Learning Low-Rank Models From Compressive Measurements for Efficient Projection Design

Fraser K. Coutts, John Thompson, and Bernard Mulgrew

Institute for Digital Communications
University of Edinburgh
fraser.coutts@ed.ac.uk

11th Sensor Signal Processing for Defence Conference
14th September 2022

Presentation Overview

1. Background, motivation, and signal model.
2. Learning low-rank source distributions from compressed data.
3. Efficiently adapting compression strategies to accommodate new sources.
4. Adaptive algorithm pairing source learning with optimised compression.
5. Experimental results using real radar data and conclusions.

Background and Motivation

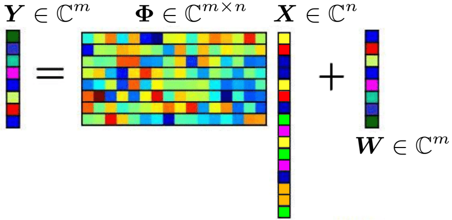
- Designing effective compression strategies is an important problem for both civilian and defence applications.
- In general, such strategies must dispose of some information to reduce complexity and memory requirements down the signal processing chain.
- In particular, when designing solutions that are constrained by low size, weight, and power (SWAP) requirements, data reduction is a primary step.

Background and Motivation

- Dimensionality reduction methods based on linear random projections — i.e., compressive sensing (CS) — have gained significant attention recently.

$$\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{X} + \boldsymbol{W}$$

$$m \ll n$$

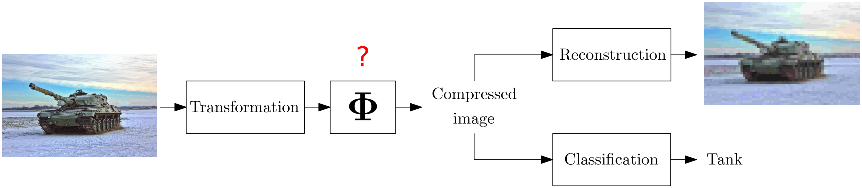


- However, random projections may not be the best choice for $\boldsymbol{\Phi}$ if we know the statistical properties of the signal $\boldsymbol{X}$.

Background and Motivation

$$\boldsymbol{Y} = \boldsymbol{\Phi}\boldsymbol{X} + \boldsymbol{W}$$

- Lower dimensionality brings memory and computational benefits.
- Signal model has various applications in defence:
  - Example: $\boldsymbol{X}$ represents transformed high dimensional image data.
  - How to find the $\boldsymbol{\Phi}$ that best facilitates the reconstruction or classification of $\boldsymbol{X}$?

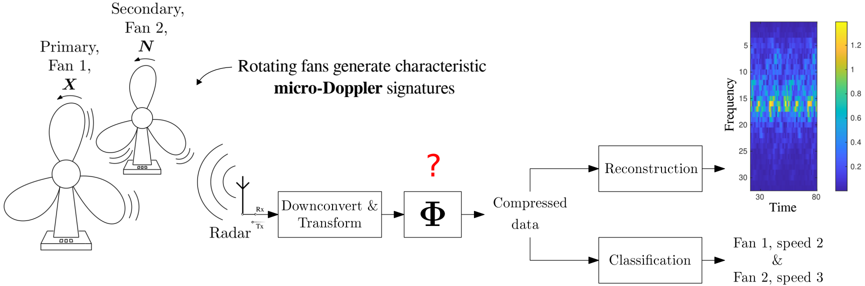


Background and Motivation

- Our recent work has focussed on finding the optimal $\boldsymbol{\Phi}$ in scenarios with noise $\boldsymbol{N}$ present on the input signal $\boldsymbol{X}$:

$$\boldsymbol{Y} = \boldsymbol{\Phi}(\boldsymbol{X} + \boldsymbol{N}) + \boldsymbol{W}$$

- Example: $\boldsymbol{X}$ represents a source generating radar return data; $\boldsymbol{N}$ can be random noise, a secondary source, or clutter.



Background and Motivation

$$\boldsymbol{Y} = \boldsymbol{\Phi}(\boldsymbol{X} + \boldsymbol{N}) + \boldsymbol{W}$$

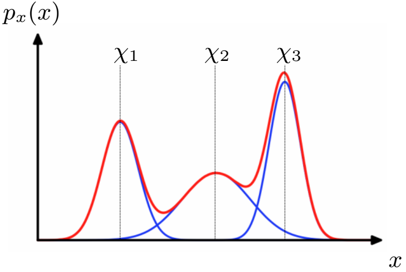
- By employing an information-theoretic approach, we design a linear projection $\boldsymbol{\Phi}$ that balances:
  - the reconstruction error for $\boldsymbol{X}$ and $\boldsymbol{N}$;
  - the classification accuracy for $\boldsymbol{X}$ and $\boldsymbol{N}$.
- Naturally, there is a trade off: we cannot increase performance on all fronts without compromise.
- Model:
  - Measurement noise $\boldsymbol{W}$ is Gaussian distributed.
  - Inputs $\boldsymbol{X}$ and $\boldsymbol{N}$ are **Gaussian Mixture (GM)** distributed.

Background and Motivation

- In general, a single Gaussian does not provide a sufficiently accurate description of source signals.

- Instead, the distribution of a non-Gaussian signal can be approximated by a mixture of several Gaussians, e.g.,

$$X \sim \sum_{c=1}^{J_x} \pi_c \mathcal{N}(x; \chi_c, \Omega_c)$$



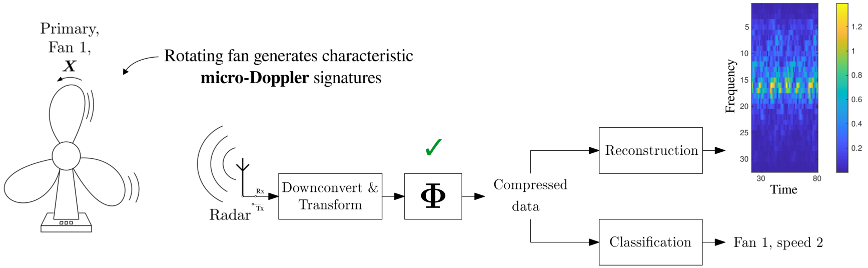
- In CS, such models have been proven to be effective and in some cases superior to sparse signal models (Yu and Sapiro, 2011).

Background and Motivation

- With good GM approximations to the distributions of $\boldsymbol{X}$ and $\boldsymbol{N}$, we can therefore design an appropriate $\boldsymbol{\Phi}$ for a given application.
- In defence applications, we might assume a good understanding of $\boldsymbol{X}$, but expect $\boldsymbol{N}$ to change through the appearance of new — potentially adversarial — secondary sources.
- Adequately extracting information from or mitigating such secondary sources could be vital.

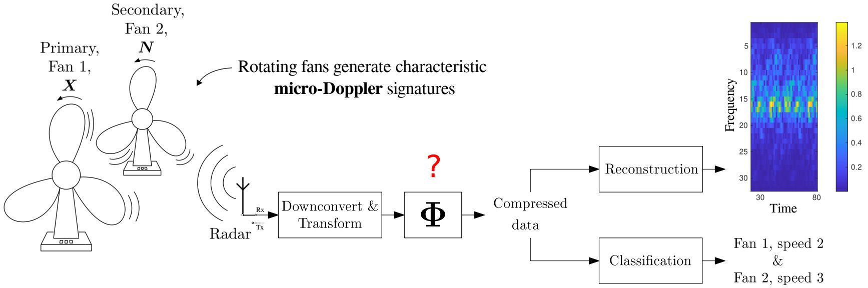
Background and Motivation

- Example scenario: we have a well-characterised primary source $\boldsymbol{X}$ measured in isolation.
- With *a priori* knowledge of the distributions, we can design $\boldsymbol{\Phi}$ such that we can use $\boldsymbol{Y}$ to accurately reconstruct/classify $\boldsymbol{X}$.



Background and Motivation

- Example scenario: we have a well-characterised primary source $\boldsymbol{X}$ measured in isolation.
- However, the appearance of a new secondary source, $\boldsymbol{N}$, will disrupt the performance of the designed $\boldsymbol{\Phi}$.



Background and Motivation

- Our recent work (Coutts *et al.*, 2021) addressed this issue by giving specific attention to the learning of secondary information sources via compressive measurements $\boldsymbol{Y}$
  - i.e., **without accessing the source data directly**.
- After updating our approximation to the distribution of the secondary source $\boldsymbol{N}$, we can redesign our compression operator $\boldsymbol{\Phi}$.
- In low SWAP applications, we want to **efficiently** learn a GM approximation to the distribution of $\boldsymbol{N}$, and **quickly** redesign $\boldsymbol{\Phi}$.

Background and Motivation

- Our recent adaptive projection design approach can deal with new or changing secondary sources.
- However, its memory and computational complexity requirements are not ideal for online, **low SWAP** implementations.
- Here, we explore novel extensions of existing methods to test if lower complexity options are available for GM-based source learning and information-theoretic projection design.

Background and Motivation

- Three main novel contributions:
  1. Techniques to learn low-rank GM approximations to secondary source distributions from compressive measurements.
  2. Insight into the complexity reductions possible during projection design when incorporating low-rank GM distributions.
  3. Two alternative projection design strategies, which we test against our established strategy to determine if cost savings can be achieved via algorithms with faster convergence.

Learning a Secondary Source From Compressive Measurements

- Learning the distribution of a GM distributed $\boldsymbol{X}$ from compressive measurements has been covered by Yang *et al.* in 2015 for a signal model without input $\boldsymbol{N}$:

$$\boldsymbol{Y}' = \boldsymbol{\Phi}\boldsymbol{X} + \boldsymbol{W}$$

- Measurement noise $\boldsymbol{W}$ is Gaussian distributed.

- We extended this approach to our chosen signal model last year (Coutts *et al.*, 2021) to learn an unknown GM-distributed $\boldsymbol{N}$ with known GM-distributed $\boldsymbol{X}$:

$$\overset{\checkmark}{\boldsymbol{Y}} = \overset{\checkmark}{\boldsymbol{\Phi}}(\overset{\checkmark}{\boldsymbol{X}} + \overset{?}{\boldsymbol{N}}) + \overset{\checkmark}{\boldsymbol{W}}$$

Learning a Secondary Source From Compressive Measurements

- To learn the GM distribution of a secondary source $\boldsymbol{N}$, we used an expectation-maximisation (EM) approach.

- We obtain the parameters $\theta = \{s_k, \boldsymbol{\mu}_k, \boldsymbol{\Gamma}_k\}$ of $\boldsymbol{N}$ that best fit our data, with

$$\boldsymbol{N} \sim \sum_{k=1}^{K} s_k \, \mathcal{CN}(\boldsymbol{n}; \boldsymbol{\mu}_k, \boldsymbol{\Gamma}_k)$$

- s_k: weight of kth Gaussian (likelihood of this component).
- $\boldsymbol{\mu}_k$: mean vector of kth Gaussian.
- $\boldsymbol{\Gamma}_k$: covariance matrix of kth Gaussian.

Learning a Secondary Source From Compressive Measurements

- The computational complexity and memory requirements of this approach increase with the signal dimension n, due to the required manipulation of the GM covariance matrices.

- We therefore impose a near-low-rank structure on the covariance matrices of the learned secondary source $\boldsymbol{N}$, such that we have

$$\boldsymbol{\Gamma}_k = \mathbf{F}_k \mathbf{F}_k^{\mathrm{H}} + \eta\, \mathbf{I}_n\,, \quad k = 1, \ldots, K\,,$$

  where $\mathbf{F}_k \in \mathbb{C}^{n \times r_k}$, $r_k \ll n$, and $0 < \eta \ll 1$.

- By manipulating the 'tall' matrix $\mathbf{F}_k$ instead of $\boldsymbol{\Gamma}_k$, we reduce our memory footprint and computational costs.

Learning a Secondary Source From Compressive Measurements

- We capture N_s compressive measurements using randomly generated projection matrices $\mathbf{\Phi}_i$:

$$\boldsymbol{y}_i = \mathbf{\Phi}_i(\boldsymbol{x}_i + \boldsymbol{n}_i) + \boldsymbol{w}_i \ , \quad i = 1, \ldots, N_s$$

and seek to obtain the parameters $\theta = \{s_k, \boldsymbol{\mu}_k, \mathbf{F}_k\}$ of $\boldsymbol{N}$ that best fit our data.

Learning a Secondary Source From Compressive Measurements

- We capture N_s compressive measurements using randomly generated projection matrices $\mathbf{\Phi}_i$:

$$\boldsymbol{y}_i = \mathbf{\Phi}_i(\boldsymbol{x}_i + \boldsymbol{n}_i) + \boldsymbol{w}_i \ , \quad i = 1, \ldots, N_s$$

and seek to obtain the parameters $\theta = \{s_k, \boldsymbol{\mu}_k, \mathbf{F}_k\}$ of $\boldsymbol{N}$ that best fit our data.

- We seek the θ that maximises the marginal log-likelihood

$$\ell(\theta|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{N_s}) = \sum_{i=1}^{N_s} \log p_{\boldsymbol{y}|\theta}(\boldsymbol{y}_i|\theta)$$

Learning a Secondary Source From Compressive Measurements

- Since the marginal log-likelihood is difficult to maximise directly, we take an iterative **expectation-maximisation** approach.
- At iteration $(t+1)$:
  1. Find the likelihood of the unobserved variables given access to compressive measurements $\boldsymbol{y}_i$ and previous system parameters $\theta^{(t)}$.
  2. Update the system parameters such that the expected value of the complete log-likelihood is maximised.

Learning a Secondary Source From Compressive Measurements

- For the maximisation in step 2, we have identified closed-form solutions for the update of the GM parameters for $\boldsymbol{N}$.
- The new parameters are guaranteed to satisfy

$$\ell(\theta^{(t+1)}|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{N_s}) \geq \ell(\theta^{(t)}|\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{N_s}),$$

i.e., the likelihood is always increasing (until a local maximum is reached).

Alternative Projection Design Algorithms

- We can now efficiently learn low-rank GM approximations to the distribution of N via compressive measurements.
- Low-rank optimisations can also result in reduced computational complexity during projection design.
- However, it would still be worthwhile to explore alternative projection design strategies to determine if faster convergence can be attained.
- Algorithms that converge more quickly than existing methods (Coutts *et al.*, 2020) may yield a lower system complexity.

Alternative Projection Design Algorithms

- Applying the singular value decomposition and eigenvalue decomposition to the projection matrix and measurement noise covariance matrix, respectively, yields

$$\mathbf{\Phi} = \mathbf{U}_{\mathbf{\Phi}}\mathbf{D}_{\mathbf{\Phi}}\mathbf{V}_{\mathbf{\Phi}}^{\mathrm{H}}, \qquad \mathbf{\Lambda} = \mathbf{U}_{\mathbf{\Lambda}}\mathbf{D}_{\mathbf{\Lambda}}\mathbf{U}_{\mathbf{\Lambda}}^{\mathrm{H}}.$$

Alternative Projection Design Algorithms

- Applying the singular value decomposition and eigenvalue decomposition to the projection matrix and measurement noise covariance matrix, respectively, yields

$$\mathbf{\Phi} = \mathbf{U}_{\mathbf{\Phi}} \mathbf{D}_{\mathbf{\Phi}} \mathbf{V}_{\mathbf{\Phi}}^{\mathrm{H}}\,, \qquad \mathbf{\Lambda} = \mathbf{U}_{\mathbf{\Lambda}} \mathbf{D}_{\mathbf{\Lambda}} \mathbf{U}_{\mathbf{\Lambda}}^{\mathrm{H}}\,.$$

- If $\mathbf{U}_{\mathbf{\Phi}} = \mathbf{U}_{\mathbf{\Lambda}}$, we can redefine our signal model:

$$\bar{\boldsymbol{Y}} = \mathbf{D}_{\mathbf{\Lambda}}^{-1/2} \mathbf{U}_{\mathbf{\Lambda}}^{\mathrm{H}} \boldsymbol{Y} = \mathbf{H} \mathbf{D}_{\mathbf{\Phi}} \mathbf{\Theta} (\boldsymbol{X} + \boldsymbol{N}) + \bar{\boldsymbol{W}}\,,$$

with $\mathbf{H} = \mathbf{D}_{\mathbf{\Lambda}}^{-1/2}$ and $\mathbf{\Theta} = \mathbf{V}_{\mathbf{\Phi}}^{\mathrm{H}}$.

Alternative Projection Design Algorithms

$$\boldsymbol{Y} = \overset{?}{\boldsymbol{\Phi}}(\overset{\checkmark}{\boldsymbol{X}} + \overset{\checkmark}{\boldsymbol{N}}) + \overset{\checkmark}{\boldsymbol{W}} \longrightarrow \bar{\boldsymbol{Y}} = \overset{\checkmark}{\mathbf{H}}\overset{?}{\mathbf{D}}_{\boldsymbol{\Phi}}\overset{?}{\boldsymbol{\Theta}}(\overset{\checkmark}{\boldsymbol{X}} + \overset{\checkmark}{\boldsymbol{N}}) + \overset{\checkmark}{\bar{\boldsymbol{W}}}$$

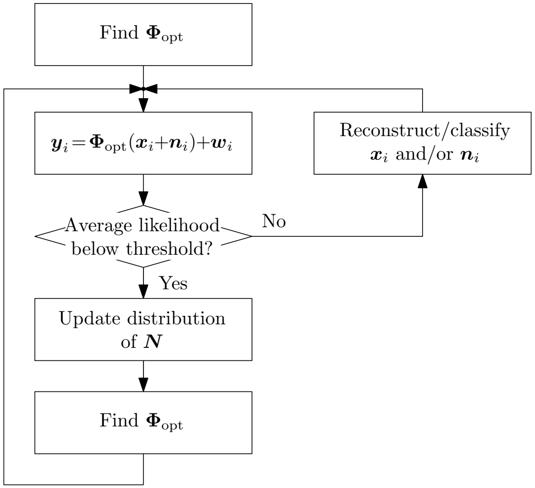
- Via expressions provided in the paper, we establish two alternative projection design strategies that rely on the iterative update of $\mathbf{D}_{\boldsymbol{\Phi}}$ and $\boldsymbol{\Theta}$ such that our objective function is maximised.
- Both approaches update $\mathbf{D}_{\boldsymbol{\Phi}}$ via gradient ascent.
- For the update of $\boldsymbol{\Theta}$, we have two options:
  1. Use gradient ascent and project to the set of orthonormal matrices.
  2. Use a Given's rotation-based approach that optimises the discrete rotation operations that exist within $\boldsymbol{\Theta}$.

Adaptive Algorithm

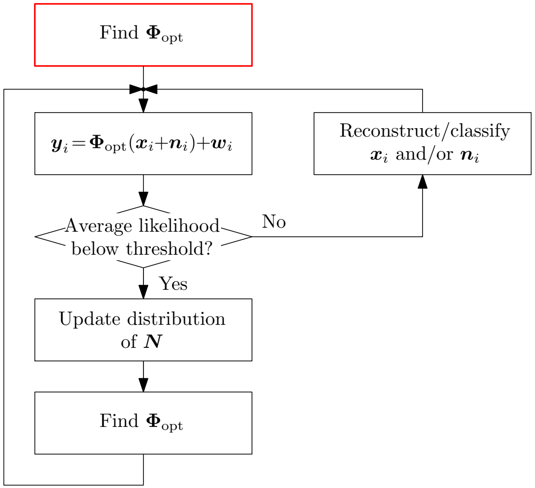
$$\boldsymbol{Y} = \boldsymbol{\Phi}_{\text{opt}}(\boldsymbol{X} + \boldsymbol{N}) + \boldsymbol{W}$$

- We now have access to:
  - Three algorithms that can seek the optimal projection matrix $\boldsymbol{\Phi}_{\text{opt}}$ given accurate estimations of the source distributions.
  - Techniques to update a low-rank estimate of the GM distribution for $\boldsymbol{N}$ if the current distribution is found to be innaccurate (via a likelihood test).
- We can now create an adaptive algorithm that updates $\boldsymbol{\Phi}_{\text{opt}}$ to account for a changing secondary source distribution.

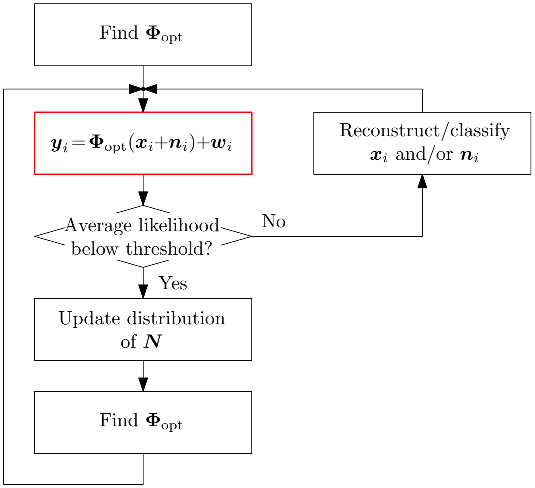
Adaptive Algorithm



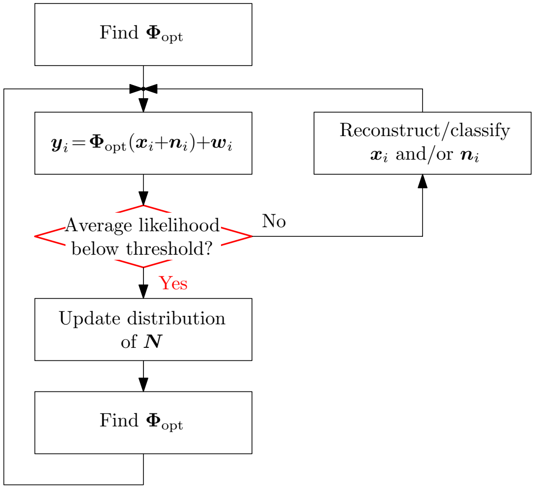
Adaptive Algorithm



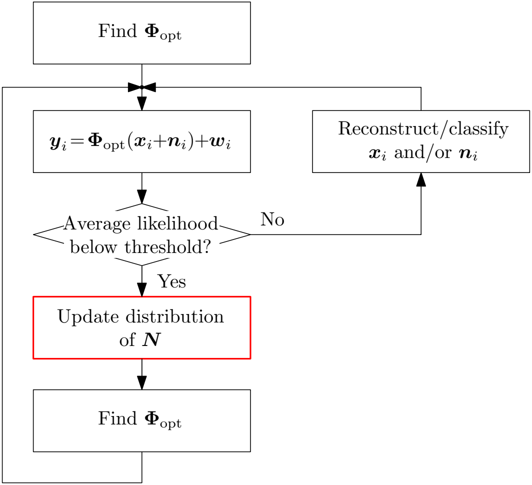
Adaptive Algorithm



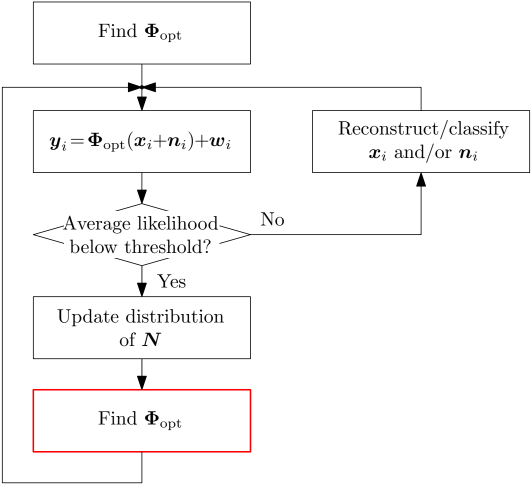
Adaptive Algorithm



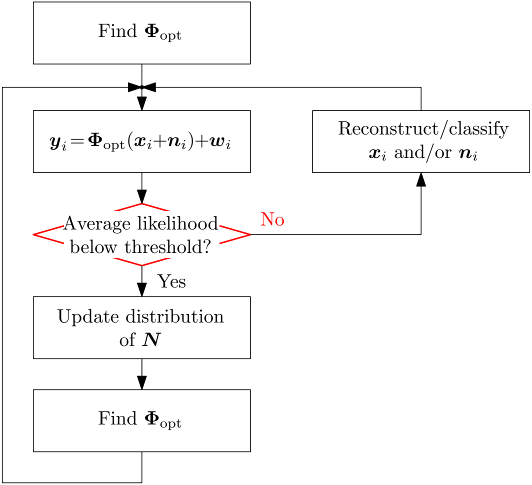
Adaptive Algorithm



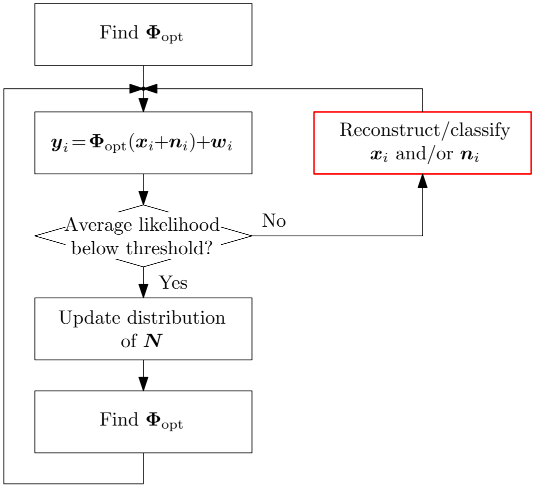
Adaptive Algorithm



Adaptive Algorithm



Adaptive Algorithm



Results Using Synthetic Data — Quality of Low-Rank Approximations

- Objectives: estimate distribution of $\boldsymbol{N}$ given 10^3 measurements; investigate impact of under/overestimating true rank, r:

Results Using Synthetic Data — Quality of Low-Rank Approximations

- Objectives: estimate distribution of $\boldsymbol{N}$ given 10^3 measurements; investigate impact of under/overestimating true rank, r:

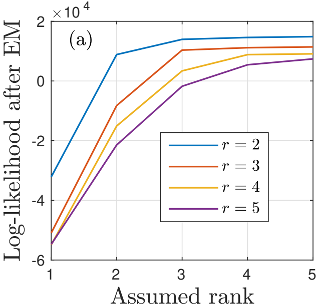


Figure: (a) Log-likelihood of compressive measurements after source learning via EM versus assumed rank of covariance matrices $\mathbf{\Gamma}_k$ for actual ranks $r \in \{2, 3, 4, 5\}$. Source data has dimensionality $n = 16$.

Results Using Synthetic Data — Quality of Low-Rank Approximations

- Objectives: estimate distribution of $\boldsymbol{N}$ given 10^3 measurements; investigate impact of under/overestimating true rank, r:

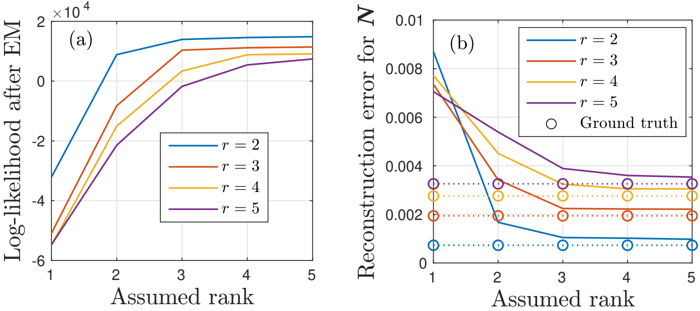


Figure: (a) Log-likelihood of compressive measurements after source learning via EM and (b) reconstruction error for $\boldsymbol{N}$ versus assumed rank of covariance matrices $\boldsymbol{\Gamma}_k$ for actual ranks $r \in \{2, 3, 4, 5\}$. Source data has dimensionality $n = 16$.

Results Using Synthetic Data — Quality of Low-Rank Approximations

- Objectives: estimate distribution of $\boldsymbol{N}$ given 10^3 measurements; investigate impact of under/overestimating true rank, r:
- Increasing the assumed rank improves performance until true rank exceeded.
- What is the impact of assumed rank on source learning complexity?

Table: Source learning run time versus assumed rank

Rank	1	2	3	4	5
Time	61.6 s	69.6 s	76.0 s	82.4 s	90.3 s

Results Using Synthetic Data — Projection Design Complexity

- Objective: evaluate projection design complexity for different covariance matrix ranks and compressive measurement dimensions, m:

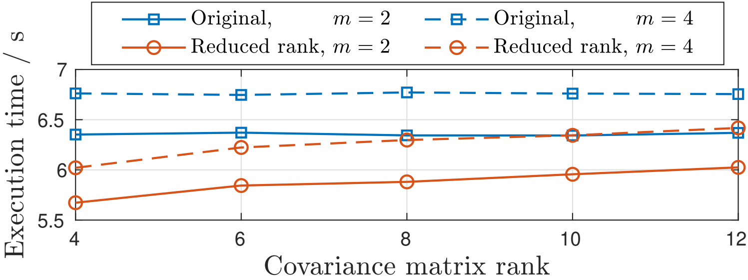
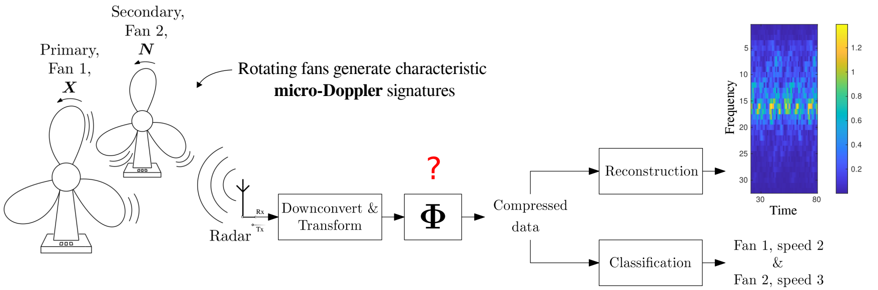


Figure: GM rank versus required projection design execution times. Our original method is compared with a low-rank-optimised version. Source data has dimensionality $n = 16$.

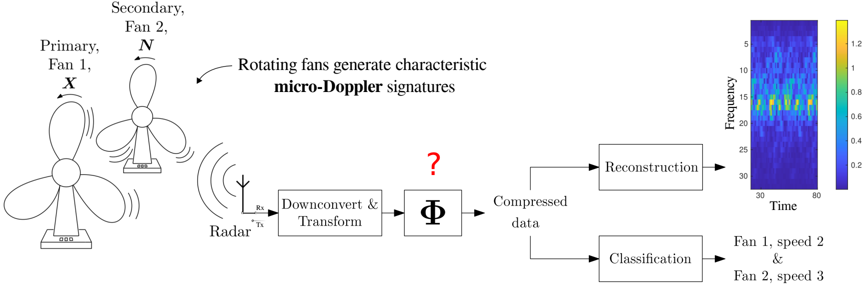
Results Using Real Radar Data

$$\boldsymbol{Y} = \Phi(\boldsymbol{X} + \boldsymbol{N}) + \boldsymbol{W}$$

- Example: $\boldsymbol{X}$ & $\boldsymbol{N}$ represent 2 sources of radar return data.
- 3 fan speeds represent 3 classes.



Results Using Real Radar Data



- Initial scenario: $\boldsymbol{X}$ well characterised, $\boldsymbol{N}$ absent (unknown).
- Simulation: $\boldsymbol{N}$ is **fleeting** & has single class.
- Objective: learn low-rank estimate of distribution of $\boldsymbol{N}$ then update optimal $\boldsymbol{\Phi}_{\text{opt}}$ — e.g., to prioritise classification of $\boldsymbol{X}$.

Results Using Real Radar Data

- Objective: learn low-rank estimate of distribution of $\boldsymbol{N}$ then update optimal $\boldsymbol{\Phi}_{\text{opt}}$ — e.g., to prioritise classification of $\boldsymbol{X}$.
  - Increasing rank improves our estimation of the distribution of $\boldsymbol{N}$, which has dimensionality $n = 32$.
  - With a good estimate, we can obtain a better $\boldsymbol{\Phi}_{\text{opt}}$.

Table: Classification accuracy (CA) for $\boldsymbol{X}$ versus assumed rank

Rank	8	10	12	14	16
CA	60.3%	65.8%	70.6%	71.1%	71.2%

- The classification accuracy when using the original $\boldsymbol{\Phi}_{\text{opt}}$ — without accounting for the presence of $\boldsymbol{N}$ — was 29.7%.

Results Using Synthetic Data — Projection Design Performance

- Objective: evaluate projection design performance for original and proposed update schemes:

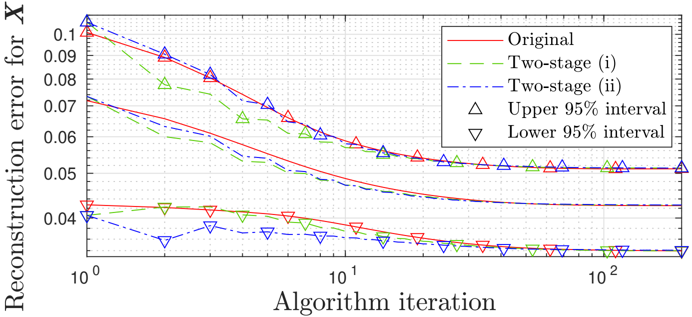


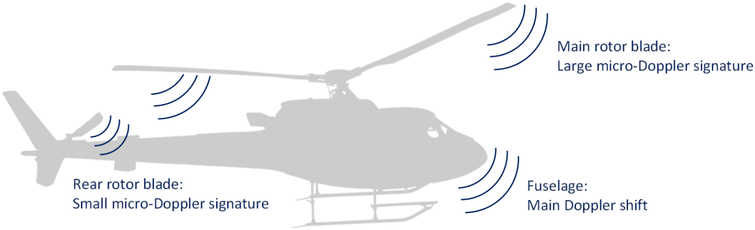
Figure: Reconstruction error for $\boldsymbol{X}$ versus algorithm iteration for the original gradient ascent approach and the developed (i) dual gradient ascent and (ii) Given's rotation-based two-stage implementations.

Conclusions

- Low-rank GM approximations reduce complexity and memory requirements during source learning and projection design.
  - Reducing the rank can decrease computational complexity for low SWAP applications while only slightly lowering performance.
  - The proposed techniques can be extended to applications in which unseen secondary sources of information might appear.
- Two novel projection design strategies introduced.
  - Results indicate that the simple gradient ascent approach from our earlier work is the best choice.
- *Learning Low-Rank Models From Compressive Measurements for Efficient Projection Design*, fraser.coutts@ed.ac.uk.

Future Work

- Additional techniques to identify changes in source distributions.
- Fully online training of source parameters and compression strategies for reconfigurable signal processing.
- Defence applications: e.g., tail rotor blades classification via micro-Doppler recognition.



References

- F. Coutts, J. Thompson, and B. Mulgrew, "Gradient of mutual information in linear vector Gaussian channels in the presence of input noise," *Proc. 28th Eur. Signal Process. Conf.*, 2021, pp. 2264–2268.
- F. Coutts, J. Thompson, and B. Mulgrew, "Information-theoretic compressive measurement design for micro-Doppler signatures," *Proc. Sens. Signal Process. Defence*, 2020, pp. 1–5.
- F. Coutts, J. Thompson, and B. Mulgrew, "Learning a secondary source from compressive measurements for adaptive projection design," *Proc. Sens. Signal Process. Defence*, 2021, pp. 1–5.
- G. Yu and G. Sapiro, "Statistical compressed sensing of Gaussian mixture models," *IEEE Trans. on Signal Process.*, vol. 59, no. 12, pp. 5842–5858, Dec. 2011.
- J. Yang, X. Liao, X. Yuan, P. Llull, 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.
- V. C. Chen, *The micro-Doppler effect in radar.* Artech House, 2019.
- 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.

Background and Motivation

$$\boldsymbol{Y} = \boldsymbol{\Phi}(\boldsymbol{X} + \boldsymbol{N}) + \boldsymbol{W}$$

- By employing an information-theoretic approach, one can design a linear projection $\boldsymbol{\Phi}$ that maximises the mutual information (MI):
  - between $\boldsymbol{Y}$ and the source signal $\boldsymbol{X}$ — i.e., $I(\boldsymbol{X}; \boldsymbol{Y})$;
  - between $\boldsymbol{Y}$ and the discrete classes C of $\boldsymbol{X}$ — i.e., $I(C; \boldsymbol{Y})$;
  - between $\boldsymbol{Y}$ and the source signal $\boldsymbol{N}$ or its discrete classes.
- Intuitively, as the respective MI terms increase, the recovery of the source signal or class information improves.

Background and Motivation

- By optimising the MI, we design a $\mathbf{\Phi}$ that balances signal recovery and classification accuracy for two sources (Coutts *et al.*, 2020).
- Goal: with $\boldsymbol{Y} = \mathbf{\Phi}(\boldsymbol{X} + \boldsymbol{N}) + \boldsymbol{W}$, design $\mathbf{\Phi}$ that maximises

$$F(\mathbf{\Phi}, \boldsymbol{\beta}) = \beta_1 I(\boldsymbol{X}; \boldsymbol{Y}) + \beta_2 I(C; \boldsymbol{Y}) + \beta_3 I(\boldsymbol{N}; \boldsymbol{Y}) + \beta_4 I(G; \boldsymbol{Y})$$

- C and G represent classes of $\boldsymbol{X}$ and $\boldsymbol{N}$, respectively.
- Large β_1 and/or β_3: prioritise **reconstruction**.
- Large β_2 and/or β_4: prioritise **classification**.

Background and Motivation

- By optimising the MI, we design a $\mathbf{\Phi}$ that balances signal recovery and classification accuracy for two sources (Coutts *et al.*, 2020).
- Goal: with $\boldsymbol{Y} = \mathbf{\Phi}(\boldsymbol{X} + \boldsymbol{N}) + \boldsymbol{W}$, design $\mathbf{\Phi}$ that maximises

$$F(\mathbf{\Phi}, \boldsymbol{\beta}) = \beta_1 I(\boldsymbol{X}; \boldsymbol{Y}) + \beta_2 I(C; \boldsymbol{Y}) + \beta_3 I(\boldsymbol{N}; \boldsymbol{Y}) + \beta_4 I(G; \boldsymbol{Y})$$

- Method: iterative update of $\mathbf{\Phi}$ via gradient ascent.
- Model:
  - Measurement noise $\boldsymbol{W}$ is Gaussian distributed.
  - Inputs $\boldsymbol{X}$ and $\boldsymbol{N}$ are Gaussian Mixture (GM) distributed.

Learning a Secondary Source From Compressive Measurements

- Each class c of $\boldsymbol{X}$ is described by a weighted sum of O Gaussian distributions and $\boldsymbol{W}$ is Gaussian:

$$\boldsymbol{X} \sim \sum_{c=1}^{J_{\boldsymbol{x}}} z_c \, p_{\boldsymbol{x}|c}(\boldsymbol{x}|c)$$

$$p_{\boldsymbol{x}|c}(\boldsymbol{x}|c) = \sum_{o=1}^{O} \pi_{c,o} \, \mathcal{CN}(\boldsymbol{x}; \boldsymbol{\chi}_{c,o}, \boldsymbol{\Omega}_{c,o})$$

Learning a Secondary Source From Compressive Measurements

- Each class c of $\boldsymbol{X}$ is described by a weighted sum of O Gaussian distributions and $\boldsymbol{W}$ is Gaussian:

$$\boldsymbol{X} \sim \sum_{c=1}^{J_{\boldsymbol{x}}} z_c \, p_{\boldsymbol{x}|c}(\boldsymbol{x}|c)$$

$$p_{\boldsymbol{x}|c}(\boldsymbol{x}|c) = \sum_{o=1}^{O} \pi_{c,o} \, \mathcal{CN}(\boldsymbol{x}; \boldsymbol{\chi}_{c,o}, \boldsymbol{\Omega}_{c,o})$$

- With $\boldsymbol{W} \sim \mathcal{CN}(\boldsymbol{w}; \boldsymbol{\nu}, \boldsymbol{\Lambda})$, we rearrange to obtain

$$\boldsymbol{Y} = \boldsymbol{\Phi}(\boldsymbol{X} + \boldsymbol{N}) + \boldsymbol{W} = \boldsymbol{\Phi}\boldsymbol{N} + \hat{\boldsymbol{W}}$$

$$\hat{\boldsymbol{W}} \sim \sum_{d=1}^{D} \tau_d \, \mathcal{CN}(\hat{\boldsymbol{w}}; \boldsymbol{\nu}_d, \boldsymbol{\Lambda}_d)$$

Learning a Secondary Source From Compressive Measurements

$$\boldsymbol{y}_i = \boldsymbol{\Phi}_i \boldsymbol{n}_i + \hat{\boldsymbol{w}}_i \ , \quad i = 1, \ldots, N_s$$

- Important considerations when estimating the distribution of $\boldsymbol{N}$:
  - Unique $\{\boldsymbol{\Phi}_i\}_{i=1}^{N_s}$ will improve estimation but add cost.

Learning a Secondary Source From Compressive Measurements

$$\boldsymbol{y}_i = \boldsymbol{\Phi}_i \boldsymbol{n}_i + \hat{\boldsymbol{w}}_i \ , \quad i = 1, \ldots, N_s$$

- Important considerations when estimating the distribution of $\boldsymbol{N}$:
  - Unique $\{\boldsymbol{\Phi}_i\}_{i=1}^{N_s}$ will improve estimation but add cost.
  - Increasing m (the number of rows in $\boldsymbol{\Phi}_i \in \mathbb{C}^{m \times n}$) will improve estimation.

Learning a Secondary Source From Compressive Measurements

$$\boldsymbol{y}_i = \boldsymbol{\Phi}_i \boldsymbol{n}_i + \hat{\boldsymbol{w}}_i \ , \quad i = 1, \ldots, N_s$$

- Important considerations when estimating the distribution of $\boldsymbol{N}$:
  - Unique $\{\boldsymbol{\Phi}_i\}_{i=1}^{N_s}$ will improve estimation but add cost.
  - Increasing m (the number of rows in $\boldsymbol{\Phi}_i \in \mathbb{C}^{m \times n}$) will improve estimation.
  - Increasing the number of samples N_s will improve estimation.

Learning a Secondary Source From Compressive Measurements

$$\boldsymbol{y}_i = \boldsymbol{\Phi}_i \boldsymbol{n}_i + \hat{\boldsymbol{w}}_i \ , \quad i = 1, \ldots, N_s$$

- Important considerations when estimating the distribution of $\boldsymbol{N}$:
  - Unique $\{\boldsymbol{\Phi}_i\}_{i=1}^{N_s}$ will improve estimation but add cost.
  - Increasing m (the number of rows in $\boldsymbol{\Phi}_i \in \mathbb{C}^{m \times n}$) will improve estimation.
  - Increasing the number of samples N_s will improve estimation.
  - Increasing the power of $\boldsymbol{N}$ relative to $\boldsymbol{X}$ will improve estimation.