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
fraser.coutts@ed.ac.uk

10th Sensor Signal Processing for Defence Conference
15th September 2021

Engineering and Physical Sciences Research Council

Presentation Overview

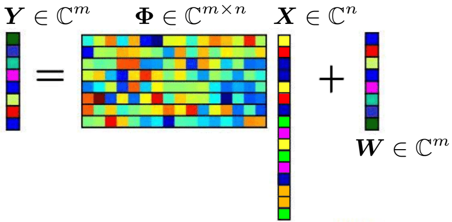
1. Background, motivation, and signal model.
2. Existing information-theoretic optimisation framework.
3. Learning a secondary source distribution from compressive measurements.
4. Adaptive information-theoretic algorithm combining 2 and 3.
5. Experimental results using real radar data and conclusions.

Background and Motivation

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

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

$$m \ll n$$

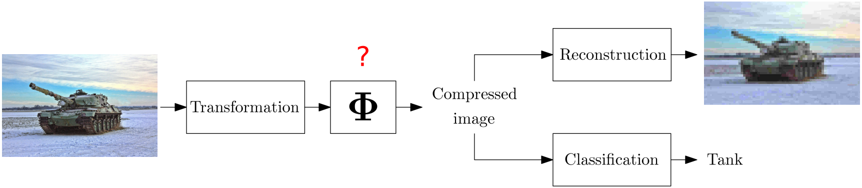


- However, random projections may not be the best choice if we know the statistical properties of the signal $\boldsymbol{X}$.

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 $\boldsymbol{X}$?



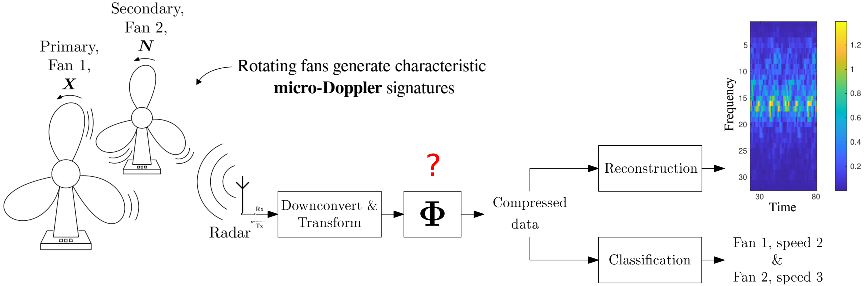
dstl
EPSRC
Engineering and Physical Sciences Research Council

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, 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.

Existing Optimisation Strategy

- Recent work utilises MI optimisation to design a linear projection that can trade-off signal recovery and classification accuracy for two sources (Coutts *et al.*, 2020).
- Goal: with $\boldsymbol{Y} = \boldsymbol{\Phi}(\boldsymbol{X} + \boldsymbol{N}) + \boldsymbol{W}$, design $\boldsymbol{\Phi}$ that maximises

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

- 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**.

Existing Optimisation Strategy

- Goal: with $\boldsymbol{Y} = \boldsymbol{\Phi}(\boldsymbol{X} + \boldsymbol{N}) + \boldsymbol{W}$, design $\boldsymbol{\Phi}$ that maximises

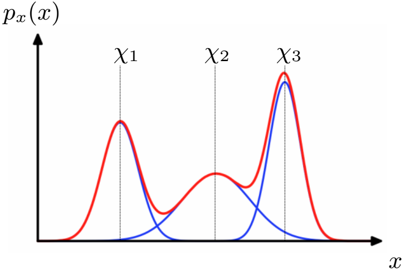
$$F(\boldsymbol{\Phi}, \boldsymbol{\beta}) = \beta_1 I(\boldsymbol{X}; \boldsymbol{Y}) + \beta_2 I(C; \boldsymbol{Y}) + \beta_3 I(\boldsymbol{N}; \boldsymbol{Y}) + \beta_4 I(G; \boldsymbol{Y})$$

- C and G represent classes of $\boldsymbol{X}$ and $\boldsymbol{N}$, respectively.
- Method: iterative gradient ascent.
  - Compute gradient of each information term w.r.t. $\boldsymbol{\Phi}$.
  - $\boldsymbol{\Phi} \leftarrow \boldsymbol{\Phi} + \delta \nabla_{\boldsymbol{\Phi}} F(\boldsymbol{\Phi}, \boldsymbol{\beta})$ for some step size $\delta > 0$; repeat.
- Model: $\boldsymbol{X}$ & $\boldsymbol{N}$ are Gaussian Mixture distributed; $\boldsymbol{W}$ Gaussian.

Existing Optimisation Strategy

- 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.,

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



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

Existing Optimisation Strategy

- Example scenario: we have a well-characterised primary source $\boldsymbol{X}$ measured in the presence of a fleeting secondary source $\boldsymbol{N}$.
- With *a priori* knowledge of the distributions, we can design $\boldsymbol{\Phi}$ such that we are able to accurately reconstruct/classify $\boldsymbol{X}$ and/or $\boldsymbol{N}$ given $\boldsymbol{Y}$.
- Designing $\boldsymbol{\Phi}$ to recognise both sources would be useful if the operational circumstances of the system were to change, with $\boldsymbol{N}$ becoming a signal of interest.
- Problem: what if the distribution of $\boldsymbol{N}$ is not constant?

Learning a Secondary Source From Compressive Measurements

- Learning the distribution of a Gaussian mixture (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 extend this approach to our chosen signal model:

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

Learning a Secondary Source From Compressive Measurements

- We assign GM distributions to $\boldsymbol{X}$ and $\boldsymbol{N}$ to match the optimisation framework.
- If we assume that the distribution of our primary source $\boldsymbol{X}$ is known *a priori*, can we learn the distribution of $\boldsymbol{N}$ from compressive measurements $\boldsymbol{Y}$?

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

- Here, we assume a known Gaussian distribution for the measurement noise $\boldsymbol{W}$.

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

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

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

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

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

Learning a Secondary Source From Compressive Measurements

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

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

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

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

- 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

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

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

where

$$p_{\boldsymbol{y}|\theta}(\boldsymbol{y}_i|\theta) = \sum_{k,d} \int p_{\boldsymbol{y},\boldsymbol{n},k,d|\theta}(\boldsymbol{y}_i, \boldsymbol{n}, k, d|\theta)\, d\boldsymbol{n}$$

and $\boldsymbol{n}$, k, and d are unobserved, 'latent' variables:

  - $\boldsymbol{n}$: instances of secondary source.
  - k: GM component $\boldsymbol{n}$ is drawn from.
  - d: GM component $\hat{\boldsymbol{w}}$ is drawn from (remembering $\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{n} + \hat{\boldsymbol{w}}$).

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 $(\boldsymbol{n}, k, d)$ given access to compressive measurements $\boldsymbol{y}_i$ and the previous system parameters $\theta^{(t)}$.

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 $(\boldsymbol{n}, k, d)$ given access to compressive measurements $\boldsymbol{y}_i$ and the previous system parameters $\theta^{(t)}$.
  2. Update the system parameters such that

$$\theta^{(t+1)} = \arg\max_{\theta} \sum_{i=1}^{N_s} \mathbb{E}\Big[\log p_{\boldsymbol{y},\boldsymbol{n},k,d|\theta}(\boldsymbol{y}_i, \boldsymbol{n}, k, d|\theta) \Big]$$

where the expectation is taken over the likelihood from step 1.

Learning a Secondary Source From Compressive Measurements

- For the maximisation in step 2, we are able to find 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).

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

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

- 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

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

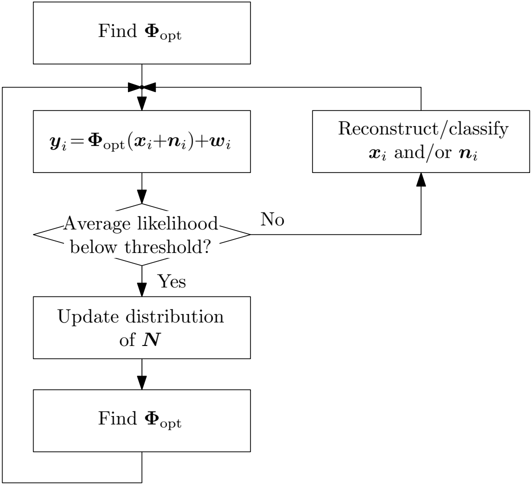
- Important considerations when estimating the distribution of $\boldsymbol{N}$:
  - Unique $\{\boldsymbol{\Phi}_i\}_{i=1}^{N_s}$ will improve estimation but add cost.
  - Increasing m (the number of rows in $\boldsymbol{\Phi}_i \in \mathbb{C}^{m \times n}$) will improve estimation.
  - Increasing the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

Adaptive Algorithm

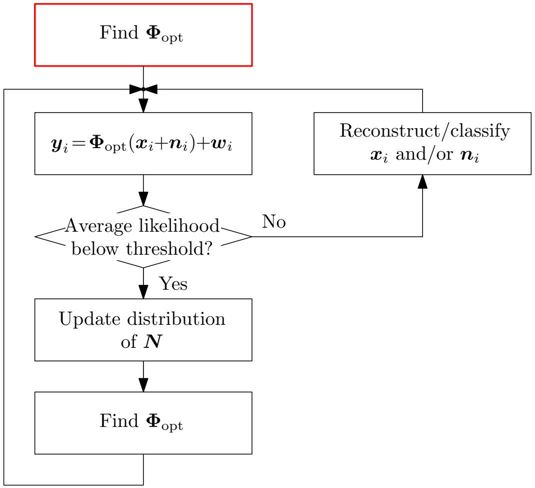
$$Y = \Phi_{\text{opt}}(X + N) + W$$

- We now have access to:
  - An algorithm that can identify the optimal projection matrix Φ_{opt} given accurate estimations of the source distributions.
  - Techniques to update the distribution of N if the current distribution is found to be innaccurate (via a likelihood test).
- We can now create an adaptive algorithm that updates Φ_{opt} to account for a changing secondary source distribution.

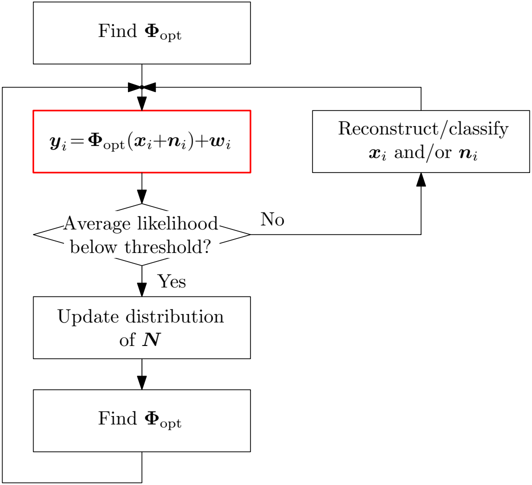
Adaptive Algorithm



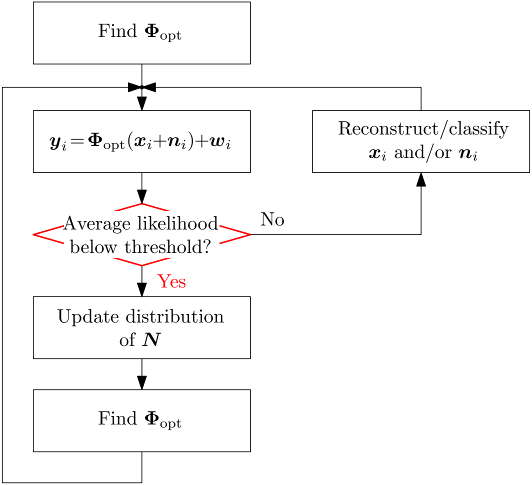
Adaptive Algorithm



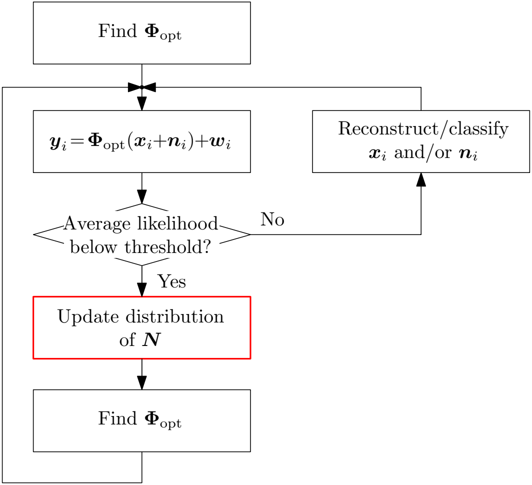
Adaptive Algorithm



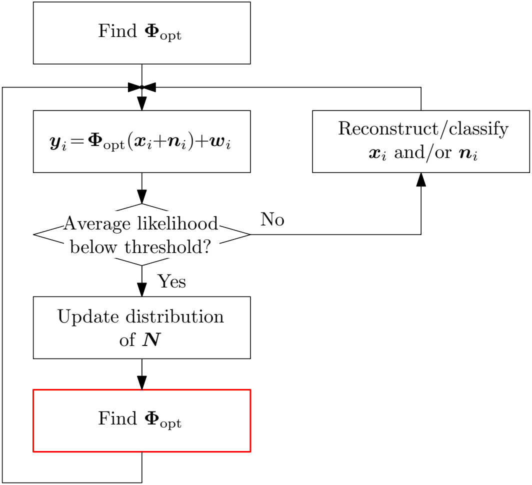
Adaptive Algorithm



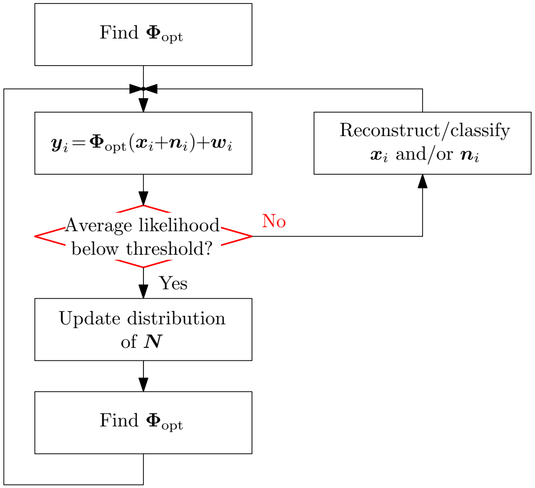
Adaptive Algorithm



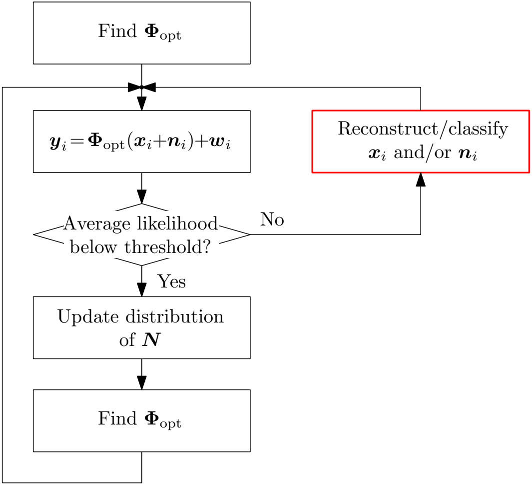
Adaptive Algorithm



Adaptive Algorithm



Adaptive Algorithm

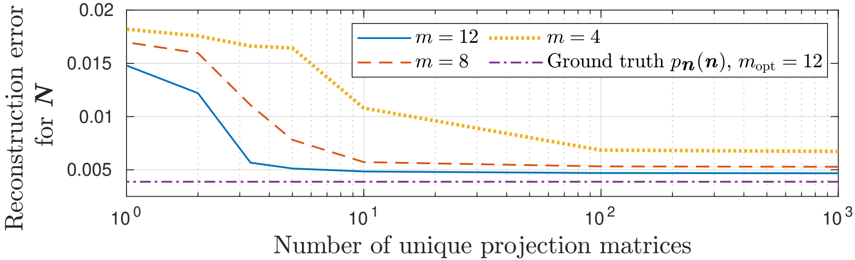


Results Using Synthetic Data

- Objective: estimate the distribution of $\boldsymbol{N}$ given $N_s = 10^3$ measurements $\{\boldsymbol{y}_i = \boldsymbol{\Phi}_i \boldsymbol{n}_i + \hat{\boldsymbol{w}}_i\}_{i=1}^{N_s}$:
  - Unique $\{\boldsymbol{\Phi}_i\}_{i=1}^{N_s}$ will improve estimation but add cost.
  - Increasing m ($\boldsymbol{\Phi}_i \in \mathbb{C}^{m \times n}$) will improve estimation but add cost.
  - Better estimation will result in lower reconstruction error for $\boldsymbol{N}$.

Results Using Synthetic Data

- Objective: estimate the distribution of $\boldsymbol{N}$ given $N_s = 10^3$ measurements $\{\boldsymbol{y}_i = \boldsymbol{\Phi}_i \boldsymbol{n}_i + \hat{\boldsymbol{w}}_i\}_{i=1}^{N_s}$:
  - Unique $\{\boldsymbol{\Phi}_i\}_{i=1}^{N_s}$ will improve estimation but add cost.
  - Increasing m ($\boldsymbol{\Phi}_i \in \mathbb{C}^{m \times n}$) will improve estimation but add cost.
  - Better estimation will result in lower reconstruction error for $\boldsymbol{N}$.

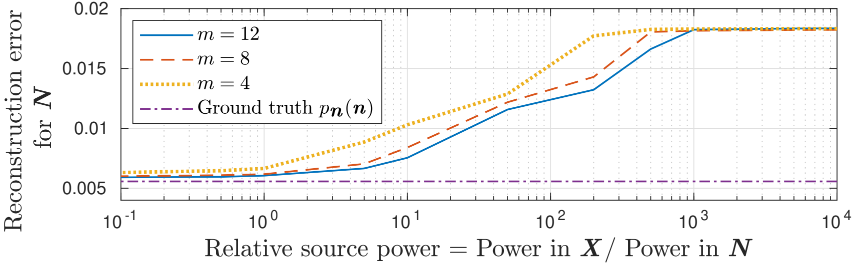


Results Using Synthetic Data

- Objective: estimate the distribution of $\boldsymbol{N}$ given $N_s = 10^3$ measurements $\{\boldsymbol{y}_i = \boldsymbol{\Phi}_i(\boldsymbol{x}_i + \boldsymbol{n}_i) + \boldsymbol{w}_i\}_{i=1}^{N_s}$:
  - Increasing the power of $\boldsymbol{X}$ relative to $\boldsymbol{N}$ will worsen estimation.
  - Increasing m ($\boldsymbol{\Phi}_i \in \mathbb{C}^{m \times n}$) will improve estimation but add cost.
  - Better estimation will result in lower reconstruction error for $\boldsymbol{N}$.

Results Using Synthetic Data

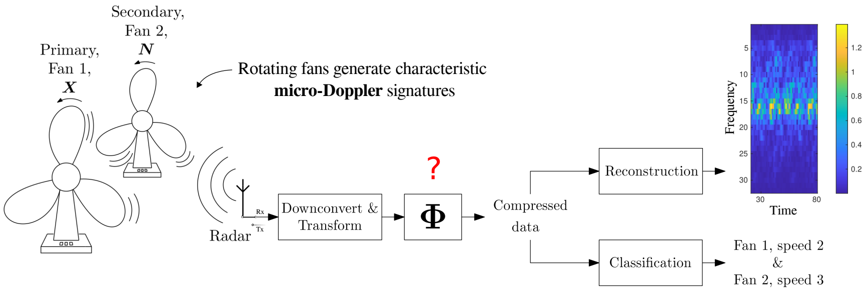
- Objective: estimate the distribution of $\boldsymbol{N}$ given $N_s = 10^3$ measurements $\{\boldsymbol{y}_i = \boldsymbol{\Phi}_i(\boldsymbol{x}_i + \boldsymbol{n}_i) + \boldsymbol{w}_i\}_{i=1}^{N_s}$:
  - Increasing the power of $\boldsymbol{X}$ relative to $\boldsymbol{N}$ will worsen estimation.
  - Increasing m ($\boldsymbol{\Phi}_i \in \mathbb{C}^{m \times n}$) will improve estimation but add cost.
  - Better estimation will result in lower reconstruction error for $\boldsymbol{N}$.



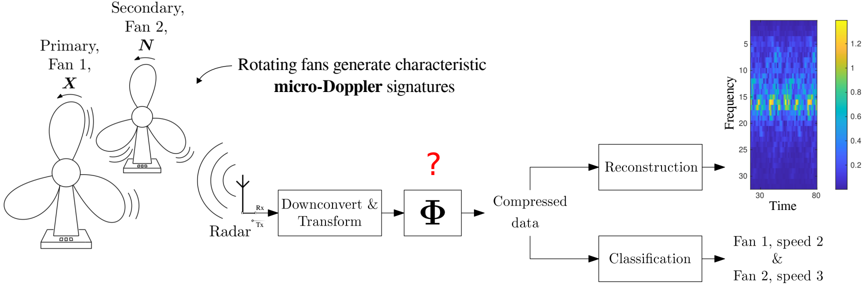
Results Using Real Radar Data

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

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



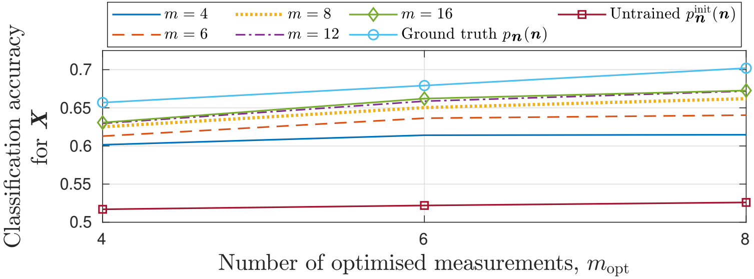
Results Using Real Radar Data



- Initial scenario: X well characterised, N absent (unknown).
- Simulation: N is **fleeting** & has variable class.
- Objective: Learn class distributions of N when source present then update optimal Φ_{opt} — e.g., to prioritise classification of X.

Results Using Real Radar Data

- Objective: Learn class distributions of $\boldsymbol{N}$ when source present then update optimal $\boldsymbol{\Phi}_{\text{opt}}$ — e.g., to prioritise classification of $\boldsymbol{X}$.
  - Increasing m improves our estimation of the distribution of $\boldsymbol{N}$.
  - With a good estimate, we can obtain a better $\boldsymbol{\Phi}_{\text{opt}} \in \mathbb{C}^{m_{\text{opt}} \times n}$.

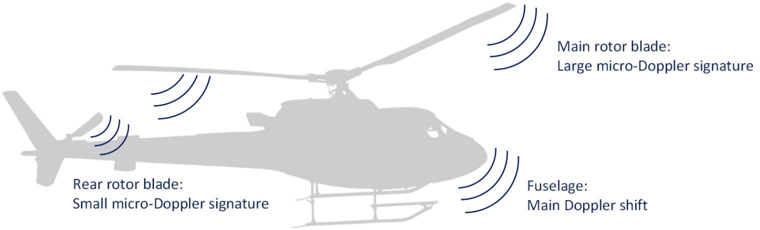


Conclusions

- Derived a methodology for the training of the GM distribution of a secondary input via compressive measurements.
- Increasing the number of compressive measurements can:
  - aid the characterisation of weak secondary sources;
  - reduce the number of unique projection matrices required.
- Well-estimated distributions yield designed projection matrices that are more able to control the input-output MI of a system.
- Framework could be extended to applications in which the operational parameters are liable to change such that a secondary source of information becomes more important.

Future Work

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



References

- M. Hellman and J. Raviv, "Probability of error, equivocation, and the Chernoff bound," *IEEE Trans. Inf. Theory*, vol. 16, no. 4, pp. 368–372, July 1970.
- 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.
- 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.