Detection and Estimation Using Regularized Least Squares:

Performance Analysis and Optimal Tuning Under Uncertainty

Tareq Y. Al-Naffouri

King Abdullah University of Science and Technology
(KAUST)

Communications Research Lab
Ilmenau University of Technology

Aug 9, 2018

Objective

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}.$$

- $\mathbf{H}$ is a linear transformation which might be uncertain $(\mathbf{H} + \tilde{\mathbf{H}})$
- $\mathbf{H}$ could be i.i.d random or ill-conditioned
- $\mathbf{z}$ is the additive noise of unknown variance $\sigma_{\mathbf{z}}^2$
- $\mathbf{x}$ is the desired that we want to estimate or detect
- $\mathbf{x}$ can be deterministic or random with unknown statistics

We will focus on regularized least-squares (and variants) for detection/estimation

$$\min_{\mathbf{x}} \; ||\mathbf{y} - \mathbf{H}\mathbf{x}||^2 + \gamma \; ||\mathbf{x}||^2$$

Optimal Tuning of Regularized Least Squares

Joint work with Mohamed Suliman & Tarig Ballal

- Data model:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}.$$

- $\mathbf{H} \in \mathbb{C}^{m \times n}$ is the linear transformation matrix. (Known)
- $\mathbf{y} \in \mathbb{C}^{m \times 1}$ is the observation vector. (Known)
- $\mathbf{x} \in \mathbb{C}^{n \times 1}$ is the desired signal. (Unknown)
    - Stochastic: $\mathbf{R}_{\mathbf{x}} \triangleq \mathbb{E}\left(\mathbf{x}\mathbf{x}^H\right)$.
    - Deterministic: $\mathbf{R}_{\mathbf{x}} \triangleq \mathbf{x}\mathbf{x}^H$.

- Data model:

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}.$$

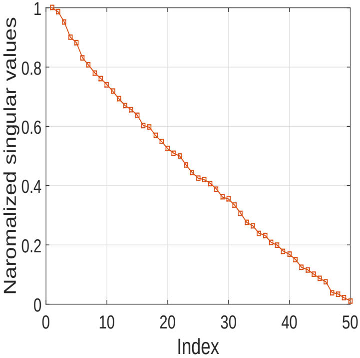
- $\mathbf{H} \in \mathbb{C}^{m\times n}$ is the linear transformation matrix. (Known)
- $\mathbf{y} \in \mathbb{C}^{m\times 1}$ is the observation vector. (Known)
- $\mathbf{x} \in \mathbb{C}^{n\times 1}$ is the desired signal with covarince matrix $\mathbf{R}_{\mathbf{x}}$. (Unknown)
- $\mathbf{z} \in \mathbb{C}^{m\times 1}$ is AWGN with variance $\sigma_{\mathbf{z}}^2$. (Unknown)
- $\mathbf{z}$ and $\mathbf{x}$ are independent.

Problems

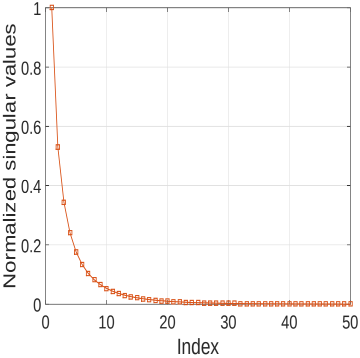
- Given $\mathbf{y}$ and $\mathbf{H}$, find an estimate of $\mathbf{x}$.
- Optimally tune γ

What type of $\mathbf{H}$?

(1) $\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$.

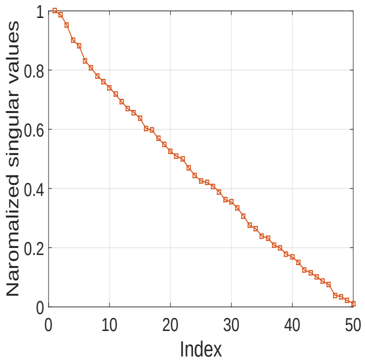


(2) $\mathbf{H}$ is highly ill-conditioned matrix.



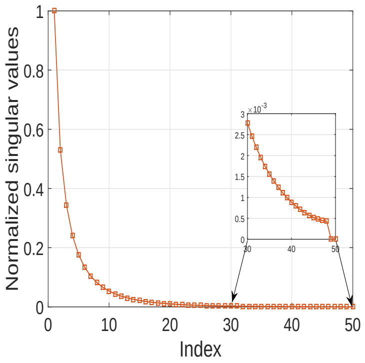
What type of $\mathbf{H}$?

(1) $\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$.



Condition Number $= 1.821 \times 10^3$.

(2) $\mathbf{H}$ is highly ill-conditioned matrix.



Condition Number $= 2.49 \times 10^{28}$.

Optimal Regularizer if Statistics are Known

- Noise variance $\sigma_{\mathbf{z}}^2$ is available
- Desired signal statistics are available
  - Stochastic: $\mathbf{R}_{\mathbf{x}} \triangleq \mathbb{E}\left(\mathbf{x}\mathbf{x}^H\right)$.
  - Deterministic: $\mathbf{R}_{\mathbf{x}} \triangleq \mathbf{x}\mathbf{x}^H$.

- Minimize MSE

$$\text{MSE} = \mathbb{E}[||\mathbf{x} - \hat{\mathbf{x}}||^2]$$

$$\gamma_{\text{o}} \approx \frac{m\sigma_{\mathbf{z}}^2}{\text{Tr}\left(\mathbf{R}_{\mathbf{x}}\right)} \tag{1}$$

Random matrix scenario

- Use deterministic equivalents

Discrete ill-posed scenario

- Use some trace bounds approximations

Relation Between γ_o and the LMMSE

- Our optimal regularizer is $\gamma_o \approx \frac{m\sigma_{\mathbf{z}}^2}{\mathrm{Tr}(\mathbf{R}_{\mathbf{x}})}$.
- Note that the LMMSE is given by

$$\hat{\mathbf{x}}_{\mathsf{LMMSE}} = \left(\mathbf{H}^H\mathbf{H} + \sigma_{\mathbf{z}}^2\mathbf{R}_{\mathbf{x}}^{-1}\right)^{-1}\mathbf{H}^H\mathbf{y}. \tag{2}$$

- When $\mathbf{x}$ is i.i.d. with zero mean, $\mathbf{R}_{\mathbf{x}} = \sigma_{\mathbf{x}}^2\mathbf{I}$.
- $\gamma_o = \frac{\sigma_{\mathbf{z}}^2}{\mathrm{Tr}(\mathbf{R}_{\mathbf{x}})/m} = \frac{\sigma_{\mathbf{z}}^2}{\sigma_{\mathbf{x}}^2}$.
- This shows that γ_o is optimal when the input is white.

Proposed Approach

- Recall the model: $$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}.$$
- Recall how the SV structure affects the result.

Proposed Approach

- Recall the model: $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}$.
- Recall how the SV structure affects the result.
- **We propose adding perturbation $\Delta\mathbf{H} \in \mathbb{C}^{m\times n}$ to $\mathbf{H}$.**
- We will impose bound on $\Delta\mathbf{H}$ (i.e., $0 \leq ||\Delta\mathbf{H}||_2 \leq \lambda$), why ?
- Perturbed model:

$$\mathbf{y} \approx (\mathbf{H} + \Delta\mathbf{H})\mathbf{x} + \mathbf{z}. \tag{3}$$

Proposed Approach

- Recall the model: $\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}.$
- Recall how the SV structure affects the result.
- **We propose adding perturbation $\Delta\mathbf{H} \in \mathbb{C}^{m \times n}$ to $\mathbf{H}$.**
- We will impose bound on $\Delta\mathbf{H}$ (i.e., $0 \le ||\Delta\mathbf{H}||_2 \le \lambda$), why ?
- Perturbed model:

$$\mathbf{y} \approx (\mathbf{H} + \Delta\mathbf{H})\mathbf{x} + \mathbf{z}. \tag{3}$$

Problem We know neither $\Delta\mathbf{H}$ nor λ.

- Judicious choice of λ is necessary.

Proposed Approach

- We call the proposed approach COnstrained Perturbation Regularization Approach (COPRA).

COPRA

- For now, let us assume we know the *best* choice of λ.

- We propose bounding the worst-case residual error

$$\min_{\hat{\mathbf{x}}} \max_{\Delta\mathbf{H}} ||\mathbf{y} - (\mathbf{H} + \Delta\mathbf{H})\,\hat{\mathbf{x}}||_2$$
$$\text{subject to: } ||\Delta\mathbf{H}||_2 \leq \lambda. \tag{4}$$

COPRA

- For now, let us assume we know the *best* choice of λ.
- We propose bounding the worst-case residual error

$$\min_{\hat{\mathbf{x}}} \max_{\Delta\mathbf{H}} ||\mathbf{y} - (\mathbf{H} + \Delta\mathbf{H})\,\hat{\mathbf{x}}||_2$$
$$\text{subject to: } ||\Delta\mathbf{H}||_2 \leq \lambda. \tag{4}$$

- After manipulations, the problem can be reduced to

$$\min_{\hat{\mathbf{x}}} \max_{\Delta\mathbf{H}} ||\mathbf{y} - (\mathbf{H} + \Delta\mathbf{H})\,\hat{\mathbf{x}}||_2 = \min_{\hat{\mathbf{x}}} ||\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}||_2 + \lambda\, ||\hat{\mathbf{x}}||_2.$$
$$\text{subject to: } ||\Delta\mathbf{H}||_2 \leq \lambda \tag{5}$$

COPRA

- Starting from

$$\min_{\hat{\mathbf{x}}} ||\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}||_2 + \lambda \; ||\hat{\mathbf{x}}||_2. \tag{6}$$

- Solution

$$\hat{\mathbf{x}} = \left(\mathbf{H}^H\mathbf{H} + \gamma\mathbf{I}\right)^{-1}\mathbf{H}^H\mathbf{y}. \tag{7}$$

- Where

$$K(\gamma, \lambda) = -\lambda^2||\mathbf{H}\left(\mathbf{H}^H\mathbf{H} + \gamma\mathbf{I}\right)^{-1}\mathbf{H}^H\mathbf{y} - \mathbf{y}||^2 + \gamma^2||\left(\mathbf{H}^H\mathbf{H} + \gamma\mathbf{I}\right)^{-1}\mathbf{H}\mathbf{y}||^2 = 0. \tag{8}$$

- We call (8) COPRA fundamental equation.
- How to proceed further ?

COPRA

- Starting from

$$\min_{\hat{\mathbf{x}}} ||\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}||_2 + \lambda\ ||\hat{\mathbf{x}}||_2. \tag{6}$$

- Solution

$$\hat{\mathbf{x}} = \left(\mathbf{H}^H\mathbf{H} + \gamma\mathbf{I}\right)^{-1}\mathbf{H}^H\mathbf{y}. \tag{7}$$

- Where

$$K\left(\gamma, \lambda\right) = -\lambda^2||\mathbf{H}\left(\mathbf{H}^H\mathbf{H} + \gamma\mathbf{I}\right)^{-1}\mathbf{H}^H\mathbf{y} - \mathbf{y}||^2 + \gamma^2||\left(\mathbf{H}^H\mathbf{H} + \gamma\mathbf{I}\right)^{-1}\mathbf{H}\mathbf{y}||^2 = 0. \tag{8}$$

- We call (8) COPRA fundamental equation.
- How to proceed further ?
- We will use the MSE criterion to select the bound λ for
  - Random matrix scenario.‘
  - Linear discrete ill-posed scenario.

(1) Random Matrix Scenario.

How to Find the Perturbation Bound λ ? (1) Random Scenario (R-COPRA)

- Recall COPRA fundamental equation (8)

$$\gamma_o{}^2 || \left(\mathbf{H}^H\mathbf{H} + \gamma_o\mathbf{I}\right)^{-1} \mathbf{H}\mathbf{y}||^2 - \lambda_o{}^2 ||\mathbf{H}\left(\mathbf{H}^H\mathbf{H} + \gamma_o\mathbf{I}\right)^{-1} \mathbf{H}^H\mathbf{y} - \mathbf{y}||^2 = 0.$$

How to Find the Perturbation Bound λ ? (1) Random Scenario (R-COPRA)

- Recall COPRA fundamental equation (8)

$$\gamma_o{}^2 || \left(\mathbf{H}^H\mathbf{H} + \gamma_o \mathbf{I}\right)^{-1} \mathbf{H}\mathbf{y}||^2 - \lambda_o{}^2 ||\mathbf{H}\left(\mathbf{H}^H\mathbf{H} + \gamma_o \mathbf{I}\right)^{-1} \mathbf{H}^H\mathbf{y} - \mathbf{y}||^2 = 0.$$

- Consider obtaining a perturbation bound that is approximately feasible for all the cases

$$\lambda_o{}^2 \ \mathbb{E}\Big[\underbrace{\sigma_{\mathbf{z}}^2 \mathrm{Tr}\left(\mathbf{H}^H\mathbf{H}\left(\mathbf{H}^H\mathbf{H} + m\tilde{\gamma}_o\mathbf{I}\right)^{-2}\right)}_{Q(\tilde{\gamma}_o)} + \underbrace{\mathrm{Tr}\left(\left(\mathbf{H}^H\mathbf{H} + m\tilde{\gamma}_o\mathbf{I}\right)^{-2}\mathbf{H}^H\mathbf{H}\mathbf{R}_{\mathbf{x}}\right)}_{R(\tilde{\gamma}_o)}\Big]$$

$$\approx \mathbb{E}\Big[\underbrace{\sigma_{\mathbf{z}}^2 \mathrm{Tr}\left(\mathbf{H}^H\mathbf{H}\left(\mathbf{H}^H\mathbf{H} + m\tilde{\gamma}_o\mathbf{I}\right)^{-2}\right)}_{G(\tilde{\gamma}_o)} + \underbrace{\mathrm{Tr}\left(\mathbf{H}^H\mathbf{H}\left(\mathbf{H}^H\mathbf{H} + m\tilde{\gamma}_o\mathbf{I}\right)^{-2}\mathbf{H}^H\mathbf{H}\mathbf{R}_{\mathbf{x}}\right)}_{T(\tilde{\gamma}_o)}\Big]. \tag{9}$$

$$\mathbb{E}\left(Q\left(\tilde{\gamma}_o\right)\right)=\frac{\sigma_{\mathbf{z}}^2\left(\sqrt{\frac{\tilde{\gamma}_o+4}{\tilde{\gamma}_o}}-1\right)^3\left(1-\frac{1}{4}\left(\sqrt{\frac{\tilde{\gamma}_o+4}{\tilde{\gamma}_o}}-1\right)^2\tilde{\gamma}_o\right)\tilde{\gamma}_o{}^3}{8\left(\tilde{\gamma}_o{}^2-\frac{1}{16}\left(\sqrt{\frac{\tilde{\gamma}_o+4}{\tilde{\gamma}_o}}-1\right)^4\tilde{\gamma}_o{}^4\right)}+\mathcal{O}\left(m^{-2}\right). \tag{10}$$

$$\mathbb{E}\left(R\left(\tilde{\gamma}_o\right)\right)=\frac{\tilde{\gamma}_o\left(\sqrt{\frac{\tilde{\gamma}_o+4}{\tilde{\gamma}_o}}-1\right)^3\operatorname{Tr}\left(\mathbf{R}_{\mathbf{x}}\right)}{4m\left(4-\tilde{\gamma}_o\left(\sqrt{\frac{\tilde{\gamma}_o+4}{\tilde{\gamma}_o}}-1\right)\right)}+\mathcal{O}\left(m^{-2}\right). \tag{11}$$

$$\mathbb{E}\left(T\left(\tilde{\gamma}_o\right)\right)=\frac{\tilde{\gamma}_o}{4}\left(-1+\sqrt{\frac{\tilde{\gamma}_o+4}{\tilde{\gamma}_o}}\right)^2\operatorname{Tr}\left(\mathbf{R}_{\mathbf{x}}\right)-\frac{\tilde{\gamma}_o{}^2\left(-1+\sqrt{\frac{\tilde{\gamma}_o+4}{\tilde{\gamma}_o}}\right)^3\operatorname{Tr}\left(\mathbf{R}_{\mathbf{x}}\right)}{4\left(-4+\tilde{\gamma}_o\left(-1+\sqrt{\frac{\tilde{\gamma}_o+4}{\tilde{\gamma}_o}}\right)\right)}+\mathcal{O}\left(m^{-2}\right). \tag{12}$$

R-COPRA

- After manipulations, we obtain

$$\lambda_{\circ}{}^{2} \approx \frac{\frac{\sigma_{\mathbf{z}}^{2}}{\mathrm{Tr}(\mathbf{R_x})}\left(2+\tilde{\gamma}_{\circ}-\sqrt{1+4\tilde{\gamma}_{\circ}{}^{-1}}\right)+2m\tilde{\gamma}_{\circ}{}^{2}\left(\left(\sqrt{1+4\tilde{\gamma}_{\circ}{}^{-1}}-1\right)\tilde{\gamma}_{\circ}+\sqrt{1+4\tilde{\gamma}_{\circ}{}^{-1}}-3\right)}{2\frac{\sigma_{\mathbf{z}}^{2}}{\mathrm{Tr}(\mathbf{R_x})}-\tilde{\gamma}_{\circ}\left(\tilde{\gamma}_{\circ}\left(\sqrt{1+4\tilde{\gamma}_{\circ}{}^{-1}}-1\right)-2\right)}. \tag{13}$$

R-COPRA

- After manipulations, we obtain

$$\lambda_o{}^2 \approx \frac{\frac{\sigma_{\mathbf{z}}^2}{\mathrm{Tr}(\mathbf{R_x})}\left(2+\tilde{\gamma}_o-\sqrt{1+4\tilde{\gamma}_o{}^{-1}}\right)+2m\tilde{\gamma}_o{}^2\left(\left(\sqrt{1+4\tilde{\gamma}_o{}^{-1}}-1\right)\tilde{\gamma}_o+\sqrt{1+4\tilde{\gamma}_o{}^{-1}}-3\right)}{2\frac{\sigma_{\mathbf{z}}^2}{\mathrm{Tr}(\mathbf{R_x})}-\tilde{\gamma}_o\left(\tilde{\gamma}_o\left(\sqrt{1+4\tilde{\gamma}_o{}^{-1}}-1\right)-2\right)}. \tag{13}$$

- From the MSE solution (1)

$$\frac{\sigma_{\mathbf{z}}^2}{\mathrm{Tr}\left(\mathbf{R_x}\right)} \to \frac{\tilde{\gamma}_o}{m}.$$

- Recall COPRA fundamental equation (8)

$$\gamma_o{}^2||\left(\mathbf{H}^H\mathbf{H}+\gamma_o\mathbf{I}\right)^{-1}\mathbf{H}\mathbf{y}||^2-\lambda_o{}^2||\mathbf{H}\left(\mathbf{H}^H\mathbf{H}+\gamma_o\mathbf{I}\right)^{-1}\mathbf{H}^H\mathbf{y}-\mathbf{y}||^2=0.$$

- Combining (13) and (8), then solving, yields to R-COPRA characteristic equation.

R-COPRA

R-COPRA Characteristic Equation

$$S_R\left(\tilde{\gamma}_o\right) = \text{Tr}\left(\boldsymbol{\Sigma}^2\left(\boldsymbol{\Sigma}^2 + m\tilde{\gamma}_o\mathbf{I}\right)^{-2}\mathbf{b}\mathbf{b}^H\right)\left[\tilde{\gamma}_o\left(\sqrt{\frac{\tilde{\gamma}_o + 4}{\tilde{\gamma}_o}} - 1\right) - 4\right]$$
$$+ \text{Tr}\left(\left(\boldsymbol{\Sigma}^2 + m\tilde{\gamma}_o\mathbf{I}\right)^{-2}\mathbf{b}\mathbf{b}^H\right)\left[m\tilde{\gamma}_o\left(\left(\sqrt{\frac{\tilde{\gamma}_o + 4}{\tilde{\gamma}_o}} - 1\right)\tilde{\gamma}_o + 2\sqrt{\frac{\tilde{\gamma}_o + 4}{\tilde{\gamma}_o}} - 4\right)\right] = 0, \quad (14)$$

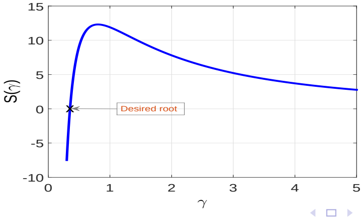
where $\mathbf{b} = \mathbf{U}^H\mathbf{y}$.

- Solving $S_R\left(\tilde{\gamma}_o\right)$ results in the regularization parameter $\tilde{\gamma}_o$.

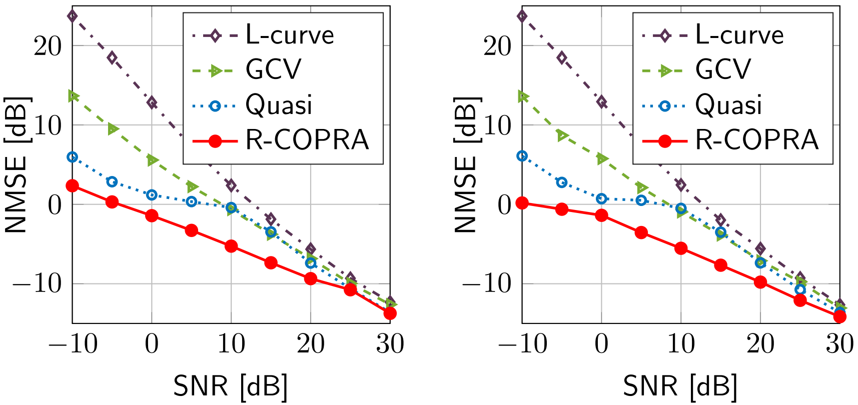
Question Can we solve (14) ?

Summary of the Properties for $S_R(\tilde{\gamma}_o)$

- $S_R(\tilde{\gamma}_o)$ is continuous over the interval $(0, +\infty)$.
- $\lim_{\tilde{\gamma}_o \to +\infty} S_R(\tilde{\gamma}_o) = 0$.
- $\lim_{\tilde{\gamma}_o \to 0^+} S_R(\tilde{\gamma}_o) = -4\ \mathrm{Tr}\left(\mathbf{\Sigma}^{-2}\mathbf{b}\mathbf{b}^H\right)$.
- $S_R(\tilde{\gamma}_o)$ is completely monotonic in the interval $(0, +\infty)$.
- Starting from $\tilde{\gamma}_o^{n=0}$, Newton's method will produce a consecutive increase estimation for $\tilde{\gamma}_o$.



Simulation Results: Stochastic $\mathbf{x}$



(a) $\mathbf{x}$ is i.i.d.

(b) $\mathbf{x}$ is i.n.d.

Figure: NMSE versus SNR for $\mathbf{H} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \mathbf{H} \in \mathbb{R}^{100\times 100}$.

Simulation Results: Deterministic $\mathbf{x}$

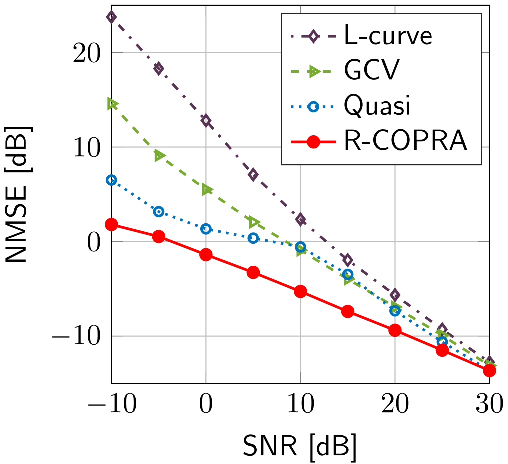
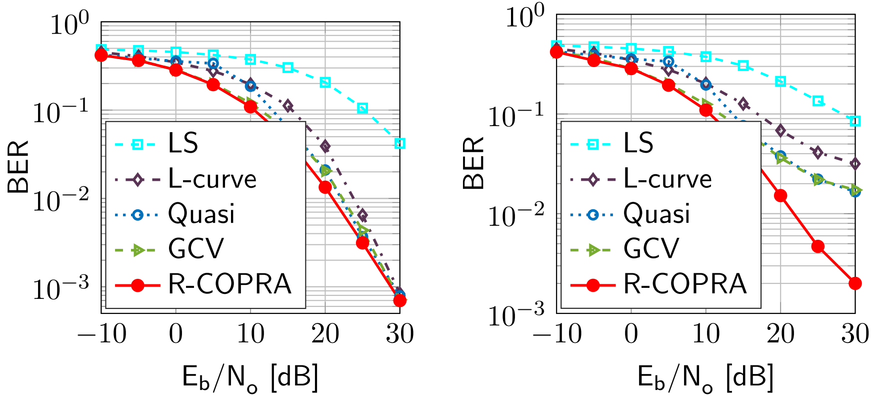


Figure: NMSE versus SNR for $\mathbf{H} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \mathbf{H} \in \mathbb{R}^{100 \times 100}$ and $\mathbf{x}$ is square pulse signal.

Simulation Results: Imperfect $\mathbf{H}$



(a) Perfect $\mathbf{H}$.

(b) Imperfect $\mathbf{H}$: $\hat{\mathbf{H}} = \mathbf{H} - e\mathbf{\Omega}$.

Figure: BER comparison when $\mathbf{H} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}), \mathbf{H} \in \mathbb{C}^{100\times 100}$ and $\mathbf{x}$ is 8-QAM signal.

Average Run Time

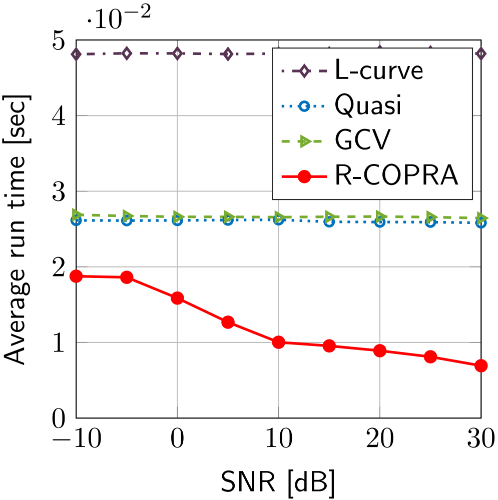


Figure: Average run time.

(2) Ill-posed Scenario.

How to Find the Perturbation Bound λ ? (2) Ill-posed Scenario (I-COPRA)

- Recall COPRA fundamental equation (8)

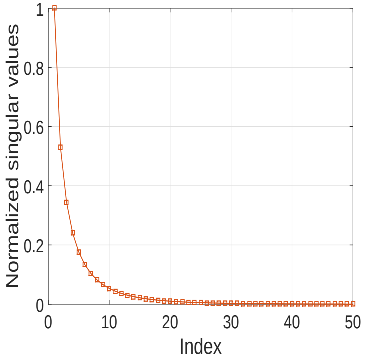
$$\gamma_o{}^2 \| \left(\mathbf{H}^H\mathbf{H} + \gamma_o\mathbf{I}\right)^{-1}\mathbf{H}\mathbf{y}\|^2 - \lambda_o{}^2 \|\mathbf{H}\left(\mathbf{H}^H\mathbf{H} + \gamma_o\mathbf{I}\right)^{-1}\mathbf{H}^H\mathbf{y} - \mathbf{y}\|^2 = 0.$$

- Manipulate to obtain

$$\lambda_o{}^2 \approx \frac{\sigma_{\mathbf{z}}^2 \mathrm{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2 + \gamma_o\mathbf{I}\right)^{-2}\right) + \mathrm{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2 + \gamma_o\mathbf{I}\right)^{-2}\mathbf{\Sigma}^2\mathbf{V}^H\mathbf{R}_{\mathbf{x}}\mathbf{V}\right)}{\sigma_{\mathbf{z}}^2 \mathrm{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2 + \gamma_o\mathbf{I}\right)^{-2}\right) + \mathrm{Tr}\left(\left(\mathbf{\Sigma}^2 + \gamma_o\mathbf{I}\right)^{-2}\mathbf{\Sigma}^2\mathbf{V}^H\mathbf{R}_{\mathbf{x}}\mathbf{V}\right)}. \tag{15}$$

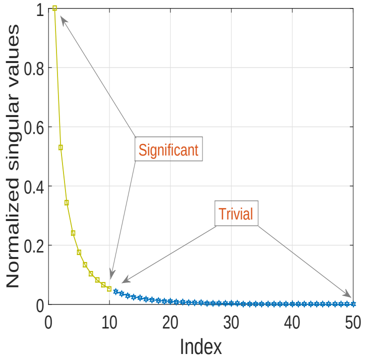
I-COPRA

- Recall the singular value structure.



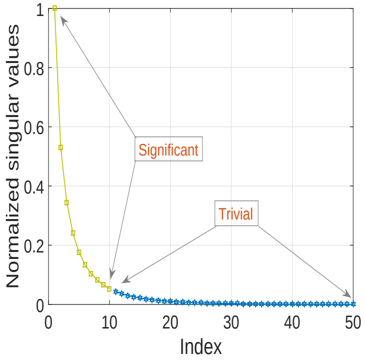
I-COPRA

- Divide $\boldsymbol{\Sigma}$ into m_1 large and m_2 small singular values.



I-COPRA

- Divide $\mathbf{\Sigma}$ into m_1 large and m_2 small singular values.
- Write $\mathbf{\Sigma} = \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma}_2 \end{bmatrix}$
- $\mathbf{\Sigma}_1 \in \mathbb{R}^{m_1 \times m_1}$ (large singular values).
- $\mathbf{\Sigma}_2 \in \mathbb{R}^{m_2 \times m_2}$ (small singular values).
- $\|\mathbf{\Sigma}_2\|^2 \ll \|\mathbf{\Sigma}_1\|^2$.



- Recall the optimal bound relation (15)

$$\lambda_o{}^2 \approx \frac{\sigma_{\mathbf{z}}^2 \mathrm{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\gamma_o\mathbf{I}\right)^{-2}\right)+\mathrm{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\gamma_o\mathbf{I}\right)^{-2}\mathbf{\Sigma}^2\mathbf{V}^H\mathbf{R}_{\mathbf{x}}\mathbf{V}\right)}{\sigma_{\mathbf{z}}^2 \mathrm{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\gamma_o\mathbf{I}\right)^{-2}\right)+\mathrm{Tr}\left(\left(\mathbf{\Sigma}^2+\gamma_o\mathbf{I}\right)^{-2}\mathbf{\Sigma}^2\mathbf{V}^H\mathbf{R}_{\mathbf{x}}\mathbf{V}\right)}.$$

- Apply the partitioning to (15), with some manipulations and reasonable approximations to obtain

$$\lambda_o{}^2 \approx \frac{\mathrm{Tr}\left(\mathbf{\Sigma}_1^2\left(\mathbf{\Sigma}_1^2+\gamma_o\mathbf{I}_1\right)^{-2}\left(\mathbf{\Sigma}_1^2+\frac{m_1\sigma_{\mathbf{z}}^2}{\mathrm{Tr}(\mathbf{R}_{\mathbf{x}})}\mathbf{I}_1\right)\right)}{\mathrm{Tr}\left(\left(\mathbf{\Sigma}_1^2+\gamma_o\mathbf{I}_1\right)^{-2}\left(\mathbf{\Sigma}_1^2+\frac{m_1\sigma_{\mathbf{z}}^2}{\mathrm{Tr}(\mathbf{R}_{\mathbf{x}})}\mathbf{I}_1\right)\right)+\frac{m_2}{\gamma_o{}^2}\frac{m_1\sigma_{\mathbf{z}}^2}{\mathrm{Tr}(\mathbf{R}_{\mathbf{x}})}}. \tag{16}$$

- From the MSE solution

$$\frac{m_1\sigma_{\mathbf{z}}^2}{\mathrm{Tr}\left(\mathbf{R}_{\mathbf{x}}\right)} \to \frac{m_1\gamma_o{}^2}{m}.$$

- Recall the optimal bound relation (15)

$$\lambda_o{}^2 \approx \frac{\sigma_{\mathbf{z}}^2 \text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\gamma_o\mathbf{I}\right)^{-2}\right)+\text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\gamma_o\mathbf{I}\right)^{-2}\mathbf{\Sigma}^2\mathbf{V}^H\mathbf{R}_{\mathbf{x}}\mathbf{V}\right)}{\sigma_{\mathbf{z}}^2 \text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\gamma_o\mathbf{I}\right)^{-2}\right)+\text{Tr}\left(\left(\mathbf{\Sigma}^2+\gamma_o\mathbf{I}\right)^{-2}\mathbf{\Sigma}^2\mathbf{V}^H\mathbf{R}_{\mathbf{x}}\mathbf{V}\right)}.$$

- Apply the partitioning to (15), with some manipulations and reasonable approximations to obtain

$$\lambda_o{}^2 \approx \frac{\text{Tr}\left(\mathbf{\Sigma}_1^2\left(\mathbf{\Sigma}_1^2+\gamma_o\mathbf{I}_1\right)^{-2}\left(\mathbf{\Sigma}_1^2+\frac{m_1\sigma_{\mathbf{z}}^2}{\text{Tr}(\mathbf{R}_{\mathbf{x}})}\mathbf{I}_1\right)\right)}{\text{Tr}\left(\left(\mathbf{\Sigma}_1^2+\gamma_o\mathbf{I}_1\right)^{-2}\left(\mathbf{\Sigma}_1^2+\frac{m_1\sigma_{\mathbf{z}}^2}{\text{Tr}(\mathbf{R}_{\mathbf{x}})}\mathbf{I}_1\right)\right)+\frac{m_2}{\gamma_o{}^2}\frac{m_1\sigma_{\mathbf{z}}^2}{\text{Tr}(\mathbf{R}_{\mathbf{x}})}}. \quad (16)$$

- From the MSE solution

$$\frac{m_1\sigma_{\mathbf{z}}^2}{\text{Tr}\left(\mathbf{R}_{\mathbf{x}}\right)} \to \frac{m_1\gamma_o{}^2}{m}.$$

- From COPRA fundamental equation (8)

$$\gamma_o{}^2||\left(\mathbf{H}^H\mathbf{H}+\gamma_o\mathbf{I}\right)^{-1}\mathbf{H}\mathbf{y}||^2-\lambda_o{}^2||\mathbf{H}\left(\mathbf{H}^H\mathbf{H}+\gamma_o\mathbf{I}\right)^{-1}\mathbf{H}^H\mathbf{y}-\mathbf{y}||^2=0.$$

- Combining (16) and (29), then solving, yields to I-COPRA characteristic equation.

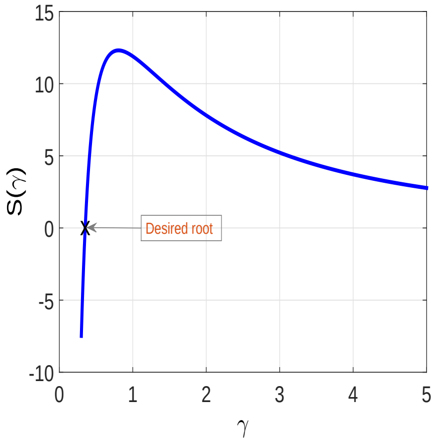
I-COPRA Characteristic Equation

$$
\begin{aligned}
S_I\left(\gamma_o\right) = \operatorname{Tr}\left(\boldsymbol{\Sigma}^2\left(\boldsymbol{\Sigma}^2+\gamma_o\mathbf{I}\right)^{-2}\mathbf{b}\mathbf{b}^H\right)\operatorname{Tr}\left(\left(\boldsymbol{\Sigma}_1^2+\gamma_o\mathbf{I}_1\right)^{-2}\left(\beta\boldsymbol{\Sigma}_1^2+\gamma_o\mathbf{I}_1\right)\right) \\
+\frac{m_2}{\gamma_o}\operatorname{Tr}\left(\boldsymbol{\Sigma}^2\left(\boldsymbol{\Sigma}^2+\gamma_o\mathbf{I}\right)^{-2}\mathbf{b}\mathbf{b}^H\right)-\operatorname{Tr}\left(\left(\boldsymbol{\Sigma}^2+\gamma_o\mathbf{I}\right)^{-2}\mathbf{b}\mathbf{b}^H\right) \\
\times\operatorname{Tr}\left(\boldsymbol{\Sigma}_1^2\left(\boldsymbol{\Sigma}_1^2+\gamma_o\mathbf{I}_1\right)^{-2}\left(\beta\boldsymbol{\Sigma}_1^2+\gamma_o\mathbf{I}_1\right)\right)=0, \quad (17)
\end{aligned}
$$

where $\mathbf{b} \triangleq \mathbf{U}^H\mathbf{y}$ and $\beta = \frac{m}{m_1}$.

- Solving $S_I\left(\gamma_o\right)$ results in the regularization parameter γ_o.
- The properties of the $S_I\left(\gamma_o\right)$ are studied and it is shown that Newton's method converges to the solution.
- We studied the special case of this function when $m_1 = n$ and

I-COPRA Properties



Simulation Results. (2) I-COPRA

- The algorithm is applied to a set of 11 real-worlds discrete ill-posed problems.

Simulation Results. (2) I-COPRA

- The algorithm is applied to a set of 11 real-worlds discrete ill-posed problems.

Regularization Tools

A Matlab Package for
Analysis and Solution of Discrete Ill-Posed Problems

Version 4.1 for Matlab 7.3

Per Christian Hansen

Informatics and Mathematical Modelling
Building 321, Technical University of Denmark
DK-2800 Lyngby, Denmark

pch@imm.dtu.dk
http://www.imm.dtu.dk/~pch

March 2008

The software described in this report was originally published in
Numerical Algorithms **6** (1994), pp. 1–35.

The *current version* is published in Numer. Algo. **46** (2007), pp. 189–194,
and it is available from www.netlib.org/numeralgo
and www.mathworks.com/matlabcentral/fileexchange

Simulation Results. (2) I-COPRA

- The algorithm is applied to a set of 11 real-worlds discrete ill-posed problems.

Table: Summary of the test problems.

Problem	Description	Condition Number
Tomo	Two-dimensional tomography	1.07×10^{3}
Shaw	One-dimensional image restoration	2.04×10^{18}
Heat	Inverse heat equation	2.94×10^{26}
Deriv2	Computation of second derivative	3.03×10^{03}
Gravity	One-dimensional gravity surveying problem	2.97×10^{11}
I-laplace	Inverse Laplace transformation	2.43×10^{33}
Baart	First kind Fredholm integral equation	4.09×10^{17}
Spikes	Test problem with a "spiky" solution	4.65×10^{18}
Wing	Test problem with a discontinuous solution	1.68×10^{18}
Foxgood	Severely ill-posed test problem	2.43×10^{18}
Phillips	Phillips "famous" test problem	1.91×10^{5}

Simulation Results. (2) I-COPRA

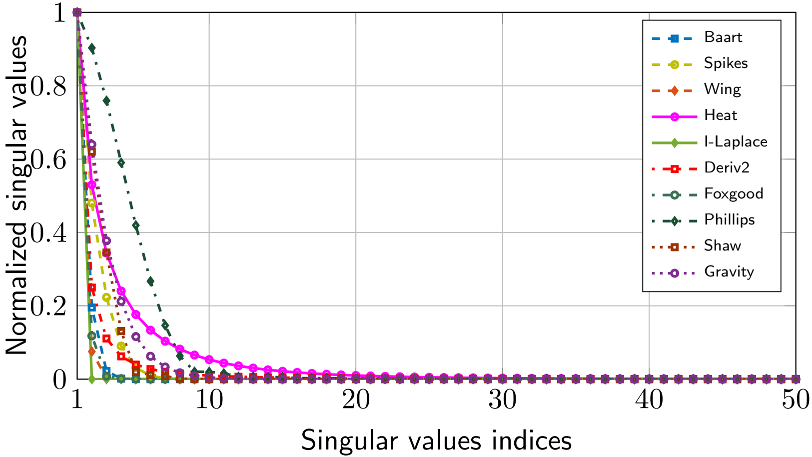
- The algorithm is applied to a set of 11 real-worlds discrete ill-posed problems.

Table: Summary of the test problems.

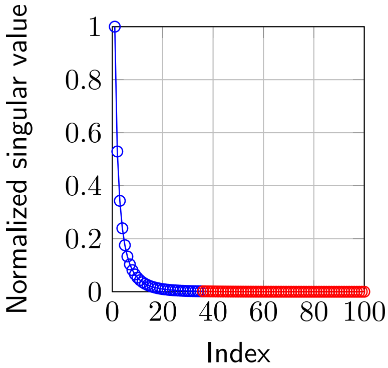
Problem	Description	Condition Number
Tomo	Image processing	1.07×10^{3}
Shaw		2.04×10^{18}
Heat	Applied Physics	2.94×10^{26}
Deriv2		3.03×10^{03}
Gravity		2.97×10^{11}
I-laplace		2.43×10^{33}
Baart	Signal Processing	4.09×10^{17}
Spikes		4.65×10^{18}
Wing		1.68×10^{18}
Foxgood		2.43×10^{18}
Phillips		1.91×10^{5}

Simulation Results. (2) I-COPRA

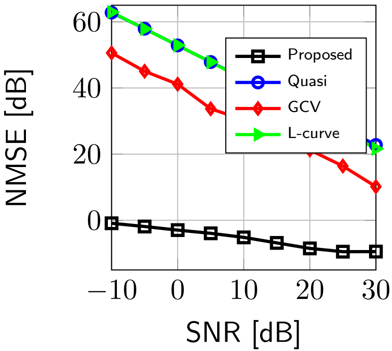
- The algorithm is applied to a set of 11 real-worlds discrete ill-posed problems.



- Heat Problem



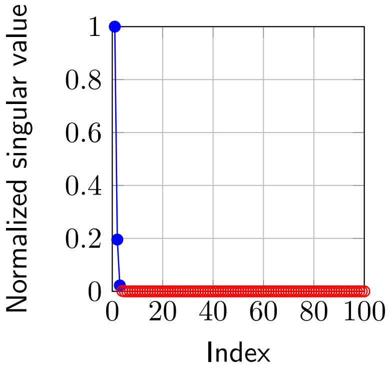
Singular values.



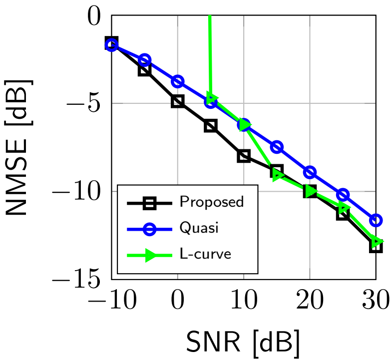
Performance.

Condition number of $\mathbf{H} = 6.8 \times 10^{36}$. $(m_1 = 10, m_2 = 40)$.

- Baart Problem



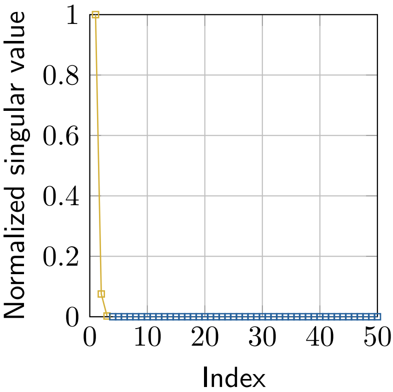
Singular values.



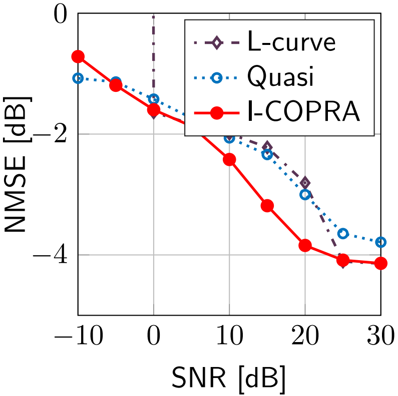
Performance.

Condition number of $\mathbf{H} = 2.89 \times 10^{18}$. $(m_1 = 3, m_2 = 47)$.

- Wing Problem



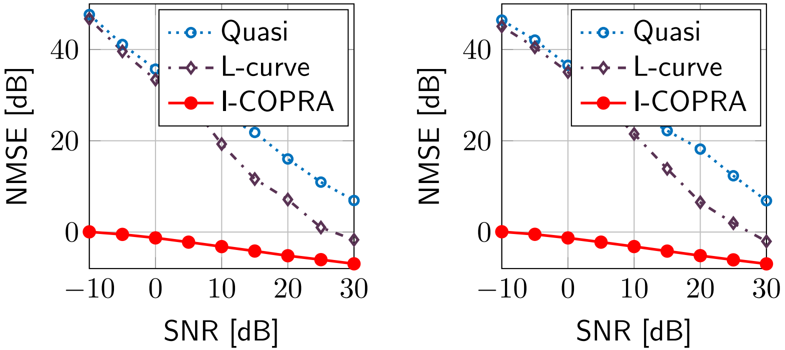
Singular values.



Performance.

Condition number of $\mathbf{H} = 1.68 \times 10^{18}$. $(m_1 = 3, m_2 = 47)$.

- Rank Deficient Matrices



(a) $\mathbf{x}$ is i.i.d. (b) $\mathbf{x}$ is i.n.d.

Figure: $\mathbf{H} = \frac{1}{50}\mathbf{B}\mathbf{B}^H$, where $\mathbf{B} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}), \mathbf{B} \in \mathbb{R}^{50\times 45}$.

- Special Case: $(m_1 = n$ and $m_2 = 0)$

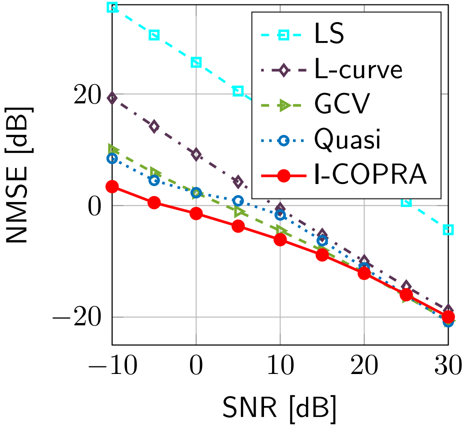


Figure: $\mathbf{H} \in \mathbb{R}^{100\times 100}$ is Toeplitz matrix and $\mathbf{x}$ is i.i.d.

Condition number of $\mathbf{H} = 389.51$.

Example of the Average Run Time

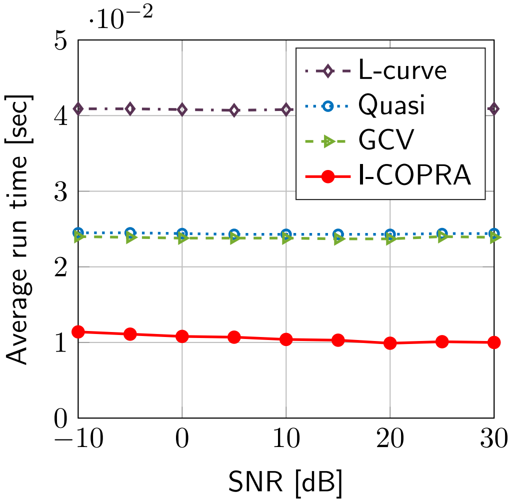
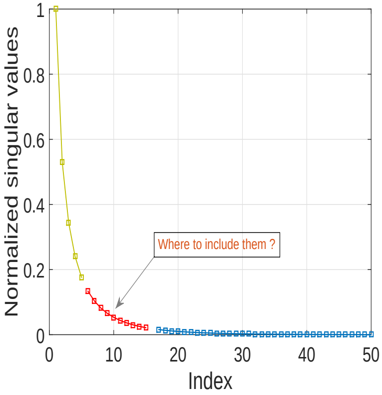


Figure: Average run time

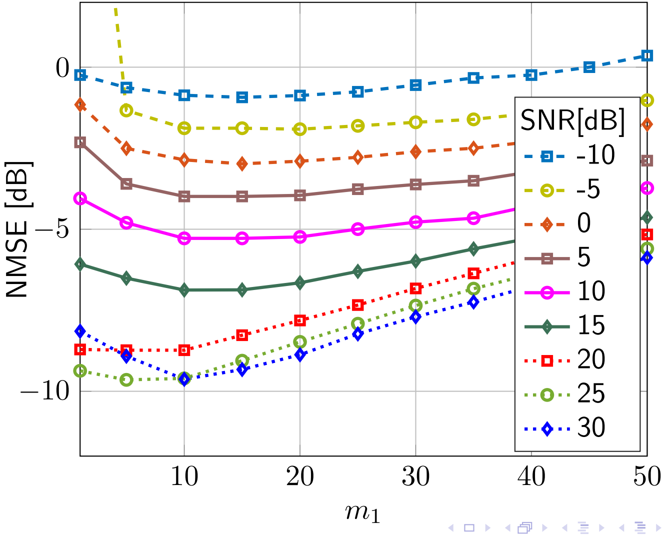
Sensitivity to the Choice of m_1

- Problem: Heat.



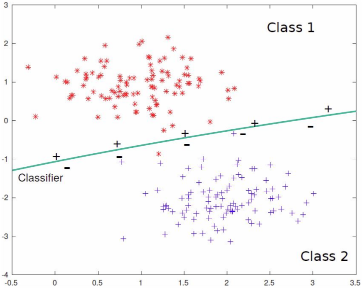
Sensitivity to the Choice of m_1

- Problem: Heat.



Discriminant Analysis

- Widely used statistical method for supervised classification
- Principle: Builds a classification rule that allows to assign for an unseen observation its corresponding class.



- Let x be the input data and f be the classification rule.

$$\text{Classifier} \triangleq \begin{cases} \text{Assign class 1 if } f(x) => 0 \\ \text{Assign class 2 if } f(x) =< 0 \end{cases}$$

Gaussian Discriminant Analysis

Gaussian mixture model for binary classification (2 classes)

- $x_1, \cdots, x_n \in \mathbb{R}^p$
- Class k is formed by $x \sim \mathcal{N}(\mu_k, \Sigma_k), k = 0, 1$

LDA Decision rule is linear in x:$\Sigma_0 = \Sigma_1$

$$W^{LDA} = (x - \tfrac{\mu_0+\mu_1}{2})^T \Sigma^{-1}(\mu_0 - \mu_1) - \tfrac{\pi_1}{\pi_0})$$

$$\begin{cases} \text{Assign } x \text{ to class 0 if} & W^{LDA} > 0 \\ \text{Assign} x \text{ to class 1 if} & \text{otherwise} \end{cases}$$

- Statistics are unknown and so need to be estimated.
- Covariance matrix will be ill-conditioned when sample size is less than the data dimension p.
- Regularization could solve the problem but the choice of the regularization parameter is an issue.

Re-write the LDA score function as

$$
\begin{aligned}
\hat{W}^{LDA}(x) &= (x - \frac{\hat{\mu}_0 + \hat{\mu}_1}{2})^T \hat{\Sigma}^{-1}(\hat{\mu_0} - \hat{\mu}_1) \\
&= a^T \hat{\Sigma}^{-1/2} \hat{\Sigma}^{-1/2} b \\
&= w^T z
\end{aligned}
$$

where

$$
w = \hat{\Sigma}^{-1/2} a \quad \& \quad z = \hat{\Sigma}^{-1/2} b
$$

which can be obtained by solving the liner systems

$$
a = \hat{\Sigma}^{1/2} w \quad \& \quad b = \hat{\Sigma}^{1/z} b
$$

Classification of digits from MINST data set

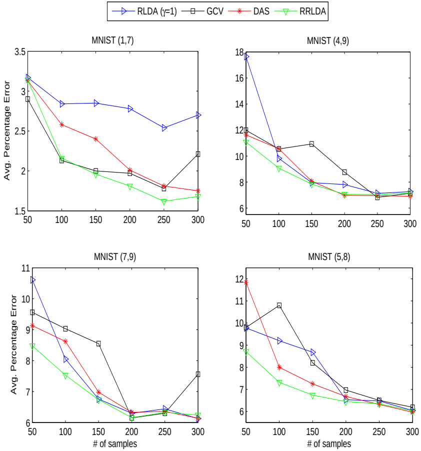


Figure: Error rate performance of different LDA classifiers using handwritten digits from MNIST dataset. The results are averaged over 50 Monte Carlo trials.

Beamforming

- The output of the beamformer can be written as

$$y_{\mathsf{BF}}[t] = \mathbf{w}^H \mathbf{y}[t], \tag{18}$$

- For the Capon/MVDR beamformer, the weighing coefficients are given by

$$\mathbf{w}_{\mathsf{MVDR}} = \frac{\hat{\mathbf{C}}_{\mathbf{yy}}^{-1}\mathbf{a}}{\mathbf{a}^H \hat{\mathbf{C}}_{\mathbf{yy}}^{-1}\mathbf{a}}, \tag{19}$$

where $\mathbf{a}$ is the array *steering vector* and $\hat{\mathbf{C}}_{\mathbf{yy}}$ is the sample covariance matrix of the received signals.

- Based on (18) and (19), we can write

$$\mathbf{y}_{\mathsf{BF}}[t] = \frac{\hat{\mathbf{a}}^H \hat{\mathbf{C}}_{\mathbf{yy}}^{-\frac{1}{2}} \hat{\mathbf{C}}_{\mathbf{yy}}^{-\frac{1}{2}} \mathbf{y}}{\mathbf{a}^H \ \hat{\mathbf{C}}_{\mathbf{yy}}^{-\frac{1}{2}} \hat{\mathbf{C}}_{\mathbf{yy}}^{-\frac{1}{2}} \mathbf{a}} = \frac{\mathbf{b}^H \mathbf{z}}{\mathbf{b}^H \mathbf{b}}, \tag{20}$$

where $\mathbf{b} \triangleq \hat{\mathbf{C}}_{\mathbf{yy}}^{-\frac{1}{2}} \mathbf{a}$ and $\mathbf{z} \triangleq \hat{\mathbf{C}}_{\mathbf{yy}}^{-\frac{1}{2}} \mathbf{y}$.

Application: Beamforming

- The two relationships of $\mathbf{a}$ and $\mathbf{b}$ can be thought of as

$$\mathbf{a} = \hat{\mathbf{C}}_{\mathbf{yy}}^{\frac{1}{2}} \mathbf{b}, \tag{21}$$

and

$$\mathbf{y} = \hat{\mathbf{C}}_{\mathbf{yy}}^{\frac{1}{2}} \mathbf{z}. \tag{22}$$

- Since $\hat{\mathbf{C}}_{\mathbf{yy}}^{\frac{1}{2}}$ is ill-conditioned, direct inversion does not provide a viable solution.

- Our regularization approach can be used to obtain estimates of $\mathbf{b}$ and $\mathbf{z}$ given that they are noisy.

Application: Beamforming

- Recall (20)

$$\mathbf{y}_{\mathsf{BF}}[t] = \frac{\hat{\mathbf{a}}^H \hat{\mathbf{C}}_{\mathbf{yy}}^{-\frac{1}{2}} \hat{\mathbf{C}}_{\mathbf{yy}}^{-\frac{1}{2}} \mathbf{y}}{\mathbf{a}^H \ \hat{\mathbf{C}}_{\mathbf{yy}}^{-\frac{1}{2}} \hat{\mathbf{C}}_{\mathbf{yy}}^{-\frac{1}{2}} \mathbf{a}} = \frac{\mathbf{b}^H \mathbf{z}}{\mathbf{b}^H \mathbf{b}}, \tag{23}$$

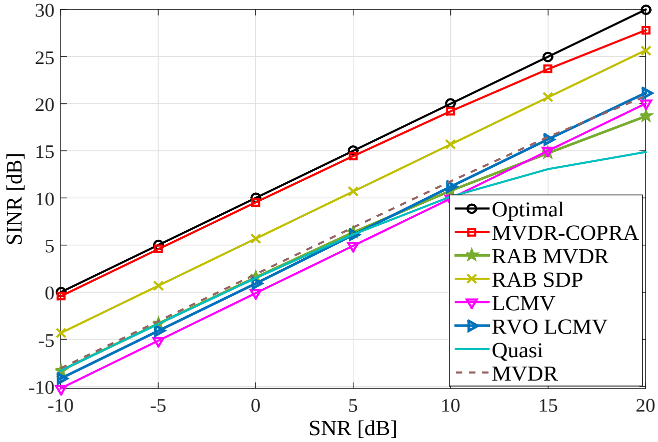
- Using regularization we can write

$$y_{\mathsf{BF\text{-}RLS}} = \frac{\mathbf{a}^H \mathbf{U} \left(\boldsymbol{\Sigma}^2 + \gamma_b \mathbf{I}\right)^{-1} \left(\boldsymbol{\Sigma}^2 + \gamma_z \mathbf{I}\right)^{-1} \boldsymbol{\Sigma}^2 \mathbf{U}^H \mathbf{y}}{\mathbf{a}^H \mathbf{U} \left(\boldsymbol{\Sigma}^2 + \gamma_b \mathbf{I}\right)^{-2} \boldsymbol{\Sigma}^2 \mathbf{U}^H \mathbf{a}}, \tag{24}$$

- Equation (24) suggests that the weighting coefficients for the RLS approach are given by

$$w_{\mathsf{BF\text{-}RLS}} = \frac{\mathbf{a}^H \mathbf{U} \left(\boldsymbol{\Sigma}^2 + \gamma_b \mathbf{I}\right)^{-1} \left(\boldsymbol{\Sigma}^2 + \gamma_z \mathbf{I}\right)^{-1} \boldsymbol{\Sigma}^2 \mathbf{U}^H}{\mathbf{a}^H \mathbf{U} \left(\boldsymbol{\Sigma}^2 + \gamma_b \mathbf{I}\right)^{-2} \boldsymbol{\Sigma}^2 \mathbf{U}^H \mathbf{a}}. \tag{25}$$

Beamforming: Simulation result



Conclusion of Part I

- We proposed a new regularization approach for linear least-square problems based on allowing a bounded perturbation into the linear transformation matrix.
- We chose the perturbation bound based on the MSE criteria and as a result, the proposed approach minimizes the MSE approximately.
- The solution of the proposed approach characteristic equation does not require knowledge of the signal and noise statistics.
- Solution performs well compared to other methods over a wide SNR range.
- The proposed approach is shown to have the lowest run time.

Regularized Least Squares for Massive MIMO: Precise Analysis and Optimal Tuning

Joint work with Ismail Atitallah, Ayed Alrashdi , and & Christos Thrampoulidis

MIMO System model: AWGN channel

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$$

- $\mathbf{y} \in \mathbb{R}^m$ is the measurement vector at the receive antennas.
- $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the channel matrix, with *iid* Gaussian entries, with zero mean and variance $\frac{1}{n}$.
- $\mathbf{x}_0 \in \{-1, 1\}^n$ is a BPSK signal.
- $\mathbf{z} \in \mathbb{R}^m$ is a additive white Gaussian noise vector with variance σ_z^2 $\Rightarrow$ SNR$= \frac{1}{\sigma_z^2}$.
- $\delta = \frac{m}{n}$ is the ratio of the number of receive/transmit antennas.

MIMO System model: AWGN channel

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$$

- $\mathbf{y} \in \mathbb{R}^m$ is the measurement vector at the receive antennas.
- $\mathbf{A} \in \mathbb{R}^{m \times n}$ is the channel matrix, with *iid* Gaussian entries, with zero mean and variance $\frac{1}{n}$.
- $\mathbf{x}_0 \in \{-1, 1\}^n$ is a BPSK signal.
- $\mathbf{z} \in \mathbb{R}^m$ is a additive white Gaussian noise vector with variance σ_z^2 $\Rightarrow$ SNR$= \frac{1}{\sigma_z^2}$.
- $\delta = \frac{m}{n}$ is the ratio of the number of receive/transmit antennas.

Optimum Receiver: Maximum Likelihood:

$$\widehat{\mathbf{x}}_{\mathrm{ML}} = \underset{\mathbf{x} \in \{-1,1\}^n}{\arg\min} \ \|\mathbf{A}\mathbf{x} - \mathbf{y}\|$$

$\Rightarrow$ computationally prohibitive in a massive MIMO context

Low-Complexity Receivers (1)

Two-step implementation of low-complexity receivers:

- Solve a convex optimization.
- Hard-threshold.

Examples of common low-complexity receivers:

- Least Squares (LS), aka Zero-Forcing receiver,

$$\widehat{\mathbf{x}}_{\mathrm{LS}} = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y},$$

$$\mathbf{x}^*_{\mathrm{LS}} = \mathsf{sign}(\widehat{\mathbf{x}}_{\mathrm{LS}}).$$

Low-Complexity Receivers (1)

Two-step implementation of low-complexity receivers:

- Solve a convex optimization.
- Hard-threshold.

Examples of common low-complexity receivers:

- Least Squares (LS), aka Zero-Forcing receiver,

$$\widehat{\mathbf{x}}_{\mathrm{LS}} = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2 = (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y},$$
$$\mathbf{x}^*_{\mathrm{LS}} = \mathsf{sign}(\widehat{\mathbf{x}}_{\mathrm{LS}}).$$

- Regularized Least Squares (RLS),

$$\widehat{\mathbf{x}}_{\mathrm{RLS}} = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \|\mathbf{A}\mathbf{x}-\mathbf{y}\|^2 + \lambda\|\mathbf{x}\|^2 = (\mathbf{A}^T\mathbf{A}+\lambda\mathbf{I})^{-1}\mathbf{A}^T\mathbf{y},$$
$$\mathbf{x}^*_{\mathrm{RLS}} = \mathsf{sign}(\widehat{\mathbf{x}}_{\mathrm{RLS}}).$$

Low-Complexity Receivers (2)

- RLS with Box Relaxation Optimization (RLS-BRO)

$$\widehat{\mathbf{x}}_{\mathrm{BRO}} = \underset{\mathbf{x}\in[-1,1]^n}{\arg\min} \|\mathbf{A}\mathbf{x}-\mathbf{y}\| + \lambda\|\mathbf{x}\|^2,$$
$$\mathbf{x}^*_{\mathrm{BRO}} = \mathsf{sign}(\widehat{\mathbf{x}}_{\mathrm{BRO}}).$$

  - No closed-form expression
  - quadratic program $\Rightarrow$ the complexity is also cubic.

Low-Complexity Receivers (2)

- RLS with Box Relaxation Optimization (RLS-BRO)

$$\widehat{\mathbf{x}}_{\mathrm{BRO}} = \arg\min_{\mathbf{x}\in[-1,1]^n} \|\mathbf{A}\mathbf{x} - \mathbf{y}\| + \lambda\|\mathbf{x}\|^2,$$
$$\mathbf{x}^*_{\mathrm{BRO}} = \mathsf{sign}(\widehat{\mathbf{x}}_{\mathrm{BRO}}).$$

  - No closed-form expression
  - quadratic program $\Rightarrow$ the complexity is also cubic.
- Aim:
  - Derive precise BER expression
  - Find optimum regularizer λ
  - Find optimum Box threhsold

Relevant literature

$\text{BER} := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{\mathbf{x}*_i \neq \mathbf{x}_{0,i}\}}.$

Receiver	BER approach	Reference
LS	Exact	Exact non-asymptotic formula, e.g. Tse and Viswanath [1]
RLS	RMT	Tulino and Verdu [2]
LS-BRO	CGMT	Thrampoulidis and Hassibi[3]
RLS-BRO	CGMT	This Talk

RMT: Random Matrix Theory
CGMT: Convex Gaussian Min-max Theorem

Asymptotic BER Analysis

- LS: $\lim_{n\to\infty} \mathrm{BER}_{\mathrm{LS}} = Q\left((\delta-1)\mathsf{snr}\right), \quad (\text{for } \delta > 1).$
- RLS: $\lim_{n\to\infty} \mathrm{BER}_{\mathrm{RLS}} = Q\left(\sqrt{\frac{\delta - \frac{1}{(1+\Upsilon(\lambda,\delta))^2}}{\left(\frac{\Upsilon(\lambda,\delta)}{1+\Upsilon(\lambda,\delta)}\right)^2 + \frac{1}{\mathsf{snr}}}}\right)$, where

$$\Upsilon(\lambda,\delta) = \frac{1-\delta+\lambda+\sqrt{(1-\delta+\lambda)^2+4\lambda\delta}}{2\delta}.$$

Asymptotic BER Analysis

- LS: $\lim_{n\to\infty} \mathrm{BER}_{\mathrm{LS}} = Q\left((\delta-1)\mathsf{snr}\right), \quad (\text{for } \delta > 1).$
- RLS: $\lim_{n\to\infty} \mathrm{BER}_{\mathrm{RLS}} = Q\left(\sqrt{\frac{\delta - \frac{1}{(1+\Upsilon(\lambda,\delta))^2}}{\left(\frac{\Upsilon(\lambda,\delta)}{1+\Upsilon(\lambda,\delta)}\right)^2 + \frac{1}{\mathsf{snr}}}}\right)$, where

$$\Upsilon(\lambda,\delta) = \frac{1-\delta+\lambda+\sqrt{(1-\delta+\lambda)^2+4\lambda\delta}}{2\delta}.$$

- The optimal λ that minimizes the asymptotic $\mathrm{BER}_{\mathrm{RLS}}$ is $\frac{1}{SNR}$ $\Rightarrow$ LMMSE receiver is also optimal in the BER sense.
- A high SNR approximation of the BER of LMMSE is $Q\left(\left(\delta - 1 + \frac{1}{(\delta-1)\mathrm{SNR}}\right)\mathrm{SNR}\right) \simeq Q((\delta-1)\mathrm{SNR})$.

Convex Gaussian Min-max Theorem

Convex-Gaussian Min-Max Theorem (CGMT)

[a]Consider the following two min-max problems:

Primary Optimization (PO) Problem: $\Phi(\mathbf{G}) := \min_{\mathbf{w} \in \mathcal{S}_w} \max_{\mathbf{u} \in \mathcal{S}_u} \mathbf{u}^T \mathbf{G} \mathbf{w} + \psi(\mathbf{w}, \mathbf{u})$

Auxilary Optimization (AO) Problem: $\phi(\mathbf{g}, \mathbf{h}) := \min_{\mathbf{w} \in \mathcal{S}_w} \max_{\mathbf{u} \in \mathcal{S}_u} \|\mathbf{w}\| \mathbf{g}^T \mathbf{u} - \|\mathbf{u}\| \mathbf{h}^T \mathbf{w} + \psi(\mathbf{w}, \mathbf{u})$

- ψ is convex-concave.
- $\mathbf{w}_\Phi$ any optimal minimizers in the (PO).
- $\mathbf{w}_\phi$ any optimal minimizers in the (AO).

Then, if $\lim_{n \to \infty} Pr(\mathbf{w}_\phi \in \mathcal{S}) = 1$, it also holds $\lim_{n \to \infty} Pr(\mathbf{w}_\Phi \in \mathcal{S}) = 1$.

[a]C. Thrampoulidis, E. Abbasi and B. Hassibi "Precise error analysis of regularized M-estimators in high-dimensions"- arXiv preprint arXiv:1601.06233, 2016

- We apply the CGMT to the set in which the BER concentrates, *i.e.*

$$\mathcal{S} = \left\{ \mathbf{w}; |\mathrm{BER} - \mathbb{E}[\mathrm{BER}]| < \epsilon \right\}. \quad (26)$$

Precise Bit Error Rate (BER) Analysis

Theorem (BER of RLS-BRO)

As $n, m \to \infty$, such that $\frac{m}{n} \to \delta \in (0, \infty)$, it holds in probability

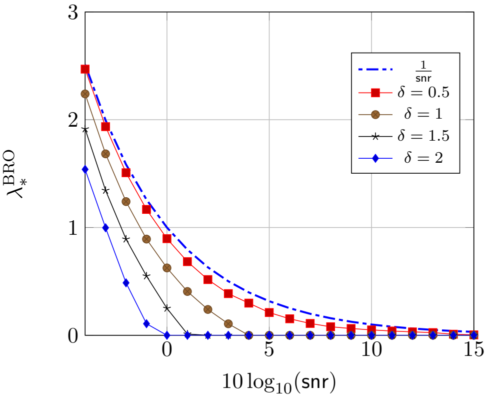
$$\lim_{n \to \infty} \mathrm{BER}_{\mathrm{BRO}} = Q(\frac{1}{\tau_*}),$$

where τ_* is the unique solution to the following

$$\min_{\tau>0} \max_{\beta>0} D(\tau, \beta) := \delta\tau\beta + \frac{\beta}{\mathrm{SNR}\tau} - \frac{\lambda\beta^2}{2} + \frac{4\beta}{\tau} Q\left(\frac{2}{\tau} + \frac{2}{\beta}\right) - 4\beta p\left(\frac{2}{\tau} + \frac{2}{\beta}\right) - \frac{\beta^2}{\frac{\beta}{\tau} + 2} \int_{-\frac{2}{\beta} - \frac{2}{\tau}}^{\frac{2}{\beta}} \left(h - \frac{2}{\beta}\right)^2 p(h) \mathrm{d}h.$$

- τ_* can be efficiently computed by writing the first order optimality conditions, *i.e.* $\nabla_{(\tau,\beta)} D(\tau, \beta) = 0$.
- τ_* (and hence the BER) depends on $\delta = \frac{m}{n}$, SNR and the regularizer λ.

Optimal tuning of the Regularizer



- The optimal regularizer λ_*^{BRO} is an decreasing function of the ratio $\frac{m}{n}$.
- It is always below $\frac{1}{\mathrm{SNR}}$.
- $\exists \overline{\mathsf{snr}} \in \mathbb{R}_+$, such that, $\lambda_*^{\mathrm{BRO}} = 0$ for all $\mathsf{snr} \in (\overline{\mathsf{snr}}, \infty)$.

LMMSE vs RLS-BRO

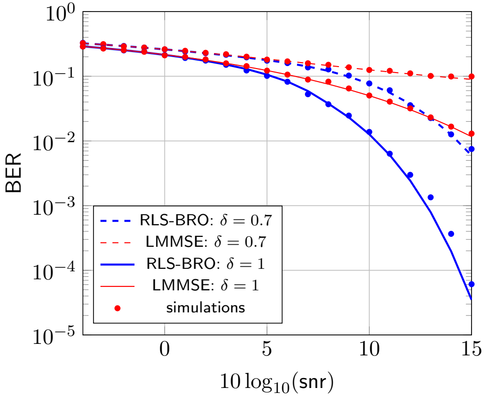


Figure: $n = 500$.

Recall the following high-SNR approximations:

- $\lim_{n\to\infty} \text{BER}_{BRO} \simeq Q((\delta - \frac{1}{2})\text{snr})$
- $\lim_{n\to\infty} \text{BER}_{RLS} \simeq Q((\delta - 1)\text{snr})$

Is $[-1, 1]$ the optimal relaxation interval?

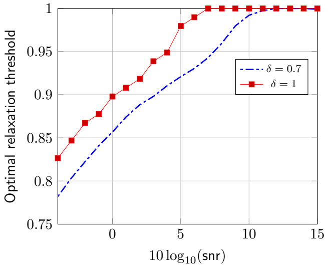
$$\widehat{\mathbf{x}}_{\mathrm{BRO}} = \arg\min_{\mathbf{x}\in[-t,t]^n} \|\mathbf{A}\mathbf{x} - \mathbf{y}\| + \lambda\|\mathbf{x}\|^2$$

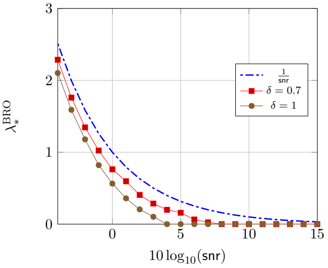
$$\mathbf{x}^*_{\mathrm{BRO}} = \mathsf{sign}(\widehat{\mathbf{x}}_{\mathrm{BRO}})$$

- In a similar fashion, we can prove that $\lim_{n\to\infty} \mathrm{BER}_{\mathrm{BRO}} = Q\left(\frac{1}{\tau_*}\right)$, where $\tau* = \arg\min_{\tau>0} \max_{\beta>0} D(\tau, \beta; t, \lambda, \delta, SNR)$.
- $\tau_*(t, \lambda, \delta, \mathrm{SNR})$ is a function of SNR, the sampling ratio δ, the regularizer λ and the relaxation threshold t.
- We select the optimal relaxation threshold t^*, and the optimal regularizer λ^*, such that:

$$(t^*, \lambda^*) \in argmin_{(t,\lambda)\in\mathbb{R}^2_+} \tau_*(t, \lambda, \delta, \mathrm{SNR})$$

Joint optimization of the regularizer and the relaxation threshold





What if we don't know the SNR?

If the signal and noise variances are not known, we use the expression of the cost function of RLS to estimate the SNR, to ultimately allow for an optimal tuning of the regularizer. Let $J(\sigma_x^2, \sigma_z^2)$ denote the asymptotic cost function of RLS.

$$\begin{aligned} J(\sigma_x^2, \sigma_z^2, \lambda, \delta) &:= \lim_{n\to\infty} \min_{\mathbf{x}} \|\mathbf{A}\mathbf{x} - \mathbf{y}\|^2 + \lambda\|\mathbf{x}\|^2. \\ &= a(\lambda, \delta)\sigma_x^2 + b(\lambda, \delta)\sigma_z^2, \end{aligned}$$

where

$$a(\lambda,\delta) = \left(\frac{\delta\lambda}{\Upsilon} - \frac{\lambda^2}{\Upsilon^2} - \frac{\lambda}{\Upsilon + \Upsilon^2}\right)\left(\frac{\Upsilon^2}{\delta(1+\Upsilon)^2 - 1}\right) - \frac{\lambda\Upsilon}{1+\Upsilon} + \lambda,$$

$$b(\lambda,\delta) = \left(\frac{\delta\lambda}{\Upsilon} - \frac{\lambda^2}{\Upsilon^2} - \frac{\lambda}{\Upsilon + \Upsilon^2}\right)\left(\frac{(1+\Upsilon)^2}{\delta(1+\Upsilon)^2 - 1}\right) + \frac{\lambda}{\Upsilon},$$

and

$$\Upsilon = \frac{1 - \delta + \lambda + \sqrt{(1-\delta+\lambda)^2 + 4\lambda\delta}}{2\delta}.$$

What if we don't know the SNR?

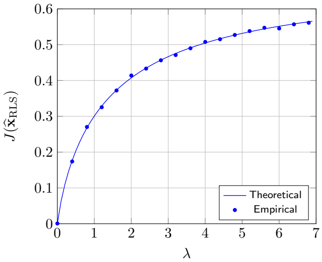


Figure: $m = 500$, $n = 800$, $\sigma_x^2 = 1$ and $\sigma_z^2 = 0.3$.

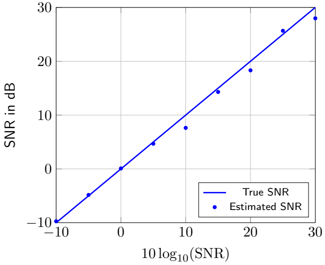


Figure: $m = 500$, $n = 800$ and $\sigma_x^2 = 1$.

What if we don't know the SNR?

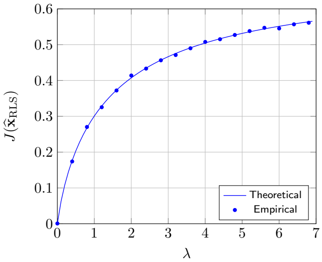


Figure: $m = 500$, $n = 800$, $\sigma_x^2 = 1$ and $\sigma_z^2 = 0.3$.

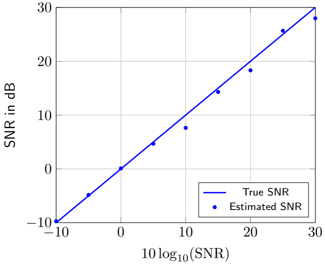
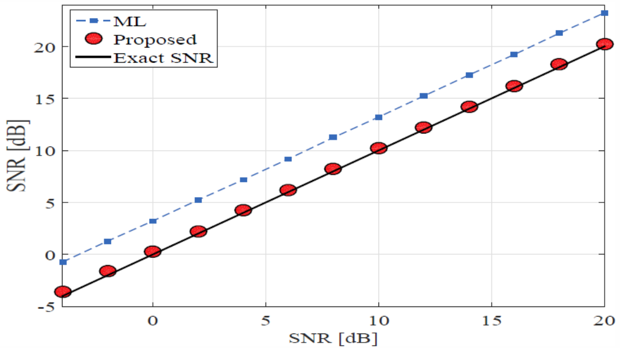


Figure: $m = 500$, $n = 800$ and $\sigma_x^2 = 1$.

- Use one observation $\mathbf{y}$ to estimate the SNR.
- Use the SNR estimate to set the regularization parameter λ and the box threshold t.

SNR Estimation under Correlation



(a) var(ML)=0.2358, var(proposed) = 0.1348.

Equalization Performance: Uncertain Channel Case

$$\mathbf{y} = (\sqrt{1-\epsilon^2}\mathbf{A} + \epsilon\boldsymbol{\Delta})\mathbf{x}_0 + \mathbf{z}$$

where

- $\epsilon \in [0, 1)$.
- $\boldsymbol{\Delta}$ is the estimation noise matrix with *iid* Gaussian entries with var. σ_δ^2.
- $\epsilon = 0$:
  $\mathrm{BER}_{\mathrm{RLS}} = Q\left(\sqrt{\frac{\delta - \frac{1}{(1+\Upsilon(\lambda,\delta))^2}}{\left(\frac{\Upsilon(\lambda,\delta)}{1+\Upsilon(\lambda,\delta)}\right)^2 + \frac{1}{\mathsf{snr}}}}\right)$
- $\epsilon \neq 0$:

$$\mathrm{BER}_{\mathrm{RLS}} = Q\left(\sqrt{\frac{\delta - \frac{1}{(1+\Upsilon(\lambda,\delta))^2}}{(1-\epsilon^2)\left(\frac{\Upsilon(\lambda,\delta)}{1+\Upsilon(\lambda,\delta)}\right)^2 + \frac{1}{\mathsf{snr}} + \epsilon^2}}\right),$$

where $\Upsilon(\lambda, \delta) = \frac{1-\delta+\lambda+\sqrt{(1-\delta+\lambda)^2+4\lambda\delta}}{2\delta}$.

Equalization Performance: Uncertain Channel Case

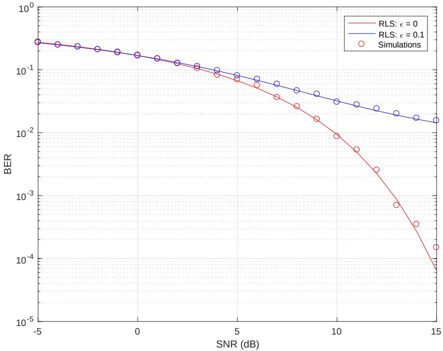
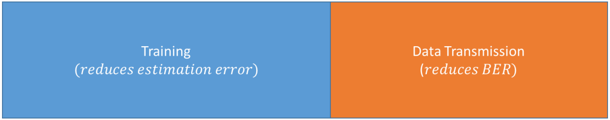


Figure: BER performance $\delta = 1.3, n = 256$

[4] Ayed M. Alrashdi, Ismail Ben Atitallah, Tareq Y. Al-Naffouri and Mohamed-Slim Alouini "Precise Performance Analysis of the LASSO Under Matrix Uncertainties", GlobalSIP, 2017.

Optimum Training

- Given a power budget at the transmitter
  - We can use some for channel estimation (reduces ϵ).
  - We can use some for data transmission (reduces BER).



- Total Energy

$$\begin{aligned} E &= E_p + E_d \\ &= \alpha E + (1-\alpha) E \end{aligned}$$

- What is the optimum trade-off?

How to find α?

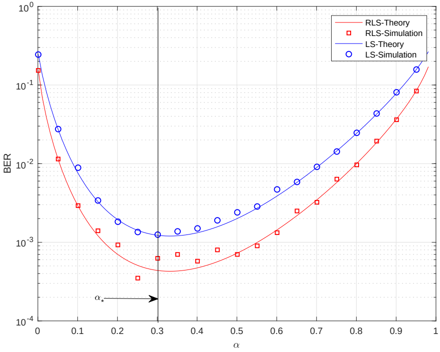


Figure: Optimal Power vs. BER for the LS and RLS equalizers.

[5] Ayed M. Alrashdi, Ismail Ben Atitallah, Tarig Ballal, Christos Thrampoulidis, Anas Chaaban and Tareq Y. Al-Naffouri "Optimum Training for MIMO BPSK Transmission", SPAWC, 2018 (submitted).

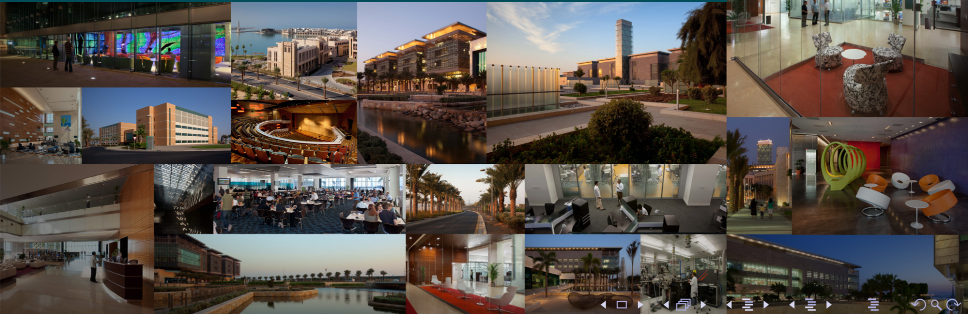
Conclusion of Part II

- Precise Asymptotic BER analysis of the Box Relaxation Optimization for BPSK signal recovery, that allow efficient optimal tuning of the paramters.
- Tuning is possible even if SNR is not known as we are able to estimate it precisely.
- Analysis is extended to the case where channel exhibits uncertainty.
- Analysis used to find optimize training power to minimize SNR.
- Future work: We are extending the work to other constellations (PAM, QAM), other equalizers, and correlated channel case.

Thank you

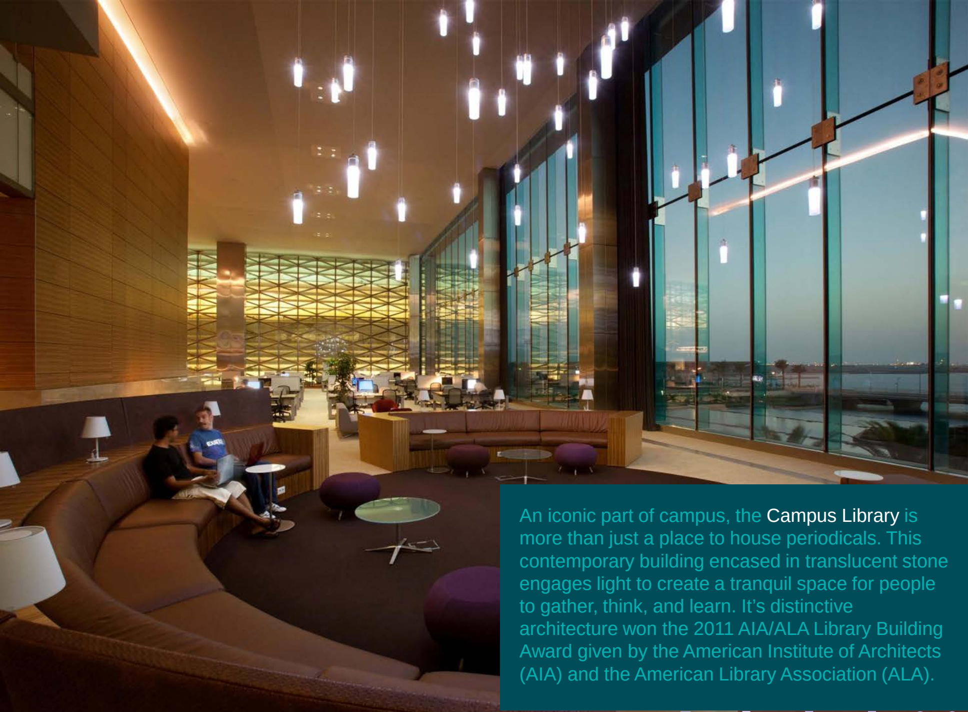
Through Inspiration, Discovery

King Abdullah University of Science and Technology



What is KAUST?

- Graduate Level research university governed by an independent Board of Trustees
- Merit based, open to all from around the world
- Research Centers as primary organizational units
- Research funding and collaborative educational programs
- Collaborative research projects, linking industry R&D and economic development
- Environmentally responsible campus



An iconic part of campus, the Campus Library is more than just a place to house periodicals. This contemporary building encased in translucent stone engages light to create a tranquil space for people to gather, think, and learn. It’s distinctive architecture won the 2011 AIA/ALA Library Building Award given by the American Institute of Architects (AIA) and the American Library Association (ALA).

The Student Center is a one-stop spot for many student-related services to support academic, personal, and professional development.