Adaptive Detection of Structured Signals in Low-Rank Interference

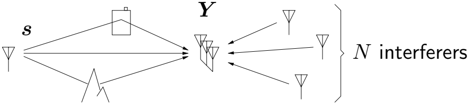
Phil Schniter and Evan Byrne

The Ohio State University

Supported in part by MIT Lincoln Labs

Asilomar 2018

Adaptive Detection of Structured Signals



Goal: Test for presence of temporal signal $\boldsymbol{s} \in \mathbb{C}^L$ using M antennas.

Challenges (typical):

- unknown steering vector $\boldsymbol{h} \in \mathbb{C}^M$ (e.g., multipath propagation)
- additive noise with unknown variance $\nu > 0$
- N additive interferers with unknown steering vectors (and unknown N)

Challenges (new):

- Signal $\boldsymbol{s}$ is known only in probability (i.e., $p(\boldsymbol{s})$ known)
  - Application: detect/synchronize using both pilots and unknown QAM symbols.
  - Traditionally, unknown symbols are ignored when synchronizing.[1]

[1] D. W. Bliss and P. A. Parker, "Temporal synchronization of MIMO wireless communication in the presence of interference," *IEEE Trans. Signal Process.*, 2010.

Binary Hypothesis Test

We consider the binary hypothesis test

$$\mathcal{H}_1 : \boldsymbol{Y} = \boldsymbol{h}\boldsymbol{s}^{\mathsf{H}} + \boldsymbol{B}\boldsymbol{\Phi}^{\mathsf{H}} + \boldsymbol{W} \in \mathbb{C}^{M \times L}$$
$$\mathcal{H}_0 : \boldsymbol{Y} = \boldsymbol{B}\boldsymbol{\Phi}^{\mathsf{H}} + \boldsymbol{W} \in \mathbb{C}^{M \times L}$$

Assumptions:

- $\boldsymbol{s} \sim p(\boldsymbol{s})$
- unknown steering vector $\boldsymbol{h} \in \mathbb{C}^M$
- unknown white Gaussian noise $\boldsymbol{W}$ with unknown variance $\nu > 0$
- unknown low-rank interference $\boldsymbol{B} \in \mathbb{C}^{M \times N}$, $\boldsymbol{\Phi} \in \mathbb{C}^{L \times N}$, $N < M$
- unknown interference rank N

Prior Work on Known-$\boldsymbol{s}$ Case

Many prior works have considered the case of known $\boldsymbol{s}$. For example...

- Kelly[2] modeled the noise-plus-interference $\boldsymbol{N} \triangleq \boldsymbol{B}\boldsymbol{\Phi}^{\mathsf{H}} + \boldsymbol{W}$ as

  $$\mathrm{vec}(\boldsymbol{N}) \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_L \otimes \boldsymbol{\Sigma}) \text{ with unknown spatial covariance } \boldsymbol{\Sigma} > 0$$

  and formulated the generalized likelihood ratio test (GLRT), i.e.,

  $$\frac{\max_{\boldsymbol{h},\boldsymbol{\Sigma}>0} p(\boldsymbol{Y}|\mathcal{H}_1; \boldsymbol{h}, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\Sigma}>0} p(\boldsymbol{Y}|\mathcal{H}_0; \boldsymbol{\Sigma})} \gtrless \eta.$$

  Using $\boldsymbol{P}_{\boldsymbol{s}}^{\perp}$ to denote orthogonal projection away from $\boldsymbol{s}$, the GLRT reduces to

  $$\frac{\prod_{m=1}^{M} \lambda_{0,m}}{\prod_{m=1}^{M} \lambda_{1,m}} \gtrless \eta \text{ where } \begin{cases} \{\lambda_{0,m}\} = \mathsf{evals}\big(\frac{1}{L}\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}\big) \\ \{\lambda_{1,m}\} = \mathsf{evals}\big(\frac{1}{L}\boldsymbol{Y}\boldsymbol{P}_{\boldsymbol{s}}^{\perp}\boldsymbol{Y}^{\mathsf{H}}\big) \end{cases}$$

[2] E. Kelly, "An adaptive detection algorithm," *IEEE Trans. Aerosp. Electron. Syst.*, 1986.

Prior Work on Known-s Case (cont.)

- Kang, Monga, and Rangaswamy[3] (KMR) modeled the noise-plus-interference $\boldsymbol{N} = \boldsymbol{B}\boldsymbol{\Phi}^{\mathsf{H}} + \boldsymbol{W}$ as

  $\mathrm{vec}(\boldsymbol{N}) \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_L \otimes \boldsymbol{\Sigma})$ with unknown $\boldsymbol{\Sigma} \in \mathcal{S}_N$

  $\mathcal{S}_N \triangleq \{\boldsymbol{R} + \nu \boldsymbol{I}_M : \mathrm{rank}(\boldsymbol{R}) = N,\ \boldsymbol{R} \geq 0,\ \nu > 0\}$ (note N assumed known)

  and formulated the GLRT, i.e.,

$$\frac{\max_{\boldsymbol{h}, \boldsymbol{\Sigma} \in \mathcal{S}_N} p(\boldsymbol{Y} | \mathcal{H}_1; \boldsymbol{h}, \boldsymbol{\Sigma})}{\max_{\boldsymbol{\Sigma} \in \mathcal{S}_N} p(\boldsymbol{Y} | \mathcal{H}_0; \boldsymbol{\Sigma})} \gtrless \eta.$$

  This GLRT reduces to

$$\frac{\prod_{m=1}^{M} \widehat{\lambda}_{0,m}}{\prod_{m=1}^{M} \widehat{\lambda}_{1,m}} \gtrless \eta, \text{ where } \{\widehat{\lambda}_{i,m}\}_{m=1}^{M} \text{ are "smoothed".}$$

  That is, $\widehat{\lambda}_{i,m} = \begin{cases} \lambda_{i,m} & m \leq N \\ \widehat{\nu}_i \triangleq \frac{1}{M-N} \sum_{m=N+1}^{M} \lambda_{i,m} & m > N \end{cases}$ for decreasing $\{\lambda_{i,m}\}$.

[3] B. Kang, V. Monga, and M. Rangaswamy, "Rank-constrained maximum likelihood estimation of structured covariance matrices," *IEEE Trans. Aerosp. Electron. Syst.*, 2014.

Prior Work on Known-$\boldsymbol{s}$ Case (cont.)

- McWhorter[4] treated temporal interference $\boldsymbol{\Phi}$ as *deterministic* (not as AWGN) in his GLRT formulation:

$$\frac{\max_{\boldsymbol{h},\boldsymbol{B},\boldsymbol{\Phi},\nu>0} p(\boldsymbol{Y}|\mathcal{H}_1;\boldsymbol{h},\boldsymbol{B},\boldsymbol{\Phi},\nu)}{\max_{\boldsymbol{B},\boldsymbol{\Phi},\nu>0} p(\boldsymbol{Y}|\mathcal{H}_0;\boldsymbol{B},\boldsymbol{\Phi},\nu)} \gtrless \eta. \quad \text{(note } N \text{ assumed known)}$$

  This GLRT reduces to

$$\frac{\widehat{\nu}_0}{\widehat{\nu}_1} = \frac{\frac{1}{M}\sum_{m=N+1}^{M}\lambda_{0,m}}{\frac{1}{M}\sum_{m=N+1}^{M}\lambda_{1,m}} \gtrless \eta',$$

  where the eigenvalues $\{\lambda_{i,m}\}$ are the same as defined earlier.

- Essentially, McWhorter uses *interference cancellation*, whereas Kelly and KMR use *interference nulling*.

[4] L. T. McWhorter, "A high resolution detector in multi-path environments," in *Proc. Workshop ASAP* (Lexington, MA), 2004.

Probabilistic $\boldsymbol{s}$ and Gaussian Interference

- We now return to the case where $\boldsymbol{s} \sim p(\boldsymbol{s})$ with known $p(\cdot)$.
- Treating the interference as Gaussian (like KMR) gives the GLRT numerator

$$\max_{\boldsymbol{h},\boldsymbol{\Sigma}\in\mathcal{S}_N} p(\boldsymbol{Y}|\mathcal{H}_1;\widehat{\boldsymbol{h}},\boldsymbol{\Sigma}) = \max_{\boldsymbol{h},\boldsymbol{\Sigma}\in\mathcal{S}_N} \int \frac{\exp(-\operatorname{tr}\{(\boldsymbol{Y}-\boldsymbol{h}\boldsymbol{s}^{\mathsf{H}})^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}(\boldsymbol{Y}-\boldsymbol{h}\boldsymbol{s}^{\mathsf{H}})\})}{\pi^{ML}|\boldsymbol{\Sigma}|^{L}}\, p(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}$$

  which is, in general, intractable.
- Thus we propose to iteratively maximize this likelihood via EM:

$$\left(\widehat{\boldsymbol{h}}^{(t+1)}, \widehat{\boldsymbol{\Sigma}}_1^{(t+1)}\right) = \operatorname*{arg\,max}_{\boldsymbol{h}\in\mathbb{C}^M,\boldsymbol{\Sigma}\in\mathcal{S}_N} \mathbb{E}\left\{ \ln p(\boldsymbol{Y},\boldsymbol{s}|\mathcal{H}_1;\boldsymbol{h},\boldsymbol{\Sigma}) \,\middle|\, \boldsymbol{Y}; \widehat{\boldsymbol{h}}^{(t)}, \widehat{\boldsymbol{\Sigma}}_1^{(t)} \right\}.$$

- After t EM iterations, the GLRT becomes $\frac{\prod_{m=1}^{M}\widehat{\lambda}_{0,m}}{\prod_{m=1}^{M}\widehat{\lambda}_{1,m}^{(t)}} \gtrless \eta$, where $\{\widehat{\lambda}_{1,m}^{(t)}\}$ are the smoothed evals of $\widehat{\boldsymbol{\Sigma}}_1^{(t)}$ and $\{\widehat{\lambda}_{0,m}\}$ are the smoothed evals of $\frac{1}{L}\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}}$.

EM Details for Gaussian Interference (cont.)

We show[5] that

$$\widehat{\boldsymbol{h}}^{(t+1)} = \boldsymbol{Y}\widehat{\boldsymbol{s}}^{(t)}/E^{(t)} \quad \text{for} \quad \begin{cases} \widehat{\boldsymbol{s}}^{(t)} \triangleq \mathbb{E}\left\{\boldsymbol{s} \middle| \boldsymbol{Y}; \widehat{\boldsymbol{h}}^{(t)}, \widehat{\boldsymbol{\Sigma}}_1^{(t)}\right\} \\ E^{(t)} \triangleq \mathbb{E}\left\{\|\boldsymbol{s}\|^2 \middle| \boldsymbol{Y}; \widehat{\boldsymbol{h}}^{(t)}, \widehat{\boldsymbol{\Sigma}}_1^{(t)}\right\} \end{cases}$$

and that minimizing $\boldsymbol{\Sigma} \in \mathcal{S}_N$ is equivalent to maximizing

$$\frac{\exp(-\operatorname{tr}\{\boldsymbol{Y}\widetilde{\boldsymbol{P}}^{\perp}_{\widehat{\boldsymbol{s}}^{(t)}}\boldsymbol{Y}^{\mathsf{H}}\boldsymbol{\Sigma}^{-1}\})}{\pi^{ML}|\boldsymbol{\Sigma}|^{L}} \quad \text{with} \quad \widetilde{\boldsymbol{P}}^{\perp}_{\widehat{\boldsymbol{s}}^{(t)}} \triangleq \boldsymbol{I}_L - \frac{\widehat{\boldsymbol{s}}^{(t)}\widehat{\boldsymbol{s}}^{(t)\mathsf{H}}}{E^{(t)}}$$

which (via Anderson'63) leads to the solution

$$\widehat{\boldsymbol{\Sigma}}_1^{(t+1)} = \boldsymbol{V}_1^{(t+1)}\operatorname{Diag}(\widehat{\lambda}_{1,1}^{(t+1)}, \dots, \widehat{\lambda}_{1,M}^{(t+1)})\boldsymbol{V}_1^{(t+1)\mathsf{H}}$$

$$\widehat{\lambda}_{1,m}^{(t+1)} = \begin{cases} \lambda_{1,m}^{(t+1)} & m = 1, \dots, N \\ \widehat{\nu}_1^{(t+1)} \triangleq \frac{1}{M-N}\sum_{m=N+1}^{M}\lambda_{1,m}^{(t+1)} & m = N+1, \dots, M \end{cases}$$

where $\{\lambda_{1,m}^{(t+1)}\}$ are the decreasing-ordered eigenvalues of $\boldsymbol{Y}\widetilde{\boldsymbol{P}}^{\perp}_{\widehat{\boldsymbol{s}}^{(t)}}\boldsymbol{Y}^{\mathsf{H}}$.

[5] E. Byrne and P. Schniter, "Adaptive Detection of Structured Signals in Low-Rank Interference," *arXiv:1808.05650*.

EM Details for Gaussian Interference (cont.)

To compute $\widehat{\boldsymbol{s}}^{(t)}$ and $\mathbb{E}^{(t)}$, we focus on independent priors $p(\boldsymbol{s}) = \prod_{l=1}^{L} p_l(s_l)$. Then...

- Under $\boldsymbol{h} = \widehat{\boldsymbol{h}}^{(t)}$ and $\boldsymbol{\Sigma} = \widehat{\boldsymbol{\Sigma}}_1^{(t)}$, the model becomes

$$\boldsymbol{y}_l = \widehat{\boldsymbol{h}}^{(t)} s_l^* + \mathcal{CN}(\boldsymbol{0}, \widehat{\boldsymbol{\Sigma}}_1^{(t)}) \ \forall l.$$

- The *whitened matched filter* gives a sufficient statistic for estimating s_l:

$$\widetilde{r}_l^{(t)} \triangleq \widehat{\boldsymbol{h}}^{(t)\mathsf{H}} (\widehat{\boldsymbol{\Sigma}}_1^{(t)})^{-1} \boldsymbol{y}_l = \xi^{(t)} s_l + \mathcal{CN}(0, \xi^{(t)}) \ \text{ for } \ \xi^{(t)} \triangleq \widehat{\boldsymbol{h}}^{(t)\mathsf{H}} (\widehat{\boldsymbol{\Sigma}}_1^{(t)})^{-1} \widehat{\boldsymbol{h}}^{(t)}.$$

- The WMF outputs can be scaled to give an *unbiased estimate*

$$r_l^{(t)} \triangleq \left[\widetilde{r}_l^{(t)} / \xi^{(t)}\right]^* = s_l + \mathcal{CN}(0, 1/\xi^{(t)}).$$

- Computation of $\widehat{s}_l^{(t)} = \mathbb{E}\{s_l | r_l^{(t)}\}$ is *scalar MMSE denoising*: easy to do. Likewise, $E^{(t)} = \|\widehat{\boldsymbol{s}}^{(t)}\|^2 + \sum_l \mathrm{Cov}\{s_l | r_l^{(t)}\}$.

EM Details for Gaussian Interference (cont.)

Algorithm 1 EM update for the interference-nulling GLRT

Require: Data $\boldsymbol{Y} \in \mathbb{C}^{M \times L}$, signal prior $p(\boldsymbol{s}) = \prod_{l=1}^{L} p_l(s_l)$.

Initialize $\widehat{\boldsymbol{s}} \in \mathbb{C}^L$ and $E > 0$.

repeat

$\widehat{\boldsymbol{h}} \leftarrow \frac{1}{E} \boldsymbol{Y} \widehat{\boldsymbol{s}}$ — steering-vector estimate

$\widehat{\boldsymbol{\Sigma}}_1 \leftarrow \frac{1}{L} \boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}} - \frac{E}{L} \widehat{\boldsymbol{h}}\widehat{\boldsymbol{h}}^{\mathsf{H}}$ — estimate of interference+noise covariance $\boldsymbol{\Sigma}$

$\widehat{N} \leftarrow \text{rank\_estimate}(\widehat{\boldsymbol{\Sigma}}_1)$

$\{\overline{\boldsymbol{V}}_1, \overline{\boldsymbol{\Lambda}}_1\} \leftarrow \text{principal\_eigs}(\widehat{\boldsymbol{\Sigma}}_1, \widehat{N})$

$\widehat{\nu}_1 \leftarrow \frac{1}{M - \widehat{N}} \left(\text{tr}(\widehat{\boldsymbol{\Sigma}}_1) - \text{tr}\{\overline{\boldsymbol{\Lambda}}_1\} \right)$ — estimate of noise variance

$\boldsymbol{g} \leftarrow \frac{1}{\widehat{\nu}_1} \widehat{\boldsymbol{h}} + \overline{\boldsymbol{V}}_1 \left(\overline{\boldsymbol{\Lambda}}_1^{-1} - \frac{1}{\widehat{\nu}_1} \boldsymbol{I}_{\widehat{N}} \right) \overline{\boldsymbol{V}}_1^{\mathsf{H}} \widehat{\boldsymbol{h}}$ — $\widehat{\boldsymbol{\Sigma}}_1^{-1} \widehat{\boldsymbol{h}}$

$\xi \leftarrow \widehat{\boldsymbol{h}}^{\mathsf{H}} \boldsymbol{g}$ — precision of error on $\boldsymbol{r}$

$\boldsymbol{r} \leftarrow \boldsymbol{Y}^{\mathsf{H}} \boldsymbol{g} / \xi$ where $\boldsymbol{r} \sim \mathcal{CN}(\boldsymbol{s}, \boldsymbol{I}/\xi)$ — AWGN-corrupted pseudo-measurement of $\boldsymbol{s}$

$\widehat{s}_l \leftarrow \mathbb{E}\{s_l | r_l; \xi\} \; \forall l = 1, \dots, L$

$E \leftarrow \sum_{l=1}^{L} \mathbb{E}\{|s_l|^2 | r_l; \xi\}$

until Terminated

Estimation of Interference Rank N

Although the previous approaches assumed known interference rank N, standard model-order selection methods[6] can be used to estimate N:

$$\widehat{N} = \underset{N \in \{0,\dots,N_{\max}\}}{\arg\max} \ln p(\boldsymbol{Y}|\mathcal{H}_1, \widehat{\boldsymbol{\Theta}}_N) - J(N)$$

where

$$\widehat{\boldsymbol{\Theta}}_N = \text{ML parameter estimate under rank hypothesis } N\text{, i.e.,}$$

$$\boldsymbol{\Theta}_N = \begin{cases} \{\boldsymbol{h}, \boldsymbol{\Sigma} \in \mathcal{S}_N\} & \text{KMR} \\ \{\boldsymbol{h}, \boldsymbol{B} \in \mathbb{C}^{M \times N}, \boldsymbol{\Phi} \in \mathbb{C}^{Q \times N}, \nu\} & \text{McWhorter} \end{cases}$$

$$J(N) = \text{penalty with respect to the degrees-of-freedom ``}D_{\mathsf{oF}}(N)\text{,'' e.g.,}$$

$$J(N) = \begin{cases} D_{\mathsf{oF}}(N) & \text{Akaike's Information Criterion (AIC)} \\ \frac{2MQ}{2MQ - D_{\mathsf{oF}}(N) - 1} D_{\mathsf{oF}}(N) & \text{corrected AIC (AICC)} \\ 0.5 D_{\mathsf{oF}}(N) \ln(2MQ) & \text{Bayesian Information Criterion (BIC)} \\ G\, D_{\mathsf{oF}}(N) & \text{Generalized Info Criterion (GIC)} \end{cases}$$

[6] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust. Speech & Signal Process.*, 1986.

EM Initialization for Gaussian Interference

- The EM algorithm benefits from a good initialization of $\widehat{\boldsymbol{s}}$.
- Let's focus on $\boldsymbol{s}^{\mathsf{H}} = \left[\boldsymbol{s}_{\mathsf{t}}^{\mathsf{H}} \;\; \boldsymbol{s}_{\mathsf{d}}^{\mathsf{H}}\right]$ with known training $\boldsymbol{s}_{\mathsf{t}} \in \mathbb{C}^{Q}$.
  Our goal is then to MMSE-estimate $\widehat{\boldsymbol{s}}_{\mathsf{d}} = \mathbb{E}\{\boldsymbol{s}_{\mathsf{d}} | \boldsymbol{Y}_{\mathsf{d}}; \widehat{\boldsymbol{h}}, \widehat{\boldsymbol{\Sigma}}\}$ for some $(\widehat{\boldsymbol{h}}, \widehat{\boldsymbol{\Sigma}})$.
- One option is training-based ML estimation. With full rank $N = M$, we'd get
$$\widehat{\boldsymbol{h}}_{\mathsf{t}} \triangleq \boldsymbol{Y}_{\mathsf{t}} \boldsymbol{s}_{\mathsf{t}} / \|\boldsymbol{s}_{\mathsf{t}}\|^2 \quad \text{and} \quad \widehat{\boldsymbol{\Sigma}}_{\mathsf{t}} \triangleq \boldsymbol{Y}_{\mathsf{t}} \boldsymbol{P}_{\boldsymbol{s}_{\mathsf{t}}}^{\perp} \boldsymbol{Y}_{\mathsf{t}}^{\mathsf{H}} / Q.$$
- When $N < M$, could try to estimate N via model-order selection, but this leads to problems in estimating the bias of the WMF outputs.
- We instead suggest to use "diagonal loading"
$$\widehat{\boldsymbol{\Sigma}}_{\mathsf{t}}^{(\alpha)} = (1-\alpha)\widehat{\boldsymbol{\Sigma}}_{\mathsf{t}} + \alpha \frac{\operatorname{tr}\{\widehat{\boldsymbol{\Sigma}}_{\mathsf{t}}\}}{M} \boldsymbol{I}, \quad \alpha \in (0,1],$$
  where α is chosen via leave-one-out cross-validation (LOOCV).[7]

[7] J. Tong, P. J. Schreier, Q. Guo, S. Tong, J. Xi, and Y. Yu, "Shrinkage of covariance matrices for linear signal estimation using cross-validation," *IEEE Trans. Signal Process.*, 2016.

Relation to Forsythe's Iterative Method

- When $p(\boldsymbol{s}) = \prod_{l=1}^{L} p_l(s_l)$ and rank $N = M$, our EM algorithm becomes

$$\begin{aligned} \boldsymbol{w} &\leftarrow (\boldsymbol{Y}\boldsymbol{Y}^{\mathsf{H}})^{-1}\boldsymbol{Y}\widehat{\boldsymbol{s}}\frac{\|\widehat{\boldsymbol{s}}\|^2}{\|\boldsymbol{P}_{\boldsymbol{Y}^{\mathsf{H}}}\widehat{\boldsymbol{s}}\|^2} \\ \boldsymbol{r} &\leftarrow \boldsymbol{Y}^{\mathsf{H}}\boldsymbol{w} \\ \widehat{\boldsymbol{s}} &\leftarrow \mathbb{E}\{\boldsymbol{s}|\boldsymbol{r}\} \ \text{ where } \ \boldsymbol{r} = \boldsymbol{s} + \mathcal{CN}(\boldsymbol{0}, \xi^{-1}\boldsymbol{I}) \end{aligned}$$

  where $\boldsymbol{w}$ plays the role of a beamformer.

- The above becomes equivalent to Forsythe's Iterative ML scheme[8] if our MMSE signal estimate is replaced with the "hard" ML estimate

$$\widehat{\boldsymbol{s}}_{\mathsf{ML}} \leftarrow \arg\min_{\boldsymbol{s}\in\mathcal{A}^L} \|\boldsymbol{s} - \boldsymbol{r}\|^2.$$

  Thus our EM scheme is the "soft" and low-rank $(N < M)$ counterpart of Forsythe's scheme.

---

[8] K. W. Forsythe, "Utilizing waveform features for adaptive beamforming and direction finding with narrowband signals," *Lincoln Lab. J.*, 1997.

Probabilistic $\boldsymbol{s}$ and Deterministic Interference

- Now let's treat $\boldsymbol{\Phi}$ as deterministic interference (like McWhorter). In this case, the GLRT numerator becomes

$$\max_{\boldsymbol{\Theta}} p(\boldsymbol{Y}|\mathcal{H}_1;\boldsymbol{\Theta}) = \max_{\boldsymbol{\Theta}} \int \frac{\exp(-\|\boldsymbol{Y}-\boldsymbol{B}\boldsymbol{\Phi}^{\mathsf{H}}-\boldsymbol{h}\boldsymbol{s}^{\mathsf{H}}\|_F^2/\nu)}{(\pi\nu)^{ML}} p(\boldsymbol{s})\,\mathrm{d}\boldsymbol{s}$$

$$\text{for} \quad \boldsymbol{\Theta} \triangleq \{\boldsymbol{h}, \boldsymbol{B}, \boldsymbol{\Phi}, \nu\}$$

  which, is in general, intractable.

- Again we propose to iteratively maximize via EM:

$$\widehat{\boldsymbol{\Theta}}^{(t+1)} = \arg\max_{\boldsymbol{\Theta}} \mathbb{E}\left\{ \ln p(\boldsymbol{Y}, \boldsymbol{s}|\mathcal{H}_1;\boldsymbol{\Theta}) \,\middle|\, \boldsymbol{Y}; \widehat{\boldsymbol{\Theta}}^{(t)} \right\}$$

- After t EM iterations, the resulting GLRT becomes

$$\frac{\widehat{\nu}_0}{\widehat{\nu}_1^{(t)}} = \frac{\frac{1}{M}\sum_{m=N+1}^{M} \lambda_{0,m}}{\frac{1}{M}\sum_{m=N+1}^{M} \lambda_{1,m}^{(t)}} \gtrless \eta',$$

EM Details for Deterministic Interference

- Similar to before, $\{\lambda_{1,m}^{(t+1)}\}$ are the eigenvalues of $\boldsymbol{Y}\widetilde{\boldsymbol{P}}_{\widehat{\boldsymbol{s}}^{(t)}}^{\perp}\boldsymbol{Y}^{\mathsf{H}}/L$ with

$$\begin{aligned}\widetilde{\boldsymbol{P}}_{\widehat{\boldsymbol{s}}^{(t)}}^{\perp} &= \boldsymbol{I}_L - \frac{\widehat{\boldsymbol{s}}^{(t)}\widehat{\boldsymbol{s}}^{(t)\mathsf{H}}}{E^{(t)}}\\ \widehat{\boldsymbol{s}}^{(t)} &= \mathbb{E}\left\{\boldsymbol{s}\middle|\boldsymbol{Y};\widehat{\boldsymbol{\Theta}}^{(t)}\right\}\\ E^{(t)} &= \mathbb{E}\left\{\|\boldsymbol{s}\|^2\middle|\boldsymbol{Y};\widehat{\boldsymbol{\Theta}}^{(t)}\right\}\end{aligned}$$

- Can initialize $\widehat{\boldsymbol{s}}$ as before, using diagonal loading of $\widehat{\boldsymbol{\Sigma}}_{\mathsf{t}}$ and LOOCV.

- Can estimate rank N as before, but now with a different $D_{\mathsf{oF}}(N)$.

EM Details for Deterministic Interference

Algorithm 2 EM update for the interference-canceling GLRT

Require: Data $\boldsymbol{Y} \in \mathbb{C}^{M \times L}$, signal prior $p(\boldsymbol{s}) = \prod_{l=1}^{L} p(s_l)$.

Initialize $\widehat{\boldsymbol{s}} \in \mathbb{C}^L$ and $E > 0$.

repeat

- $\zeta \leftarrow \sqrt{1 - \|\widehat{\boldsymbol{s}}\|^2 / E}$ — softness factor; $\zeta = 0$ for hard estimate $\widehat{\boldsymbol{s}}$
- $\boldsymbol{g} \leftarrow \boldsymbol{Y}\widehat{\boldsymbol{s}} / \|\widehat{\boldsymbol{s}}\|^2$ — steering-vector estimate before IC
- $\overline{\boldsymbol{Y}} \leftarrow \boldsymbol{Y} + (\zeta - 1)\boldsymbol{g}\widehat{\boldsymbol{s}}^{\mathsf{H}}$ — estimate of noise+interference samples
- $\widehat{N} \leftarrow \text{rank\_estimate}(\overline{\boldsymbol{Y}})$
- $\{\overline{\boldsymbol{V}}, \overline{\boldsymbol{D}}_1, \overline{\boldsymbol{U}}^{\mathsf{H}}\} \leftarrow \text{principal\_svd}(\overline{\boldsymbol{Y}}, \widehat{N})$
- $\widehat{\nu}_1 \leftarrow \frac{1}{ML}\left(\|\overline{\boldsymbol{Y}}\|_F^2 - \operatorname{tr}\{\overline{\boldsymbol{D}}_1^2\}\right)$ — estimate of noise variance
- $\widehat{\boldsymbol{h}} \leftarrow \frac{1}{E}\left(\|\widehat{\boldsymbol{s}}\|^2 \boldsymbol{g} - \frac{1}{\zeta}\overline{\boldsymbol{V}}\,\overline{\boldsymbol{D}}_1\,\overline{\boldsymbol{U}}^{\mathsf{H}}\widehat{\boldsymbol{s}}\right)$ — steering-vector estimate after IC
- $\xi \leftarrow \frac{\|\widehat{\boldsymbol{h}}\|^2}{\widehat{\nu}}$ — precision of error on $\boldsymbol{r}$
- $\boldsymbol{r} \leftarrow \frac{1}{\|\widehat{\boldsymbol{h}}\|^2}\left(\overline{\boldsymbol{Y}}^{\mathsf{H}}\widehat{\boldsymbol{h}} - \overline{\boldsymbol{U}}\,\overline{\boldsymbol{D}}_1\overline{\boldsymbol{V}}^{\mathsf{H}}\widehat{\boldsymbol{h}}\right) + \frac{1}{1+\zeta}\widehat{\boldsymbol{s}}$ — AWGN-corrupted pseudo-measurement of $\boldsymbol{s}$
  where $\boldsymbol{r} \sim \mathcal{CN}(\boldsymbol{s}, \frac{1}{\xi}\boldsymbol{I})$
- $\widehat{s}_l \leftarrow \mathbb{E}\{s_l | r_l; \xi\}\ \forall l = 1, \ldots, L$
- $E \leftarrow \sum_{l=1}^{L} \mathbb{E}\{|s_l|^2 | r_l; \xi\}$

until Terminated

Setup for Numerical Experiments

Model $|\mathcal{H}_1$:

$$\begin{aligned} \boldsymbol{Y} &= \boldsymbol{h}\boldsymbol{s}^{\mathsf{H}} + \boldsymbol{B}\boldsymbol{\Phi}^{\mathsf{H}} + \boldsymbol{W} \in \mathbb{C}^{M\times L} \\ \boldsymbol{s}^{\mathsf{H}} &= [\boldsymbol{s}_{\mathsf{t}}^{\mathsf{H}}, \boldsymbol{s}_{\mathsf{d}}^{\mathsf{H}}], \quad \mathrm{rank}(\boldsymbol{B}\boldsymbol{\Phi}^{\mathsf{H}}) = N \end{aligned}$$

Dimensions:

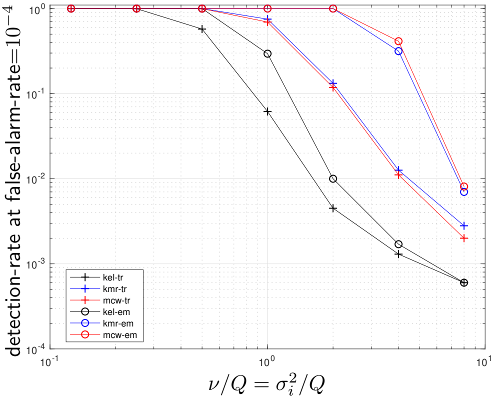
M	$= 64$	antennas
Q	$= 32$	training symbols in $\boldsymbol{s}_{\mathsf{t}}$ (QPSK)
L	$= 1024$	total symbols in $\boldsymbol{s}$ (QPSK)
N	$= 5$	interferers

Monte-Carlo:

$\boldsymbol{s}$:	i.i.d. QPSK	$\mathbb{E}\{\lvert s_l\rvert^2\} = 1$
$\boldsymbol{h}$:	random on 2D-UPA manifold	$\mathbb{E}\{\lvert h_m\rvert^2\} = 1$
$\boldsymbol{B}$:	2D-UPA sidelobe peaks	$\mathbb{E}\{\lvert[\boldsymbol{B}\boldsymbol{\Phi}]_{ml}\rvert^2\} = \sigma_i^2$
$\boldsymbol{W}$:	AWGN	$\mathbb{E}\{\lvert w_{ml}\rvert^2\} = \sigma_w^2 \propto Q$

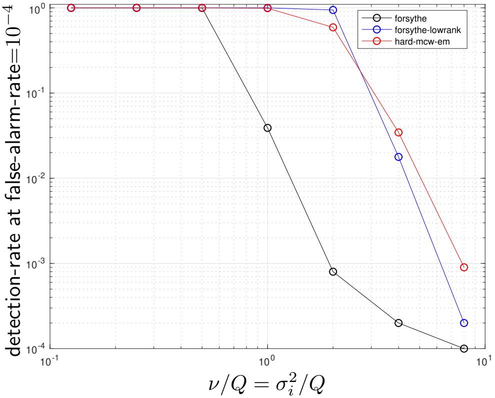
Performance: $\Pr\{\text{detection}\}$ under $\Pr\{\text{false-alarm}\} = 10^{-4}$

Pr(detection) versus SNR — EM & Training-only



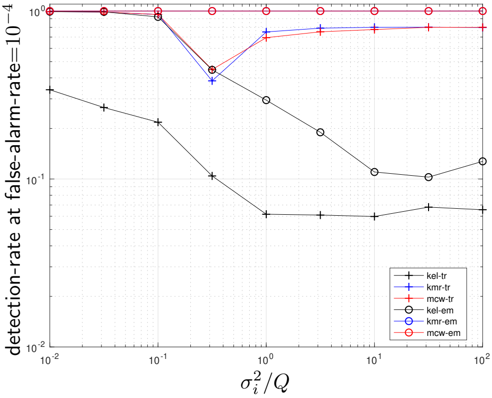
- Note: noise variance = interference variance
- Ranking:
  1. EM-based low-rank: oo
  2. training-only low-rank: ++
  3. EM-based full-rank: o
  4. training-only full-rank: +

Pr(detection) versus SNR — Iterative Hard Estimation



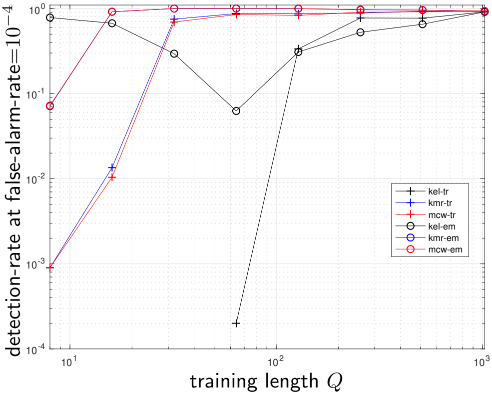
- Again: noise variance = interference variance
- Ranking:
  1. hard low-rank: oo
  2. hard full-rank: o
- These "hard" detectors are outperformed by their "soft" counterparts in the previous figure.

Pr(detection) versus SIR at fixed SNR



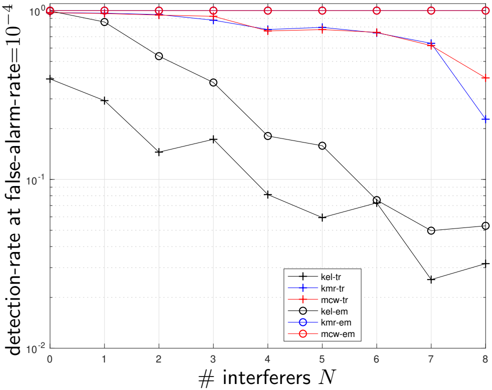
- Now noise variance fixed at $\nu = Q$
- Ranking:
  1. EM-based low-rank: oo
  2. training-only low-rank: ++
  3. EM-based full-rank: o
  4. training-only full-rank: +
- Dip in ++ results from mis-estimating N.

Pr(detection) versus training length Q for fixed $L = 1024$



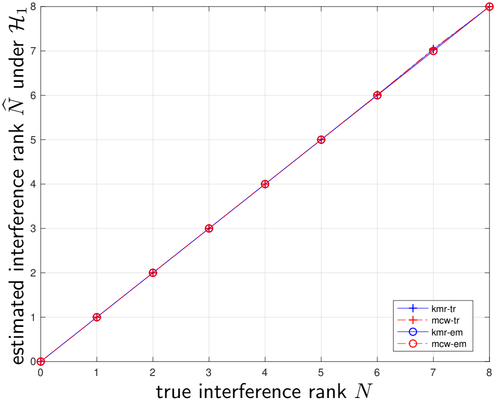
- Now variances fixed at $\nu = \sigma_i^2 = Q$
- Ranking for $Q > M/2$:
  1. EM-based low-rank: oo
  2. training-only low-rank: ++
  3. EM-based full-rank: o
  4. training-only full-rank: +

Pr(detection) versus # Interferers N



- Now variances are $\nu = Q$ and $\sigma_i^2 = QN$
- Ranking:
  1. EM-based low-rank: oo
  2. training-only low-rank: ++
  3. EM-based full-rank: o
  4. training-only full-rank: +

$\widehat{N}$ versus # Interferers N



- Again, variances are $\nu = Q$ and $\sigma_i^2 = QN$
- Rank-estimation successful in all cases.

Conclusions

- For adaptive detection of *known* signals in unknown-interference/noise environments, prior work includes:
  - Kelly'86: full-rank interference
  - Kang-Monga-Rangaswamy'14: low-rank interference nulling
  - McWhorter'04: low-rank interference cancellation
- For detection/synchronization of wireless communications signals, the common approach is to *ignore* the unknown data symbols, as in Bliss-Parker'10.
- For probabilistic signals $\boldsymbol{s} \sim p(\boldsymbol{s})$ we proposed three EM-based schemes:
  - full-rank interference (inspired by Kelly)
  - low-rank interference nulling (inspired by KMR)
  - low-rank interference cancellation (inspired by McWhorter)
- Numerical experiments suggest that
  - the EM-based methods outperform training-only methods
  - low-rank methods outperform full-rank methods
  - soft/EM-iterative methods outperform hard-iterative methods