (1) assume that computations can be performed with infinite (or near-infinite) precision. While such precision can be achieved in standard computation devices, power-efficient platforms like FPGAs, which are commonly deployed in battery-operated equipment, have much lower precision. Solving problems like (1) under these scenarios often requires completely new strategies [4]. For example, if we solve (1) with standard algorithms, e.g., proximal-gradient or interior-point methods, the resulting solution will satisfy the finite precision constraints of the computing machine rather than infinitely precise solutions satisfying optimal convergence rates. Early termination of iterative algorithms and reduced precision (RP) via finite precision arithmetic can save computational time or power while tolerating losses in accuracy in resource-constrained systems. Furthermore, many optimization software solvers are approximate and this must be accounted for in convergence analysis [5].

Problem statement. The aforementioned approximation techniques come at a cost of reduced accuracy and increased algorithmic perturbations. Given the convex composite optimization problem (1), we define the approximate gradient step operator $T_k^G : \mathbb{R}^n \rightarrow \mathbb{R}^n$ at iteration k as

$$T_k^G(x) := x - s_k \nabla^{\epsilon_1^k} g(x), \quad (2)$$

where $s_k > 0$ is the stepsize, $\nabla^{\epsilon_1^k} g := \nabla g + \epsilon_1^k$, and $\epsilon_1^k \in \mathbb{R}^n$ is the gradient error. We also define the approximate proximal operator of h , $T_k^P : \mathbb{R}^n \rightarrow \mathbb{R}^n$, as

$$T_k^P(x) := \text{prox}_h^{\epsilon_2^k}(x), \quad (3)$$

where

$$\text{prox}_{\frac{\epsilon_2^k}{s_k} h}(y) := \left\{ x \in \mathbb{R}^n : h(x) + \frac{1}{2s_k} \|x - y\|_2^2 \leq \epsilon_2^k + \inf_z h(z) + \frac{1}{2s_k} \|z - y\|_2^2 \right\}, \quad (4)$$

where $\epsilon_2^k \in \mathbb{R}$ is the error associated to the proximal computation. Then, the approximate proximal gradient operator is given by the following operator product

$$T_k^{PG} = T_k^P T_k^G. \quad (5)$$

The approximate proximal gradient algorithm sequence is generated by sequentially applying the mapping sequence $\{T_k^{PG}\}_{k>0}$, i.e.,

$$x^{k+1} = T_k^{PG}(x^k) = \text{prox}_h^{\epsilon_2^k}(x^k - s_k \nabla^{\epsilon_1^k} g(x^k)). \quad (6)$$

More precisely, our goal is to establish conditions on the problem and on the errors ϵ_1^k and ϵ_2^k under which (6) converges. In such cases, we also aim to obtain the respective rate of convergence. **Summary of prior work.** It is known that when (1) is convex and g has an L -Lipschitz-continuous gradient, then the exact proximal method, i.e., with $\epsilon_1^k = 0_n$ and $\epsilon_2^k = 0$ for all k , and its accelerated counterpart, require, respectively, $O(1/\rho)$ and $O(\sqrt{1/\rho})$ iterations to achieve an error ρ in the objective function [6], [7]. Although this seems promising in noise-free applications, running the same type of algorithms in resource-constrained environments often leads to unexpected outcomes, as the well-known optimal convergence bounds no longer hold in the presence of gradient and proximal computation errors.

Following [8], the work in [9] showed that the above nearly optimal rates can still be achieved when the computation of the gradients and proximal operators are approximate. This variant is also known as the *Inexact Proximal-Gradient* algorithm. The analysis in [9] requires the errors to decrease with iterations k at rates $O(1/k^{a+1})$ for the basic PG, and $O(1/k^{a+2})$ for the accelerated PG for any $a > 0$ in order to satisfy the summability assumptions of both error terms. The work in [9] established the following ergodic convergence bound in terms of function values of the averaged iterates for the basic approximate PG:

$$f\left(\frac{1}{k} \sum_{i=1}^k x^i\right) - f(x^*) \leq \frac{L}{2k} \left[\|x^* - x^0\|_2 + 2A_k + \sqrt{2B_k} \right]^2, \quad (7)$$

where x^* is any optimal solution of (1), L is the Lipschitz constant of ∇g , x^0 is the initialization vector, and

$$A_k = \sum_{i=1}^k \left(\frac{\|\epsilon_1^i\|_2}{L} + \sqrt{\frac{2\epsilon_2^i}{L}} \right) \quad B_k = \sum_{i=1}^k \frac{\epsilon_2^i}{L}.$$

Our approach. In the case of deterministic errors ϵ_1^k and ϵ_2^k , we get inspiration from [7] to derive, using simple arguments, upper bounds on $f(\frac{1}{k} \sum_{i=1}^k x^i) - f(x^*)$ throughout the iterations. The resulting bounds not only are simpler and tighter than (7), but also decouple the contribution of the two types of errors, ϵ_1^k and ϵ_2^k . In the case of random errors, we show that we can bypass the need to assume that ϵ_1^k and ϵ_2^k converge to zero. We believe this line of reasoning is novel in the analysis of approximate PG algorithms.

Contributions. We summarize our contributions as:

- We establish convergence bounds for the approximate PG in the presence of deterministic errors.
- We extend the analysis to incorporate random errors and propose new parameterized probabilistic convergence bounds with a tuning parameter.
- We propose new models for the proximal and gradient errors that satisfy interesting martingale properties in consistence with experimental results.

II. MAIN RESULTS

All the proofs of the results in this section will appear in a subsequent publication.

Consider the approximate PG algorithm in (6). Before stating our convergence guarantees for approximate PG, we specify our main assumptions on the problem and describe the class of algorithms that our analysis covers.

All of our results assume the following:

Assumption II.1 (Assumptions on the problem).

- The function $h : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{+\infty\}$ is closed, proper, and convex.
- The function $g : \mathbb{R}^n \rightarrow \mathbb{R}$ is convex and differentiable, and its gradient $\nabla g : \mathbb{R}^n \rightarrow \mathbb{R}^n$ is Lipschitz-continuous with constant $L > 0$, that is,

$$\|\nabla g(y) - \nabla g(x)\|_2 \leq L \|y - x\|_2, \quad (8)$$

for all $x, y \in \mathbb{R}^n$, where $\|\cdot\|_2$ stands for the standard Euclidean norm.

- The set of optimal solutions of (1) is nonempty.

The above assumptions are standard in the analysis of PG algorithms and are actually required for convergence to an optimal solution from an arbitrary initialization.

Error models and assumptions. Our analysis assumes two different scenarios:

- 1) The sequences of errors $\{\epsilon_1^k\}_{k \geq 1}$ and $\{\epsilon_2^k\}_{k \geq 1}$ are deterministic, or
- 2) The sequences of errors $\{\epsilon_1^k\}_{k \geq 1}$ and $\{\epsilon_2^k\}_{k \geq 1}$ are discrete stochastic processes, in which case we use $\epsilon_{1\Omega}^k$ and $\epsilon_{2\Omega}^k$ to denote their respective realizations at iteration k .

In scenario 2), the sequences $\{x^k\}_{k \geq 1}$ and $\{y^k\}_{k \geq 1}$ become random as well. And we also use x_{Ω}^k and y_{Ω}^k to denote the respective random vectors at iteration k , where Ω denotes the sample space of a given probability space. We make the following assumption in this case:

Assumption II.2. In scenario 2), we assume that each random vector $\epsilon_{1\Omega}^k$, for $k \geq 1$, satisfies

$$\mathbb{E}[\epsilon_{1\Omega}^k \mid \epsilon_{1\Omega}^1, \dots, \epsilon_{1\Omega}^{k-1}] = \mathbb{E}[\epsilon_{1\Omega}^k] = 0, \quad (9a)$$

$$\mathbb{P}(|\epsilon_{1\Omega,j}^k| \leq \delta) = 1, \quad \text{for all } j = 1, \dots, n, \quad (9b)$$

$$\mathbb{E}[\epsilon_{1\Omega}^{k\top} x_{\Omega}^k \mid \epsilon_{1\Omega}^1, \dots, \epsilon_{1\Omega}^{k-1}, x_{1\Omega}^1, \dots, x_{1\Omega}^{k-1}] = \mathbb{E}[\epsilon_{1\Omega}^{k\top} x_{\Omega}^k] = 0, \quad (9c)$$

where $\epsilon_{1\Omega,j}^k$ in (9b) denotes the j -th entry of $\epsilon_{1\Omega}^k$, and $\delta > 0$ is some finite constant.

The first assumption, (9a), states that $\epsilon_{1\Omega}^k$ is independent from past realizations and has zero mean. The second assumption, (9b), states that the absolute value of each entry of $\epsilon_{1\Omega}^k$ is bounded by δ almost surely. Finally, the third assumption, (9c), states that $\epsilon_{1\Omega}^1, \dots, \epsilon_{1\Omega}^{k-1}, \epsilon_{1\Omega}^k$ and $x_{1\Omega}^1, \dots, x_{1\Omega}^{k-1}, x_{1\Omega}^k$ are mutually independent.

Let us define the residual error vector as follows:

$$r_{\Omega}^k = x_{\Omega}^k - \bar{x}^k, \quad (10)$$

where x_Ω^k and \bar{x}^k stand for the perturbed and gradient-error-free iterates, respectively. Similar assumptions can be made about r_Ω^k , mainly:

$$\mathbb{E}[r_\Omega^k | r_\Omega^1, \dots, r_\Omega^{k-1}] = \mathbb{E}[r_\Omega^k] = 0, \quad (11a)$$

$$\mathbb{E}[r_\Omega^{k\top} x_\Omega^k | r_\Omega^1, \dots, r_\Omega^{k-1}, x_{1\Omega}^1, \dots, x_{1\Omega}^{k-1}] = \mathbb{E}[r_\Omega^{k\top} x_\Omega^k] = 0. \quad (11b)$$

Lemma II.3. *Let x^k and \bar{x}^k be the approximate and exact proximal-gradient iterates and let ϵ_2^k be the proximal error at instant k . Assume a constant stepsize $s_k = s > 0$, for all k . Then, the norm of the residual vector $r^k = x^k - \bar{x}^k$ satisfies*

$$\|r^k\|_2 \leq \sqrt{2s\epsilon_2^k}, \quad \forall k > 0. \quad (12)$$

Lemma II.3 bounds the norm of the residual vector r^k as a function of ϵ_2^k . Therefore, boundedness of the latter implicitly implies boundedness of the norm of the former.

We start by considering deterministic error sequences $\{\epsilon_1^k\}_{k \geq 1}$ and $\{\epsilon_2^k\}_{k \geq 1}$, and then we consider the case in which these sequences are random, as in Assumption II.2.

Deterministic errors. Our first result provides a bound for the ergodic convergence of the sequence of function values, and decouples the contribution of the errors in the computation of gradient and in the computation of the proximal operator.

Theorem II.4 (PG, deterministic errors). *Consider problem (1) and let Assumption II.1 hold. Then, for arbitrary error sequences $\{\epsilon_1^k\}_{k \geq 1}$ and $\{\epsilon_2^k\}_{k \geq 1}$, the sequence generated by approximate PG in (6) with constant stepsize $s_k := s \leq 1/L$, for all k , satisfies*

$$\begin{aligned} f\left(\frac{1}{k+1} \sum_{i=0}^k x^{i+1}\right) - f(x^*) &\leq \frac{1}{k+1} \left[\sum_{i=0}^k \epsilon_2^i \right. \\ &+ \sum_{i=0}^k \left(\epsilon_1^i - \frac{1}{s} r^{i+1} \right)^\top (x^* - x^{i+1}) + \frac{1}{2s} \|x^* - x^0\|_2^2 \\ &\left. - \frac{1}{k+1} \left[\frac{1}{2s} \sum_{i=0}^k \|r^{i+1}\|_2^2 + \frac{1}{2s} \|x^* - x^{k+1}\|_2^2 \right], \quad (13) \end{aligned}$$

where x^* is any solution of (1) and r^i is the residual vector associated with error ϵ_2^i defined in (10).

This result implies that the well-known $O(1/k)$ convergence rate for the gradient method without errors still holds when both ϵ_2^k and $(\epsilon_1^k - \frac{1}{s} r^{k+1})^\top (x^* - x^k)$ are summable. Note that a faster convergence of these two errors will not improve the convergence rate but will yield a better coefficient.

Consider now the case in which the sequence $\{r^k\}_{k > 0}$ cannot be observed [as \bar{x}^k in (10) is usually unobservable], but is bounded, e.g., if $\{\epsilon_2^k\}$ is bounded as in Lemma II.3. Then, to obtain a convergence bound that is independent of the particular sequences $\{x^k\}_{k \geq 0}$ and $\{r^k\}_{k > 0}$, we can apply Cauchy-Schwarz's inequality to the first term involving r^k in the right-hand side of (13) followed by Féjer's inequality (see [7, Thm. 10.23]):

Corollary II.5. *Under the same conditions as Theorem II.4, the sequence generated by approximate PG in (6) satisfies*

$$\begin{aligned} f\left(\frac{1}{k+1} \sum_{i=0}^k x^{i+1}\right) - f(x^*) &\leq \frac{1}{k+1} \left[\sum_{i=0}^k \epsilon_2^i \right. \\ &+ \sum_{i=0}^k \left\| \epsilon_1^i - \frac{1}{s} r^{i+1} \right\|_2 \|x^* - x^{i+1}\|_2 + \frac{1}{2s} \|x^* - x^0\|_2^2 \\ &\left. - \frac{1}{k+1} \left[\frac{1}{2s} \sum_{i=0}^k \|r^{i+1}\|_2^2 + \frac{1}{2s} \|x^* - x^{k+1}\|_2^2 \right], \quad (14) \end{aligned}$$

where, again, x^* is any solution of (1) and we used Lemma II.3 to bound $\|r^k\|$.

Notice that, by Féjer's inequality [7, Thm. 10.23], $\|x^* - x^0\|_2$ upper bounds all residuals $\|x^* - x^i\|_2$. Since $\{\epsilon_1^k\}_{k \geq 1}$ is a centered sequence, the use of Cauchy-Schwarz's inequality followed by Féjer's inequality yields a bound looser than the one in (13). Yet, the $O(1/k)$ convergence rate is still guaranteed with weaker summability assumptions of $\{\epsilon_2^k\}_{k \geq 1}$ and $\{\|\epsilon_1^k\|\}_{k \geq 1}$. If we set both errors to zero for all $k \geq 1$, we recover the error-free optimal upper bound $\frac{L}{2k} \|x^* - x^0\|_2^2$ [7].

Next we relax the summability assumption on ϵ_1^k and ϵ_2^k and replace it with the weaker assumption of *boundedness*.

Random errors. Let us now consider the case in which ϵ_1^k , ϵ_2^k and therefore x^k , are random, and let $\epsilon_{1\Omega}^k$, $\epsilon_{2\Omega}^k$ and x_Ω^k be the corresponding random variables/vectors, respectively.

Theorem II.6 (Random errors). *Consider problem (1) and let Assumption II.1 hold. Assume that the rounding error $\{\epsilon_{1\Omega}^k\}_{k \geq 1}$ and residual error $\{r_\Omega^k\}_{k \geq 1}$ sequences satisfy Assumption II.2 and $\mathbb{P}(\epsilon_{2\Omega}^k \leq \epsilon_0) = 1$, for all $k > 0$, and for some $\epsilon_0 \in \mathbb{R}$. Then, for any $\gamma > 0$, the sequence generated by approximate PG in (6) with constant stepsize $s_k := s \leq 1/L$, for all k , satisfies*

$$\begin{aligned} f\left(\frac{1}{k} \sum_{i=1}^k x_\Omega^i\right) - f(x^*) &\leq \frac{1}{k} \sum_{i=1}^k \epsilon_{2\Omega}^i + \\ &\frac{\gamma}{\sqrt{k}} \left(\sqrt{n} |\delta| + \sqrt{\frac{2\epsilon_0}{s}} \right) \|x^* - x^0\|_2 + \frac{1}{2sk} \|x^* - x^0\|_2^2, \quad (15) \end{aligned}$$

with probability at least $1 - 2 \exp(-\frac{\gamma^2}{2})$, where x^* is any solution of (1).

For large scale problems,¹ we typically have $n \gg \frac{1}{s} \geq L$; therefore, we obtain the following approximated bound,

$$\begin{aligned} f\left(\frac{1}{k} \sum_{i=1}^k x_\Omega^i\right) - f(x^*) &\lesssim \frac{1}{k} \sum_{i=1}^k \epsilon_{2\Omega}^i + \gamma \sqrt{\frac{n}{k}} |\delta| \|x^* - x^0\|_2 \\ &+ \frac{1}{2sk} \|x^* - x^0\|_2^2, \quad (16) \end{aligned}$$

¹And for same levels of error magnitudes δ and ϵ_0 .

with the same probability. In the absence of computational errors, (15) coincides with the results of Theorem II.4 and Corollary II.5, which reduce to the deterministic noise-free convergence upper bound, i.e., $\frac{L}{2k} \|x^* - x^0\|_2^2$. With exact proximal operation and approximate gradient computations (i.e., $\epsilon_2^k = 0$ and $\epsilon_2^k \neq 0$ for all $k \geq 0$), if we let the machine precision δ to decrease at $O(\frac{1}{k^{0.5+\delta}})$, i.e., progressively increase computation accuracy, then we obtain the optimal convergence rate $O(\frac{1}{k})$. In order to recover the same convergence rate for the approximate proximal case, we also need the sum of the ensemble means $\mathbb{E}(\epsilon_{2\Omega}^i)$ to decrease as $O(\frac{1}{k^{1+\delta}})$, which is a weaker than what [9] [cf.(7)] requires: $O(\frac{1}{k^{2+\delta}})$. This result also suggests that a slower $O(\frac{1}{\sqrt{k}})$ convergence rate (same as noise-free subgradient method) is achieved when the sequence of ensemble means $\{\mathbb{E}(\epsilon_{2\Omega}^k)\}$ is summable for all centered and bounded sequences $\{\epsilon_{1\Omega}^k\}$, and consequently the proximal error is the main contributor to any divergence from the optimal set X^* .

Notice that for a fixed machine precision δ and probability parameter γ we obtain a computable error residual constant rather than variable running error terms as in Theorem II.4, Corollary II.5 or (7) without making any summability assumptions on $\{\|\epsilon_{1\Omega}^k\|\}$ (as in Corollary II.5) or $\{\epsilon_{1\Omega}^k\}$ in general.

Moreover, the effect of the dimension n of the problem variable appears explicitly in (15), but neither in Theorem II.4, nor in Corollary II.5, nor in (7). The latter suggests that using progressively sparser gradient vectors² can potentially accelerate the convergence speed (e.g., by using $n' \ll n$), but never faster than the optimal (limit) speed of $O(\frac{1}{k})$. Overall, better design parameter selections would result in better error residuals rather than exceeding the optimal convergence rate.

The following result applies if we relax the summability of $\{\epsilon_{2\Omega}^k\}$ but still assume statistical stationarity.³

Theorem II.7 (Random errors). *Consider problem (1) and let Assumption II.1 hold. Assume that the rounding error $\{\epsilon_{1\Omega}^k\}_{k \geq 1}$ and residual error $\{r_{\Omega}^k\}_{k \geq 1}$ sequences satisfy Assumption II.2, and that the proximal computation error is upper bounded, i.e., $\epsilon_{2\Omega}^k \leq \epsilon_0$ for all $k \geq 1$, and also stationary with constant mean $\mathbb{E}(\epsilon_{2\Omega})$. Then, the sequence generated by approximate PG in (6) with constant stepsize $s_k := s \leq 1/L$, for all k , satisfies*

$$f\left(\frac{1}{k} \sum_{i=1}^k x_{\Omega}^i\right) - f(x^*) \leq E(\epsilon_{2\Omega}) + \frac{\gamma}{\sqrt{k}} \left(\frac{\epsilon_0}{2} + \sqrt{n}|\delta| \|x^* - x^0\|_2\right) + \frac{1}{2sk} \|x^* - x^0\|_2^2, \quad (17)$$

with probability at least $1 - 2 \exp(-\frac{\gamma^2}{2})$, where x^* is any solution of (1).

The following corollary applies to approximate PG with uniformly distributed proximal error.

²As is the case in *Proximal-Gradient* algorithm when applied to LASSO.

³This means that the ensemble mean and the variance are time-invariant.

Corollary II.8 (Random uniformly distributed proximal error). *Let the proximal computation error be upper bounded, i.e., $\epsilon_{2\Omega}^k \leq \epsilon_0$, for all $k \geq 1$. If the latter is stationary and uniformly distributed over its range, i.e., $\epsilon_{2\Omega}^k \sim \mathcal{U}\{0, \epsilon_0\}$, then substituting $\mathbb{E}(\epsilon_{2\Omega}) = \frac{\epsilon_0}{2}$ in the bound of Theorem II.7 gives*

$$f\left(\frac{1}{k} \sum_{i=1}^k x_{\Omega}^i\right) - f(x^*) \leq \frac{\epsilon_0}{2} + \frac{\gamma}{\sqrt{k}} \left(\frac{\epsilon_0}{2} + \sqrt{n}|\delta| \|x^* - x^0\|_2\right) + \frac{1}{2sk} \|x^* - x^0\|_2^2, \quad (18)$$

with probability at least $1 - 2 \exp(-\frac{\gamma^2}{2})$, where x^* is any solution of (1).

In terms of the proximal error $\epsilon_{2\Omega}^k$, using a concentration-based probabilistic bound, i.e., $\epsilon_{2\Omega}^k \leq (1/2 + \gamma/\sqrt{k})\epsilon_0$ results in a sharper bound than what we would have obtained if we used the more conservative deterministic upper bound $\epsilon_{2\Omega}^k \leq \epsilon_0$.

III. QUANTIZATION ERRORS

We start by briefly reviewing the theory behind hardware quantization. Quantization is a critical step between data-level and hardware-level, which can be thought of as a type of contract between the two levels of the application in order to allocate a certain (finite) number of bits (resource) to represent an infinitely precise parameter or a value from a continuous signal in the digital circuit with finite precision. This very initial step of data type/hardware design plays a major role in determining the overall precision for the application as well as the complexity of the implementation.

Quantization errors can be reduced by choosing an appropriate number of bits (usually the higher the number of used bits the less the error will be). However, this reduction comes at the cost of using more hardware resources for storage and computation. The trade-off between error attenuation and calculation precision, energy and speed (or latency), is subject to application constraints.

Fixed-point number system. Fixed-point machine representation can be interpreted in different ways, and we discuss here the two most common approaches. In the first case, all the word (W) bit elements are allocated for value representation assuming all signal values are always positive. As can be deduced from the machine representation system's name, i.e., "fixed-point," the set of bits in W are split by a fixed-point into I most significant bits (MSBs) and F least significant bits (LSBs) representing the integer and the fractional parts, respectively. This quantization is called unsigned I.F or *uI.F* for short. The quantized value is evaluated by the following formula,

$$uI.F(x) = \sum_{i=0}^{W-1} b_i(x) 2^{i-F}, \quad (19)$$

with $F, W \in \mathbb{N}_+$. The corresponding dynamic range (DR) is given by $DR_{uI.F} = [0; 2^I - 2^{-F}]$. The signed I.F, or *sI.F* can be obtained from I.F by encoding the sign of the value

using one bit and this is typically done by taking the most significant bit (MSB) of the integer part I as being a sign bit. Although this operation would reduce the integer part's number of bits I to $I - 1$, the two's complement approach handles negative numbers and therefore extends the DR in the negative direction, i.e., $DR_{sI.F} = [-2^{I-1}; 2^{I-1} - 2^{-F}]$, and the quantized value is now given by

$$sI.F(x) = \sum_{i=0}^{W-2} b_i 2^{i-F} - b_{W-1} 2^{I-1}. \quad (20)$$

Quantization Rules. Once a machine representation is defined, quantization rules need to be established in order to define the quantization behaviour under decreasing accuracy and/or overflow in subsequent operations. Although the IEEE 754 standard uses *round-to-nearest integer*⁴, other modes of rounding also exist in the literature. For instance, rounding towards the standard limits $(0, +\infty, -\infty)$ is also known as *directed rounding*. Moreover, rounding can be deterministic in a pre-defined rule or stochastic according to a random distribution.

IV. EXPERIMENTAL RESULTS

We now apply the proposed bounds to analyze the convergence of the approximate *proximal gradient* algorithm when applied to solve randomly generated LASSO problems:

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2} \|Ax - y\|_2^2 + \lambda \|x\|_1,$$

where $n = 100$ (dimension of x) and $A \in \mathbb{R}^{m \times n}$ has $m = 500$ rows. We run a total of 5 random experiments for every algorithm parameter selection. We mainly vary the bitwidth (BIT), the fraction width (FRAC.) of the fixed-point representation in (20), the CVX [10] solver's precision (PRECISION) to approximate the proximal step (4), and the tolerance bound of the approximate PG (ABSTOL in Table I). We record and take the average over all 5 experiments of the residual error in the iterates $\|x - x^*\|_2$, the residual error (suboptimality) in the function values $f - f^*$, and the total number of iterations k ($k, \textit{iters.}$). The results are summarized in Table I. We also plot the proposed convergence bounds, the error-free optimal bound as well as the original bound in (7) for the different tests as depicted in Figs. 1-3. Note that for the probabilistic bounds we tune the parameter γ to obtain 3 different bounds which hold with probabilities 1, 0.5 and 0.25, respectively. From Figs.1-3, we can clearly see that our proposed bounds give better approximations of the discrepancy caused by perturbations, and consequently we obtain better error terms. As a necessary condition for convergence, we only required the partial sums $\sum_{i=1}^k \epsilon_2^i$ and $\sum_{i=1}^k \|\epsilon_1^i\|_2$ to be in $o(k)$, in contrast to the stronger condition $o(\sqrt{k})$ of (7). For the probabilistic bounds, we do not assume summability of the error terms but only require them to be

⁴If the correct answer is exactly halfway between the two, the system chooses the output where the least significant bit of the fraction (mantissa M) is zero.

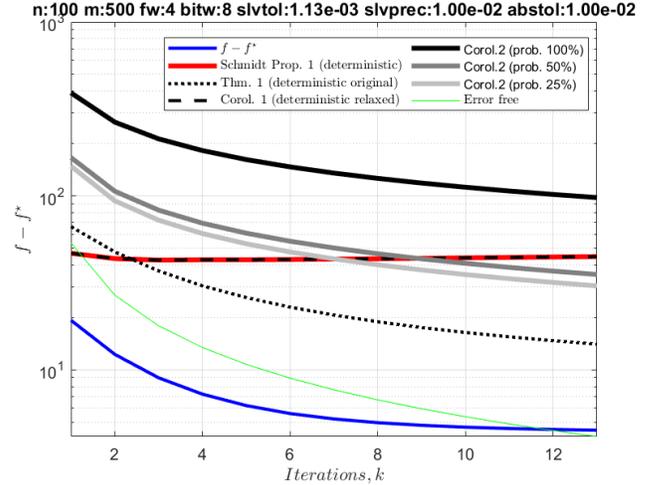


Fig. 1. Upper bounds based on Theorems II.4 & II.7 and their corresponding corollaries vs (7)

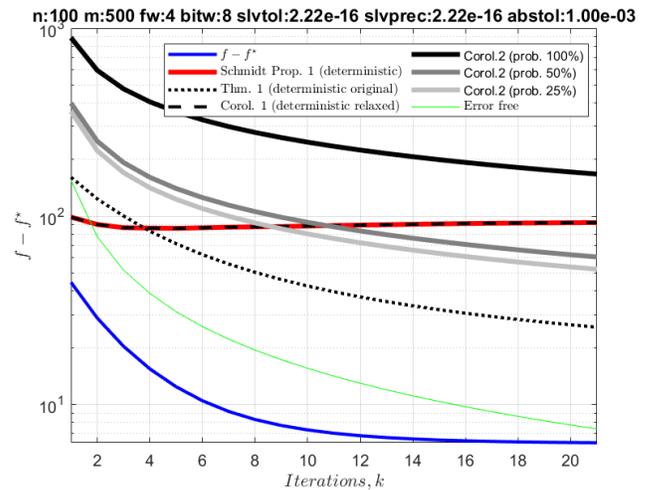


Fig. 2. Upper bounds based on Theorems II.4 & II.7 and their corresponding corollaries vs (7)

Table I
RESULTS OF OUR EXPERIMENTS.

PRECISION	BIT	FRAC.	ABSTOL	$\ x - x^*\ _2$	$f - f^*$	$k, \textit{iters.}$
2.22e-16	8	4	2.22e-16	0.052609	1.3818	85
		0.001	0.15256	6.2202	20	
	16	6	2.22e-16	0.13947	5.1224	79
0.001	8	4	0.001	0.1418	5.6894	15
		0.01	0.09152	3.1661	79	
	16	8	2.22e-16	0.09152	3.1661	79
0.01	8	4	0.001	0.14508	5.555	20
		0.01	0.14271	5.6136	14	
	16	8	2.22e-16	0.096077	4.1512	84
	8	4	0.001	0.13155	4.5246	18
		0.01	0.12847	4.4239	14	
	16	6	0.01	0.13084	4.3855	14
		8	0.01	0.15369	5.8671	15

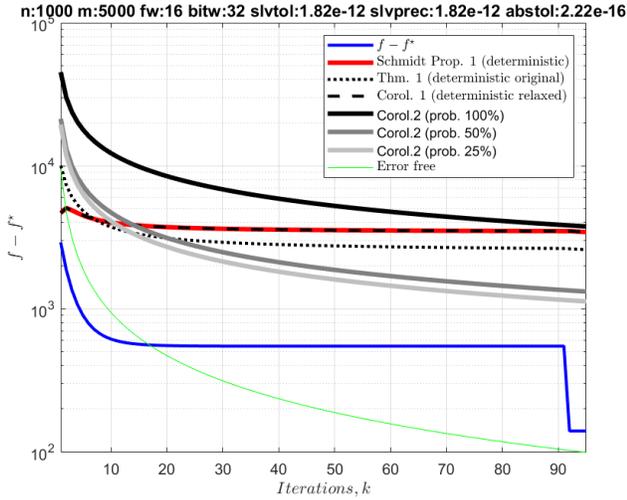


Fig. 3. Upper bounds based on Theorems II.4 & II.7 and their corresponding corollaries vs (7)

bounded. Consequently, the probabilistic bounds achieve better approximations over iterations and are less sensitive to error variations and become tighter with decreasing probability. If we relax our original bound of Theorem II.4 and use Lemma II.3 to bound the sequence of the proximal residual error $\{r^k\}$, then our bound coincides with the one in (7), as depicted by the overlapping dashed and red lines in Figs. 1-3.

Increasing the tolerance bound of the approximate PG from 2.22×10^{-16} to 10^{-3} improved the algorithm's running time by 65 iterations for the 8 bits representation and 66 iterations for the 16 bits representation without affecting too much the residuals.

Table I shows that, in general, varying the internal loop (CVX solver's) precision does not largely affect the number of outer iterations of the PG, but leads to substantial bias around the optimum when increased from 0.01 by a factor of 10.

Reducing the hardware precision from 16 to 8 bits accelerated the algorithm by 6 iterations, but slightly increased the residual in the solution $\|x - x^*\|_2$ by 8.6861×10^{-2} while added 3.7406 extra bias error to the function value. Increasing the hardware precision by allocating more bits for the fractional part in the fixed-point representation of 16 bits caused the residual error in $\|x - x^*\|_2$ to drop by 17.46% and the error in $\|f - f^*\|_2$ by 33.78% without remarkable effect on the number of iterations.

V. CONCLUSIONS

We considered the proximal-gradient algorithm in the case in which the gradient of the differentiable function and the proximal operator are computed with errors. We obtained new bounds, tighter than the ones in [9] and demonstrated their validity on a practical optimization example (LASSO) solved on a reduced-precision machine combined with reduced-precision solver. While we established worst-case performance bounds, we also established probabilistic upper bounds catering for

random computational errors. Interesting directions for future work include relaxing the assumptions in order to incorporate more general perturbations into the analysis and considering accelerated versions (i.e., Nesterov) of the approximate proximal-gradient algorithm.

REFERENCES

- [1] J. S. Cramer, "The origins of logistic regression," 2002.
- [2] D. L. Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [3] Y. Wu, J. F. Mota, and A. M. Wallace, "Approximate lasso model predictive control for resource constrained systems," in *2020 Sensor Signal Processing for Defence Conference (SSPD)*, IEEE, 2020, pp. 1–5.
- [4] P. Machart, S. Anthoine, and L. Baldassarre, "Optimal computational trade-off of inexact proximal methods," *arXiv preprint arXiv:1210.5034*, 2012.
- [5] M. Ulbrich and B. van Bloemen Waanders, "An introduction to partial differential equations constrained optimization," *Optimization and Engineering*, vol. 19, no. 3, pp. 515–520, 2018.
- [6] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM journal on imaging sciences*, vol. 2, no. 1, pp. 183–202, 2009.
- [7] A. Beck, *First-order methods in optimization*. SIAM, 2017, vol. 25.
- [8] A. Beck and M. Teboulle, "Gradient-based algorithms with applications to signal recovery," *Convex optimization in signal processing and communications*, pp. 42–88, 2009.
- [9] M. Schmidt, N. L. Roux, and F. R. Bach, "Convergence rates of inexact proximal-gradient methods for convex optimization," in *Advances in neural information processing systems*, 2011, pp. 1458–1466.
- [10] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming*, cvxr.com/cvx, 2011.

Learning a Secondary Source From Compressive Measurements for Adaptive Projection Design

Fraser K. Coutts, John Thompson, and Bernard Mulgrew

Institute for Digital Communications, University of Edinburgh, Edinburgh, EH9 3FG, UK

email: fraser.coutts@ed.ac.uk

Abstract—Recent work has established that the gradient of the mutual information (MI) in Gaussian channels with input noise can be used for projection design in a compressive sensing (CS) scenario with two independent, Gaussian mixture (GM)-distributed inputs. The resulting CS projection matrices have been shown to be capable of controlling the input-output MI terms for the two inputs. One downside of such information-theoretic strategies is their reliance on access to *a priori* knowledge of the input source statistics. In this paper, we assume that the GM distribution of a primary input is known and that the GM distribution of a secondary input is unknown. We derive a methodology for the online training of the distribution of the secondary input via compressive measurements and illustrate that once the distribution of this secondary source is known, we can use projection design to control the input-output MI of the system. We demonstrate through simulations the various performance trade-offs that exist within the proposed methodology.

I. INTRODUCTION

Dimensionality reduction methods based on linear random projections — i.e., compressive sensing [1] (CS) — have gained significant attention recently; however, random projections may not be the best choice if we know the statistical properties of the source signal [2]. By employing an information-theoretic approach, one can design a linear projection such that the mutual information (MI) between the projected signal and the source signal or its class label is maximised [3], [4]. Intuitively, as the MI increases, the recovery of the source signal or label information improves; indeed, the Bayes classification error is bounded by the MI [3].

The distribution of a non-Gaussian signal can be approximated by a mixture of several Gaussians [5]. Importantly, increasing the number of Gaussians used enables the approximation of general distributions to an arbitrary level of accuracy [6]. Such Gaussian mixture models (GMMs) have been shown to be effective [7] and in some cases superior to sparse signal models in CS scenarios [5]. Recent work [8] utilises MI maximisation within a CS framework to optimise information throughput for a Gaussian mixture (GM) source signal in the presence of GM input noise. In [9], this framework is extended to complex signal models and applied to real radar data containing micro-Doppler (m-D) signatures [10]; subsequent results highlight that the methodology is able to assist in the joint classification of the m-D signatures of a primary, always-present source and a secondary, fleeting source. By modelling two independent inputs via GMMs and treating each as a source of structured input noise for the other, both [8] and [9] employ an iterative gradient-ascent approach to design a linear projection matrix capable of controlling the

information throughput for each source. However, these works rely on *a priori* knowledge of the source statistics and are therefore limited to the case of stationary sources.

In [11], Yang *et al.* investigate a scenario in which a desired source in a CS scenario without input noise has an unknown GM distribution. They seek to learn the distribution using only knowledge of the compressive measurements, the projection matrices involved, and the parameters of the Gaussian measurement noise. Their implementation is iterative and related to the expectation-maximisation (EM) algorithm [12].

In this paper, we consider the compressive measurement of two GM-distributed inputs; these perceive each other as additive noise and experience the same linear projection. We assume that the GM distribution of the primary input is known and that the GM distribution of the secondary input is unknown. We extend the work of [11] and derive a novel methodology for the training of the GM distribution of the secondary input from compressive measurements and illustrate that once the distribution of this secondary source is known, we can use the projection design techniques of [8], [9] to control the input-output MI of the system. For this demonstration, we apply the developed adaptive projection design algorithm to real radar data containing two coincident sources of m-D information. Using synthetic data, we also show the various performance trade-offs that exist within the proposed distribution learning methodology.

Below, Sec. II establishes the signal model considered in this paper. In Sec. III, we summarise the optimisation framework from [8], [9] that we use for projection design. In Sec. IV, we introduce the theory and algorithm required to learn the GM distribution of a secondary source from compressive measurements. Sec. V then briefly explains how these projection design and source learning procedures fit together within an adaptive algorithm. Sec. VI provides a practical demonstration of the proposed approach, and conclusions are drawn in Sec. VII.

Notation: Straight bold lowercase and uppercase symbols denote vectors and matrices, respectively, and \mathbf{I}_n is an $n \times n$ identity matrix. Italicised uppercase letters such as \mathbf{Y} and \mathbf{C} denote random vectors and variables; their realisations are lowercase equivalents, such as \mathbf{y} and c . Operators $\{\cdot\}^H$, $\{\cdot\}^*$, $\mathbb{E}[\cdot]$, $\|\cdot\|_1$, and $\text{tr}\{\cdot\}$ evaluate the Hermitian transpose, complex conjugate, expectation, ℓ_1 -norm, and trace, respectively.

II. SIGNAL MODEL

We consider the following complex-valued signal model with input noise:

$$\mathbf{Y} = \Phi(\mathbf{X} + \mathbf{N}) + \mathbf{W}. \quad (1)$$

Following the CS protocol, we have measurements $\mathbf{Y} \in \mathbb{C}^m$ obtained from input signals $\mathbf{X} \in \mathbb{C}^n$ and $\mathbf{N} \in \mathbb{C}^n$ via a compressive projection matrix $\Phi \in \mathbb{C}^{m \times n}$, with $m \ll n$. We opt for \mathbf{X} and \mathbf{N} to have a GM distribution for two reasons. Firstly, GMs are known to be effective in statistical CS scenarios [5], [7]. Secondly, a GM-distributed \mathbf{X} permits a natural extension of the work of [11] to the case of estimating \mathbf{N} given knowledge of \mathbf{Y} , Φ , and the distributions of \mathbf{X} and the measurement noise \mathbf{W} .

The signal \mathbf{N} , which is independent of \mathbf{X} and — when attempting to recover the features of \mathbf{X} — can be considered as input noise, is distributed according to a complex GM; i.e.,

$$\mathbf{N} \sim p_{\mathbf{n}}(\mathbf{n}) = \sum_{g=1}^{J_n} r_g \sum_{k=1}^K s_{g,k} \mathcal{CN}(\mathbf{n}; \boldsymbol{\mu}_{g,k}, \Gamma_{g,k}), \quad (2)$$

with mean vectors $\boldsymbol{\mu}_{g,k} \in \mathbb{C}^n$, covariance matrices $\Gamma_{g,k} \in \mathbb{C}^{n \times n}$, and weights $s_{g,k}$ such that $\sum_{k=1}^K s_{g,k} = 1$. An instance of \mathbf{N} is generated by one of J_n classes, which are each characterised by a GM with K components. The classes $g = 1, \dots, J_n$ occur with probability r_g such that $\sum_{g=1}^{J_n} r_g = 1$. The random vector \mathbf{X} represents a signal of interest and is distributed as

$$\mathbf{X} \sim p_{\mathbf{x}}(\mathbf{x}) = \sum_{c=1}^{J_x} z_c \sum_{o=1}^O \pi_{c,o} \mathcal{CN}(\mathbf{x}; \boldsymbol{\chi}_{c,o}, \Omega_{c,o}). \quad (3)$$

That is, the probability distributions of classes $c = 1, \dots, J_x$ of \mathbf{X} are each characterised by a GM with O components. The vector $\mathbf{W} \sim \mathcal{CN}(\mathbf{w}; \mathbf{v}, \Lambda)$ represents additive complex Gaussian noise with mean $\mathbf{v} \in \mathbb{C}^m$ and covariance $\Lambda \in \mathbb{C}^{m \times m}$.

III. OPTIMISATION FRAMEWORK

Assume, for now, that — in addition to the distribution of our primary, always present source \mathbf{X} — we have access to a predefined GM distribution for \mathbf{N} , which represents a secondary source that may or may not be present in the system at the time of measurement. As such, in the distribution for \mathbf{N} , the J_n th class is characterised as $p_{\mathbf{n}|g}(\mathbf{n}|g = J_n) = \sum_{k=1}^K (1/K) \mathcal{CN}(\mathbf{0}, \sigma \mathbf{I}_n)$ for some arbitrarily small σ ; i.e., the value of \mathbf{N} for this class is close to zero to represent the scenario in which the secondary source is absent.

We seek the matrix Φ that maximises the objective function

$$F(\Phi, \boldsymbol{\beta}) = \beta_1 I(\mathbf{X}; \mathbf{Y}) + \beta_2 I(\mathbf{C}; \mathbf{Y}) + \beta_3 I(\mathbf{N}; \mathbf{Y}) + \beta_4 I(\mathbf{G}; \mathbf{Y}), \quad (4)$$

where $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3, \beta_4] \in \mathbb{R}^4$ controls the relative importance of the Shannon MI [13] terms and \mathbf{C} and \mathbf{G} are random variables that represent the classes of \mathbf{X} and \mathbf{N} . For our purposes, we maintain $\|\boldsymbol{\beta}\|_1 = 1$.

We use the iterative gradient ascent algorithm of [8], [9] to identify the matrix Φ that locally maximises $F(\Phi, \boldsymbol{\beta})$ by setting $\Phi \leftarrow \Phi + \delta \nabla_{\Phi} F(\Phi, \boldsymbol{\beta})$ and normalising such that $\text{tr}\{\Phi \Phi^H\} = m$ at each iteration. The step size $\delta > 0$ controls the rate of change of Φ . When computing $\nabla_{\Phi} F(\Phi, \boldsymbol{\beta})$, we evaluate the gradient terms given in [8], [9] via Monte Carlo (MC) integration and utilise the Bayesian inference model detailed in [9].

IV. LEARNING THE SECONDARY SOURCE DISTRIBUTION

Here, we extend the work of [11] such that we are able to learn the distribution of \mathbf{N} from compressive measurements.

We assume that our compressive measurements have been captured using a block of data that contains instances of only one class of \mathbf{N} . Therefore, we omit the class parameter g to simplify notation. We first rewrite (1) as $\mathbf{Y} = \Phi \mathbf{N} + \mathbf{W}$, where

$$\hat{\mathbf{W}} \sim \sum_{d=1}^D \tau_d \mathcal{CN}(\hat{\mathbf{w}}; \mathbf{v}_d, \Lambda_d), \quad (5)$$

$$\begin{aligned} \tau_d &= z_{c'} \pi_{c',o'}, & \mathbf{v}_d &= \Phi \boldsymbol{\chi}_{c',o'} + \mathbf{v}, & \Lambda_d &= \Phi \Omega_{c',o'} \Phi^H + \Lambda, \\ D &= J_x O, & c' &= \lceil \frac{d}{O} \rceil, & o' &= ((d-1) \bmod O) + 1. \end{aligned}$$

We seek the system parameters θ that maximise the log of the marginal probability; i.e., the incomplete log-likelihood:

$$\ell_{\text{inco}}(\theta|\mathbf{y}) = \log p_{\mathbf{y}|\theta}(\mathbf{y}|\theta) = \log \sum_{k,d} \int p_{\mathbf{y},\mathbf{n},k,d|\theta}(\mathbf{y}, \mathbf{n}, k, d|\theta) d\mathbf{n}.$$

Since the log of a sum is not easily separable for maximisation purposes, we take a two-stage EM approach [12] and utilise the complete log-likelihood:

$$\ell_{\text{co}}(\theta|\mathbf{y}, \mathbf{n}, k, d) = \log p_{\mathbf{y},\mathbf{n},k,d|\theta}(\mathbf{y}, \mathbf{n}, k, d|\theta). \quad (6)$$

Specifically, we consider the expected value of $\ell_{\text{co}}(\theta|\mathbf{y}, \mathbf{n}, k, d)$ under the posterior distribution of the latent variables (\mathbf{n}, k, d) . In the first stage of iteration ($t+1$), we use the previous parameters $\theta^{(t)}$ to find the posterior distribution of the latent variables given by $p_{\mathbf{n},k,d|\mathbf{y},\theta}(\mathbf{n}, k, d|\mathbf{y}, \theta^{(t)})$. We then use this to find the expectation of the complete log-likelihood

$$\ell_{\text{ex-co}}(\theta|\mathbf{y}, \theta^{(t)}) = \mathbb{E}_{\mathbf{n},k,d|\mathbf{y},\theta^{(t)}} \left[\log p_{\mathbf{y},\mathbf{n},k,d|\theta}(\mathbf{y}, \mathbf{n}, k, d|\theta) \right]$$

with respect to this posterior. In the second stage, we determine the new parameters $\theta^{(t+1)}$ by maximizing $\ell_{\text{ex-co}}(\theta|\mathbf{y}, \theta^{(t)})$:

$$\theta^{(t+1)} = \arg \max_{\theta} \ell_{\text{ex-co}}(\theta|\mathbf{y}, \theta^{(t)}). \quad (7)$$

Fortunately, we can show that maximising this function actually maximises the incomplete log-likelihood. We can write

$$p_{\mathbf{y}|\theta}(\mathbf{y}|\theta) = p_{\mathbf{y},\mathbf{n},k,d|\theta}(\mathbf{y}, \mathbf{n}, k, d|\theta) / p_{\mathbf{n},k,d|\mathbf{y},\theta}(\mathbf{n}, k, d|\mathbf{y}, \theta). \quad (8)$$

By taking the expectation of the log of both sides with respect to $p_{\mathbf{n},k,d|\mathbf{y},\theta}(\mathbf{n}, k, d|\mathbf{y}, \theta^{(t)})$, we obtain

$$\log p_{\mathbf{y}|\theta}(\mathbf{y}|\theta) = \ell_{\text{ex-co}}(\theta|\mathbf{y}, \theta^{(t)}) + h(\theta|\mathbf{y}, \theta^{(t)}), \quad (9)$$

where $h(\theta|\mathbf{y}, \theta^{(t)})$ is a conditional entropy term. The above holds for any θ , including $\theta = \theta^{(t)}$. That is,

$$\log p_{\mathbf{y}|\theta}(\mathbf{y}|\theta^{(t)}) = \ell_{\text{ex-co}}(\theta^{(t)}|\mathbf{y}, \theta^{(t)}) + h(\theta^{(t)}|\mathbf{y}, \theta^{(t)}). \quad (10)$$

Subtracting this from (9), we obtain

$$\begin{aligned} \log p_{\mathbf{y}|\theta}(\mathbf{y}|\theta) - \log p_{\mathbf{y}|\theta}(\mathbf{y}|\theta^{(t)}) &= \\ \ell_{\text{ex-co}}(\theta|\mathbf{y}, \theta^{(t)}) + h(\theta|\mathbf{y}, \theta^{(t)}) - \ell_{\text{ex-co}}(\theta^{(t)}|\mathbf{y}, \theta^{(t)}) - h(\theta^{(t)}|\mathbf{y}, \theta^{(t)}). \end{aligned}$$

By Gibbs' inequality [13], we know that for two probability distributions $p_1(\mathbf{y})$ and $p_2(\mathbf{y})$, we have

$$-\int p_1(\mathbf{y}) \log p_1(\mathbf{y}) d\mathbf{y} \leq -\int p_1(\mathbf{y}) \log p_2(\mathbf{y}) d\mathbf{y}, \quad (11)$$

with equality only when $p_1(\mathbf{y}) = p_2(\mathbf{y})$. Thus, we have $h(\theta|\mathbf{y}, \theta^{(t)}) \geq h(\theta^{(t)}|\mathbf{y}, \theta^{(t)})$ and

$$\begin{aligned} \log p_{\mathbf{y}|\theta}(\mathbf{y}|\theta) - \log p_{\mathbf{y}|\theta}(\mathbf{y}|\theta^{(t)}) &= \\ \geq \ell_{\text{ex-co}}(\theta|\mathbf{y}, \theta^{(t)}) - \ell_{\text{ex-co}}(\theta^{(t)}|\mathbf{y}, \theta^{(t)}). \end{aligned} \quad (12)$$

That is, choosing θ such that $\ell_{\text{ex-co}}(\theta|\mathbf{y}, \theta^{(t)})$ improves upon $\ell_{\text{ex-co}}(\theta^{(t)}|\mathbf{y}, \theta^{(t)})$ guarantees that the resulting improvement

from $\log p_{y|\theta}(\mathbf{y}|\theta^{(t)})$ to $\log p_{y|\theta}(\mathbf{y}|\theta)$ is at least as large.

When evaluating the posterior distribution of the latent variables, we utilise the following Bayesian inference model, where we have omitted the reliance on $\theta^{(t)}$ for brevity:

$$p_{k,d|\mathbf{y}}(k, d|\mathbf{y}) = \frac{s_k \tau_d p_{y|k,d}(\mathbf{y}|k, d)}{p_{\mathbf{y}}(\mathbf{y})}, \quad (13)$$

$$p_{y|k,d}(\mathbf{y}|k, d) = \mathcal{CN}(\mathbf{y}; \Phi \boldsymbol{\mu}_k + \mathbf{v}_d, \Phi \Gamma_k \Phi^H + \Lambda_d), \quad (14)$$

$$p_{\mathbf{n}|y,k,d}(\mathbf{n}|\mathbf{y}, k, d) = \mathcal{CN}(\mathbf{n}; \tilde{\boldsymbol{\mu}}_{k,d}, \mathbf{C}_{k,d}), \quad (15)$$

$$\mathbf{C}_{k,d} = (\Phi^H \Lambda_d^{-1} \Phi + \Gamma_k^{-1})^{-1}, \quad (16)$$

$$\tilde{\boldsymbol{\mu}}_{k,d} = \boldsymbol{\mu}_k + \mathbf{C}_{k,d} \Phi^H \Lambda_d^{-1} (\mathbf{y} - \Phi \boldsymbol{\mu}_k - \mathbf{v}_d). \quad (17)$$

In actuality, we learn the distribution using a set of samples

$$\{\mathbf{y}_i = \Phi \mathbf{n}_i + \hat{\mathbf{w}}_i\}, \quad i = 1, \dots, N_s. \quad (18)$$

Here, if we consider each Φ_i as the ‘window’ through which we observe \mathbf{n}_i , having a unique projection matrix for each sample allows us to more fully observe the characteristics of the distribution of \mathbf{N} . However, we will show later that this is not always necessary. For our N_s measurements, we have

$$\ell_{\text{ex-co}}(\theta|\theta^{(t)}) = \sum_{i=1}^{N_s} \mathbb{E}_{\mathbf{n},k,d|\mathbf{y}_i, \theta^{(t)}} \left[\log p_{\mathbf{y},\mathbf{n},k,d}(\mathbf{y}_i, \mathbf{n}, k, d|\theta) \right].$$

After a number of operations, we are able to expand this expression to obtain (19). Here, $\mathbf{C}_{k,d}^{(i)}$ and $\tilde{\boldsymbol{\mu}}_{k,d}^{(i)}$ are the per-sample equivalents of (16) and (17), respectively, and we have again omitted the reliance on $\theta^{(t)}$. If we set the expressions for the gradient of (19) with respect to s_k , $\boldsymbol{\mu}_k$, and Γ_k to zero — noting that $\sum_k s_k = 1$ — we obtain

$$s_k^{(t+1)} = \frac{\sum_{i=1}^{N_s} p_{k|\mathbf{y}}(k|\mathbf{y}_i)}{\sum_{i=1}^{N_s} \sum_{k'=1}^K p_{k'|\mathbf{y}}(k'|\mathbf{y}_i)} = \frac{\sum_{i=1}^{N_s} p_{k|\mathbf{y}}(k|\mathbf{y}_i)}{N_s}, \quad (20)$$

$$\boldsymbol{\mu}_k^{(t+1)} = \frac{\sum_{i=1}^{N_s} \sum_{d=1}^D p_{k,d|\mathbf{y}}(k, d|\mathbf{y}_i) \tilde{\boldsymbol{\mu}}_{k,d}^{(i)}}{\sum_{i=1}^{N_s} p_{k|\mathbf{y}}(k|\mathbf{y}_i)}, \quad (21)$$

$$\Gamma_k^{(t+1)} = \frac{\sum_{i=1}^{N_s} \sum_{d=1}^D p_{k,d|\mathbf{y}}(k, d|\mathbf{y}_i) \left[(\tilde{\boldsymbol{\mu}}_{k,d}^{(i)} - \boldsymbol{\mu}_k) (\tilde{\boldsymbol{\mu}}_{k,d}^{(i)} - \boldsymbol{\mu}_k)^H + \mathbf{C}_{k,d}^{(i)} \right]}{\sum_{i=1}^{N_s} p_{k|\mathbf{y}}(k|\mathbf{y}_i)}. \quad (22)$$

Thus, we are able to iteratively move towards parameters for \mathbf{N} that better fit our data. The approach ceases after a pre-determined number of iterations or if $\ell_{\text{inco}}(\theta^{(t+1)}|\mathbf{y}) - \ell_{\text{inco}}(\theta^{(t)}|\mathbf{y})$ falls below a specified threshold.

V. ADAPTIVE INFORMATION-THEORETIC ALGORITHM

We combine the pre-existing projection design methodology of Sec. III with our proposed distribution learning approach. A pseudocode representation of the resulting framework is provided in Algorithm 1. The algorithm initialises with a trained $\Phi_{\text{opt}} \in \mathbb{C}^{m_{\text{opt}} \times n}$, which has been designed subject to the objective function of (4) and some prior knowledge of the system parameters θ_{opt} . If there is initially no knowledge of a secondary source, \mathbf{N} is considered absent with only one class as defined in Sec. III.

The algorithm captures compressive samples at regular intervals. If the average log-likelihood of the past N_{avg} compressive samples falls below a predefined threshold ζ_{avg} , the algorithm begins a compressive sampling process using random matrices $\Phi_{i_{\text{CS}}} \in \mathbb{C}^{m \times n}$ with elements drawn from $\mathcal{CN}(0, 1)$. Each matrix $\Phi_{i_{\text{CS}}}$ is reused N_{rep} times. If the number of consecutive samples with low average log-likelihood exceeds a predefined minimum N_s^{min} , up to N_s^{max} of these samples will be used to learn the GM distribution of a new class of \mathbf{N} . This learning process will occur according to the iterative approach described in Sec. IV and will cease after a predefined number of iterations or if the change in log-likelihood across iterations falls below a predefined threshold. When updating the current system parameters θ with the new class of \mathbf{N} , the class probabilities for \mathbf{N} are updated to match their likelihoods in the previous N_T samples:

$$r_g \leftarrow \frac{1}{N_T} \sum_{i'=i-N_T+1}^i p_{g|\mathbf{y}}(g|\mathbf{y}_{i'}). \quad (23)$$

Here, N_T is large and defined by the user beforehand. To avoid the retraining of a class with very low likelihood in the future, a threshold ξ can be placed on the class probabilities such that if $r_{g'} < \xi$ for some g' , we set $r_{g'} = \xi$ and, for $g \neq g'$,

$$r_g \leftarrow (1 - \xi) r_g \cdot \left(\sum_{g' \neq g} r_{g'} \right)^{-1}. \quad (24)$$

Samples with an average log-likelihood above the threshold are used for classification and signal reconstruction purposes according to the inference model in [9]. If the number of samples taken is a multiple of N_T and the parameters have changed since the last execution of the projection design step, we redesign Φ_{opt} subject to the objective function of (4).

Note that, if desired, the random matrices $\Phi_{i_{\text{CS}}}$ can be of a different dimensionality to Φ_{opt} ; i.e., we can have $m_{\text{opt}} \neq m$. Using a high m to generate random projections will provide more information about the statistics of \mathbf{N} during the distribution training step; it might therefore be possible to decrease the minimum number of random samples N_s^{min} that are required to obtain a good estimate of the distribution. However, a low m_{opt} might be sufficient for reconstruction or classification purposes. Using $m \neq m_{\text{opt}}$ will, of course, require an additional (assumed known) distribution for the measurement noise $\mathbf{W}_{\text{opt}} \sim \mathcal{CN}(\mathbf{w}_{\text{opt}}; \mathbf{v}_{\text{opt}}, \Lambda_{\text{opt}})$.

VI. EXPERIMENTAL RESULTS

A. Experiments with Synthetic Data

In the following simulations, we use known distributions to generate instances of \mathbf{X} , \mathbf{N} , and \mathbf{W} . We then attempt to recover the distribution of \mathbf{N} using compressive measurements under various conditions. Initially, for simplicity, we constrain the number of classes of \mathbf{X} and \mathbf{N} to $J_x = J_n = 1$, and consider both sources active at all times. The GM distributions are limited to $O = K = 3$ components. We use $N_s = 1000$ random measurements and 500 iterations during the training of the GM for \mathbf{N} . Training ceases if the change in the incomplete log-likelihood between iterations drops below a value of one. Our input dimensionality is $n = 32$ and our compressive measurements are of dimensionality $m \in \{4, 8, 12\}$.

$$\begin{aligned} \ell_{\text{ex-co}}(\theta|\theta^{(t)}) = & \text{constant} - \sum_{i=1}^{N_s} \left\{ \mathbb{E}_{d|y_i} \left[\log \det \Lambda_d \right] + \mathbb{E}_{k|y_i} \left[\log \det \Gamma_k \right] + \mathbb{E}_{k,d|y_i} \left[\text{tr} \left\{ \Lambda_d^{-1} \Phi_i \mathbf{C}_{k,d} \Phi_i^H \right\} \right] - \mathbb{E}_{k|y_i} \left[\log s_k \right] - \mathbb{E}_{d|y_i} \left[\log \tau_d \right] \right. \\ & \left. + \mathbb{E}_{k,d|y_i} \left[\left(y_i - \Phi_i \tilde{\boldsymbol{\mu}}_{k,d}^{(i)} - \mathbf{v}_d \right)^H \Lambda_d^{-1} \left(y_i - \Phi_i \tilde{\boldsymbol{\mu}}_{k,d}^{(i)} - \mathbf{v}_d \right) \right] + \mathbb{E}_{k,d|y_i} \left[\text{tr} \left\{ \Gamma_k^{-1} \mathbf{C}_{k,d}^{(i)} \right\} \right] + \mathbb{E}_{k,d|y_i} \left[\left(\tilde{\boldsymbol{\mu}}_{k,d}^{(i)} - \boldsymbol{\mu}_k \right)^H \Gamma_k^{-1} \left(\tilde{\boldsymbol{\mu}}_{k,d}^{(i)} - \boldsymbol{\mu}_k \right) \right] \right\} \quad (19) \end{aligned}$$

Find the $\Phi_{\text{opt}} \in \mathbb{C}^{m_{\text{opt}} \times n}$ that maximises (4)

$i \leftarrow 0, i_{\text{CS}} \leftarrow 0, \theta \leftarrow \theta_{\text{opt}}$

repeat

$i \leftarrow i + 1, j \leftarrow 0$

Store new sample $y_i \leftarrow \Phi_{\text{opt}}(\mathbf{x}_i + \mathbf{n}_i) + \mathbf{w}_i^{\text{opt}}$

ζ_i^{avg} is the average of N_{avg} last $\zeta_i \leftarrow \log p_{y|\theta}(y_i|\theta)$

if $\zeta_i^{\text{avg}} < \zeta_{\text{thr}}$ **then** $j \leftarrow 1$

if $j = 1$ **and** $i_{\text{CS}} < N_s^{\text{max}}$ **then**

$i_{\text{CS}} \leftarrow i_{\text{CS}} + 1$

if $\text{mod}(i_{\text{CS}} - 1, N_{\text{rep}}) = 0$ **then**

 Generate and store random $\Phi_{i_{\text{CS}}} \in \mathbb{C}^{m \times n}$

else $\Phi_{i_{\text{CS}}} \leftarrow \Phi_{i_{\text{CS}}-1}$

 Store random sample $\tilde{y}_{i_{\text{CS}}} \leftarrow \Phi_{i_{\text{CS}}}(\mathbf{x}_i + \mathbf{n}_i) + \mathbf{w}_i$

else if $i_{\text{CS}} > N_s^{\text{min}}$ **then**

 Learn distribution of new class of N using random measurements and projection matrices

 Update distribution parameters θ , set $i_{\text{CS}} \leftarrow 0$

else

 Reconstruct/classify \mathbf{x}_i and \mathbf{n}_i , set $i_{\text{CS}} \leftarrow 0$

if $\text{mod}(i - 1, N_T) = 0$ **and** $\theta \neq \theta_{\text{opt}}$ **then**

$\theta_{\text{opt}} \leftarrow \theta$, find the Φ_{opt} that maximises (4)

Algorithm 1: Adaptive information-theoretic algorithm.

The weights $\pi_{c,o}$ and $s_{g,k}$ are drawn from the standard uniform distribution and normalised. The mean vectors $\chi_{c,o}$ and $\boldsymbol{\mu}_{g,k}$ comprise elements drawn from the complex Gaussian distribution $\mathcal{CN}(0, \sqrt{2}/10)$, and the covariance matrices $\boldsymbol{\Omega}_{c,o}$ and $\boldsymbol{\Gamma}_{g,k}$ are initially equal to instances of the product $\mathbf{Q}\mathbf{D}\mathbf{Q}^H$, where $\mathbf{Q} \in \mathbb{C}^{n \times n}$ is a random unitary matrix and $\mathbf{D} \in \mathbb{R}^{n \times n}$ is a diagonal matrix with elements drawn from the uniform distribution $\mathcal{U}(10^{-6}, 10^{-2})$. To vary the signal-to-noise ratio (SNR), we adjust the values in the diagonal elements of the matrices \mathbf{D} used to generate $\boldsymbol{\Omega}_{c,o}$. Samples of measurement noise are drawn according to $\mathbf{W} \sim \mathcal{CN}(\mathbf{w}; \mathbf{0}, 10^{-6}\mathbf{I}_m)$ and $\mathbf{W}_{\text{opt}} \sim \mathcal{CN}(\mathbf{w}_{\text{opt}}; \mathbf{0}, 10^{-6}\mathbf{I}_{m_{\text{opt}}})$. For now, elements of Φ_{opt} are drawn from $\mathcal{CN}(0, 1)$. Results are averaged over 100 instances of the simulation scenario. We use the ground truth parameters for N to generate results for comparison purposes.

Generating and storing a unique random projection matrix for each sample used for the training of the distribution of N provides the best possible insight into the source statistics. However, it may be possible to decrease computational costs and memory requirements by reusing the same projection matrix for multiple samples. In Fig. 1, we show that by decreasing the number of unique projection matrices, we increase the resulting mean-square error of reconstruction for N obtained via the inference model of [9]; i.e., our estimate of the distribution of N becomes increasingly inaccurate. However, it is clear that the reliance of the training process on unique

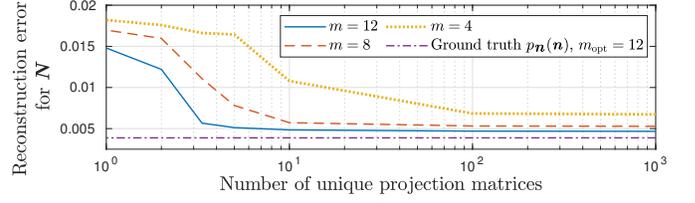


Fig. 1. Mean-square reconstruction error for N versus the number of unique random projection matrices for $m \in \{4, 8, 12\}$ and $m_{\text{opt}} = m$.

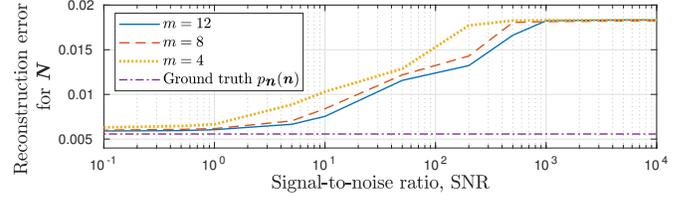


Fig. 2. Mean-square reconstruction error for N versus SNR from (25) when training the distribution of N for $m \in \{4, 8, 12\}$ and $m_{\text{opt}} = 4$.

projections depends on m , with smaller m generally requiring more unique matrices to achieve the best-case results. For a large number of measurements, e.g., for $m = 12$, we see that there are no significant disadvantages to using only 1% of the available unique matrices. We can also observe that by increasing the number of measurements m_{opt} that we use for signal recovery, we decrease the reconstruction error.

The SNR — i.e., the ratio of the power in X to the power in N — will impact our ability to learn the distribution of N . For the results of Fig. 2, we have used various values of SNR during the training of the distribution of N , with

$$\text{SNR} = \text{tr}\{\boldsymbol{\Omega}_{\text{avg}}\} / \text{tr}\{\boldsymbol{\Gamma}_{\text{avg}}\}. \quad (25)$$

Here, $\boldsymbol{\Omega}_{\text{avg}}$ and $\boldsymbol{\Gamma}_{\text{avg}}$ are the average covariance matrices for the ground truth GM distributions of X and N , respectively. All estimated distributions experienced the same test conditions; i.e., we have attempted to reconstruct N in a scenario with X and N of equal power and $m_{\text{opt}} = 4$. The resulting reconstruction error illustrates the quality of each estimate of the distribution of N . We observe that increasing the number of random measurements m improves the distribution estimation in low SNR scenarios. Furthermore, we can see that for very high or very low SNR, all m perform similarly. Significantly, we see that we are unable to estimate the distribution for SNRs of order 10^3 and above, as the reconstruction error is no longer increasing; i.e., our estimates cannot become worse.

B. Experiments with Real Radar Data

Real radar returns from two fixed-location, three-bladed fans were acquired according to the setup in [9]. The fans possessed three rotation speeds, which can be seen in Table I. Acquisitions for each speed were downsampled to 5.5 kHz. Fans 1 and 2 contribute to the primary and secondary sources, X and N , respectively. A time series \mathbf{r} is the vectorised output

TABLE I
CLASS DESCRIPTIONS BASED ON FAN SPEEDS IN ROTATIONS PER SECOND

Fan	Input	Class 1	Class 2	Class 3	Class 4
1	X	2.63 rps	4.10 rps	5.06 rps	N/A
2	N	5.68 rps	6.21 rps	6.78 rps	Absent

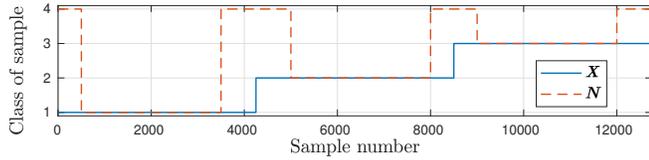


Fig. 3. Ground truth classes of X and N in the examined radar returns.

of the radar receiver system; \mathbf{r} is split into R non-overlapping ‘frames’, which are segmented into B overlapping ‘bursts’. Each burst is Hamming windowed and transformed to the frequency domain via the discrete Fourier transform.

We obtain ground truth distributions for the GMMs of $X, N \in \mathbb{C}^n$ via the EM algorithm [12]. For this, instances of X and N are obtained from transformed bursts when \mathbf{r} contains radar returns from either source in isolation. Training data for this is obtained from 50 frames of data recorded for each fan speed. For feature extraction purposes, we use a frame length of 700. As in [9], we limit the number of frequency coefficients (and therefore the dimensionalities of $X, N \in \mathbb{C}^n$) to $n = 32$. Each burst overlaps its neighbours by 75%.

We initialise Algorithm 1 by designing the matrix Φ_{opt} such that $I(C; Y)$ is maximised; i.e., we use $\beta = [0, 1, 0, 0]$ in (4). Matrix Φ_{opt} is designed over 10^3 iterations using a step size of $\delta = 0.01$ and 500 MC draws to evaluate $\nabla_{\Phi} F(\Phi, \beta)$. Our initial system parameters θ_{opt} include $\mathbf{W} \sim \mathcal{CN}(\mathbf{0}, 10^{-6}\mathbf{I}_m)$, $\mathbf{W}_{\text{opt}} \sim \mathcal{CN}(\mathbf{0}, 10^{-6}\mathbf{I}_{m_{\text{opt}}})$, and the GMM for X with class probabilities $z_c = 1/3 \forall c$ and $O = 3$ components. We assign $K=3$ components to each class of N and begin with $N \sim p_n^{\text{init}}(\mathbf{n}) = \sum_{k=1}^K (1/K) \mathcal{CN}(\mathbf{0}, 10^{-6}\mathbf{I}_n)$. The algorithm is applied to a sequence of radar return data of length $N_T = 12750$ in which the ground truth classes of X and N from Table I are changing according to Fig. 3. Note that Fan 2 is absent for class 4 of N . We use the following parameters: $m \in \{4, 6, 8, 12, 16\}$, $m_{\text{opt}} \in \{4, 6, 8\}$, $N_{\text{avg}} = 100$, $N_{\text{rep}} = 1$, $N_s^{\text{max}} = 3000$, $N_s^{\text{min}} = 1000$, and $\zeta_{\text{thr}} = 2.5$. When learning GMs for N , we use the parameters of Sec. VI-A.

Fig. 4 shows the burst classification accuracies for X obtained after applying Algorithm 1 to the examined radar returns. Note that the retraining of Φ_{opt} again used $\beta = [0, 1, 0, 0]$. Clearly, increasing m has improved our ability to classify X . This indicates that, as in Sec. VI-A, a large m provides a better estimate of $p_n(\mathbf{n})$. With a better estimate, we are able to obtain a superior Φ_{opt} . As in [9], increasing m_{opt} also improves our classification accuracy. Note that classifying on a per-frame basis [9] can further increase performance.

VII. CONCLUSIONS

In this paper, we have derived a methodology for the training of the GM distribution of a secondary input via compressive measurements. We have shown that well-estimated distributions yield designed projection matrices that are more

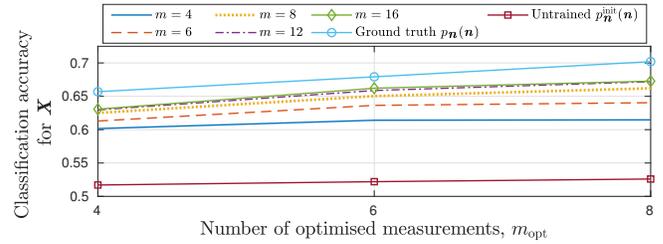


Fig. 4. Burst classification accuracy for X versus the number of optimised measurements m_{opt} for $m \in \{4, 6, 8, 12, 16\}$ random measurements.

able to control the input-output MI of a system. Furthermore, we have demonstrated that increasing the number of compressive measurements aids the characterisation of weak secondary sources and can reduce the number of unique projection matrices required for distribution training purposes.

ACKNOWLEDGEMENTS

This work was supported by the Engineering and Physical Sciences Research Council of the UK (EPSRC) Grant number EP/S000631/1 and the UK MOD University Defence Research Collaboration (UDRC) in Signal Processing. The authors would like to thank Prof. Mike Davies, Prof. Mathini Sellathurai, and Dr João Mota for their valuable input. Gratitude is also extended to Dr Domenico Gaglione, Dr Christos Ilioudis, and Dr Carmine Clemente for capturing the radar data.

REFERENCES

- [1] E. Candes and M. Wakin, ‘‘An introduction to compressive sampling,’’ *IEEE Signal Process. Mag.*, vol. 25, no. 2, pp. 21–30, Mar. 2008.
- [2] J. Duarte-Carvajalino and G. Sapiro, ‘‘Learning to sense sparse signals: Simultaneous sensing matrix and sparsifying dictionary optimization,’’ *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1395–1408, July 2009.
- [3] Z. Nenadic, ‘‘Information discriminant analysis: Feature extraction with an information-theoretic objective,’’ *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 8, pp. 1394–1407, Aug. 2007.
- [4] L. Wang *et al.*, ‘‘Information-theoretic compressive measurement design,’’ *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1150–1164, June 2017.
- [5] G. Yu and G. Sapiro, ‘‘Statistical compressed sensing of Gaussian mixture models,’’ *IEEE Trans. Signal Process.*, vol. 59, no. 12, pp. 5842–5858, Dec. 2011.
- [6] B. Paul, C. D. Chapman, A. R. Chiriyath, and D. W. Bliss, ‘‘Bridging mixture model estimation and information bounds using I-MMSE,’’ *IEEE Trans. Signal Process.*, vol. 65, no. 18, pp. 4821–4832, Sept. 2017.
- [7] F. Renna, R. Calderbank, L. Carin, and M. Rodrigues, ‘‘Reconstruction of signals drawn from a Gaussian mixture via noisy compressive measurements,’’ *IEEE Trans. Signal Process.*, vol. 62, no. 9, pp. 2265–2277, May 2014.
- [8] F. K. Coutts, J. Thompson, and B. Mulgrew, ‘‘Gradient of mutual information in linear vector Gaussian channels in the presence of input noise,’’ in *Proc. 28th Eur. Signal Process. Conf.*, 2021, pp. 2264–2268.
- [9] —, ‘‘Information-theoretic compressive measurement design for micro-Doppler signatures,’’ in *Proc. Sens. Signal Process. Defence*, 2020, pp. 1–5.
- [10] V. C. Chen, *The micro-Doppler effect in radar*, 1st ed. Norwood, MA: Artech House, 2011.
- [11] J. Yang, X. Liao, X. Yuan, P. Lull, D. J. Brady, G. Sapiro, and L. Carin, ‘‘Compressive sensing by learning a Gaussian mixture model from measurements,’’ *IEEE Trans. Image Process.*, vol. 24, no. 1, pp. 106–119, Jan 2015.
- [12] A. P. Dempster, N. M. Laird, and D. B. Rubin, ‘‘Maximum likelihood from incomplete data via the EM algorithm,’’ *J. of the Royal Statistical Society. Series B (Methodological)*, vol. 39, no. 1, pp. 1–38, 1977.
- [13] D. J. MacKay, *Information Theory, Inference, and Learning Algorithms*. Cambridge University Press, 2003.

Author Index

A		K	
Addabbo, Pia	6	Khattak, Faizan	40
Ahmad, Bashar I.	80	L	
Alinaghi, Atiyeh	90	Lamb, R.	16
Almasan, Anton-David	36	Lamont-Smith, Tim.....	50
Altmann, Y.	16	Lane, Richard O.	50
B		Leary, Paul.....	11
Balaji, Bhashyam	60	Liang, Jiaming.....	80
Bassett, Robert	11	Luong, David.....	60
Belmekki, Mohamed Amir Alaa	1	M	
Bertram, Stuart J.	50	Macdonald, Sofie	26
Biondi, Filippo	6	Maskell, Simon.....	85
Blouin, Stéphane	55	Matthews, James	65
C		McLaughlin, S.	16
Clemente, Carmine.....	6	McLaughlin, Stephen	1
Clerckx, Bruno	70	Mota, João F.C.	95
Coutts, Fraser K.	21,101	Mulgrew, Bernard.....	101
Couturier, Antonin.....	36	N	
D		Narykov, Alexey	85
Davidson, Nigel.....	45	O	
Davies, Mike E.	75	Orlando, Danilo	6
Delaosa, Connor.....	65	P	
Dizdar, Onur	70	Pawlikowska, A.	16
Dong, Jun	45	Proudler, Ian	75,90
Drummond, K.	16	Proudler, Ian K.	40,65
F		R	
Fiscante, Nicomino	6	Rajan, Sreeraman.....	60
Foster, Jacob	11	Ralph, Jason	85
G		S	
Gemba, Kay L.	11	Sarkar, Amit	31
Ghosh, Debalina	31	Smith, Kevin B.	11
Giunta, Gaetano	6	Stankovic, Vladimir	45,90
Godsill, Simon.....	80	Sun, Mengwei	75
H		T	
Halimi, Abderrahim	1	Thompson, John	21,101
Hamadouche, Anis	95	U	
Holmes, Wendy J.	50	Üney, Murat	85
Hopgood, James R.	26,75	W	
J		Wallace, Andrew M.	95
Jackson, Ben A.	65		
Jahangir, Mohammad	80		

Weiss, Stephan.....40,65,90
Wragge, Adam J.50
Wu, Yun95

Z

Zhang, Kaiyu.....21

SSPD 2022

Sensor Signal Processing for Defence Conference

13th - 14th September

IET London: Savoy Place

Important Dates:

Submission of Papers: 15 April 2022

Notification of Paper Acceptance: 1 July 2022

Final version of Paper Due: 31 July 2022

Papers are solicited from the following areas:-

- Array Signal Processing
- Image Processing
- Radar, Sonar and Acoustic
- Multimodal Signal Processing
- Multi-Target Tracking
- Signal Acquisition and Sensor Management
- Multiple-input and multiple-output (MIMO)
- Deep Learning, Machine Learning
- Information/Data Analysis
- Data Fusion
- Source Separation
- Anomaly Detection
- Distributed Signal Processing
- Low Size Weight & Power Solutions
- Target Detection and Identification
- Electro-Optic Sensing

All submitted papers will be peer reviewed. Technical sponsorship is provided by the IEEE Signal Processing Society and proceedings will be submitted to the Xplore Digital Library.

www.sspdconference.org