Constrained Perturbation Regularization Approach for Signal Estimation Using Random Matrix Theory

جامعة الملك عبدالله للعلوم والتقنية
King Abdullah University of Science and Technology

Mohamed Suliman, Tarig Ballal, Abla Kammoun, and Tareq Y. Al-Naffouri

King Abdullah University of Science and Technology (KAUST)

Emails:{mohamed.suliman, tarig.ahmed, abla.kammoun, tareq.alnaffouri}@kaust.edu.sa

1. Abstract

- This work proposes a new regularization approach for linear least-squares problems with random Gaussian matrices.
- The proposed approach is based on forcing an artificial perturbation matrix with a bounded norm into the linear model matrix to enhance the singular-value (SV) structure of the matrix and hence the solution of the estimation problem.
- Relying on the randomness of the model matrix, a number of tools from random matrix theory are applied to derive the near-optimum regularizer that minimizes the mean-squared error of the estimator.
- Simulation results demonstrate that the proposed approach outperforms a set of benchmark regularization methods for various estimated signal characteristics. In addition, simulations show that our approach is robust in the presence of model uncertainty.

2. Problem Statement

- Consider the linear system

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{z}. \quad (1)$$

  - $\mathbf{H} \in \mathbb{C}^{N\times M}$ is the linear transformation matrix. (Known)
  - $\mathbf{y} \in \mathbb{C}^{N\times 1}$ is the observation vector. (Known)
  - $\mathbf{x} \in \mathbb{C}^{M\times 1}$ is the desired signal (Unknown):
    - $\mathbf{R_x} \triangleq \mathbb{E}\left(\mathbf{x}\mathbf{x}^H\right)$ if $\mathbf{x}$ is random. (Unknown)
    - $\mathbf{R_x} \triangleq \mathbf{x}\mathbf{x}^H$ if $\mathbf{x}$ is deterministic.
  - $\mathbf{z} \in \mathbb{C}^{N\times 1}$ is the noise vector that has i.i.d. entries with zero mean with variance $\sigma_\mathbf{z}^2$. (Unknown)
  - $\mathbf{z}$ and $\mathbf{x}$ are independent.

Assumption 1 *Let $\mathbf{H} \in \mathbb{C}^{N\times M}$ have i.i.d. entries with $H_{ij} \sim \mathcal{CN}(0,1)$, and let $\mathbf{R_x}$ be a deterministic uniformly bounded real matrix of size $M \times M$.*

Assumption 2 *Consider the linear asymptotic regime in which the problem dimensions N and M grow proportionally to infinity with $\rho = N/M \in (0,\infty)$.*

Problem Given $\mathbf{y}$ and $\mathbf{H}$, find an estimate of $\mathbf{x}$.

- The simplest way to estimate $\mathbf{x}$ is by using the least-squares (LS) estimation

$$\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2. \quad (2)$$

- LS Issues:
  - Solution is potentially very sensitive to perturbations in the data
  - In many cases, LS is completely unreliable.
- Alternatives: Use regularization
- The most common and well-known form of regularization is the Tikhonov regularization

$$\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + \eta||\mathbf{x}||_2^2. \quad (3)$$

- The solution of (3) is given by

$$\hat{\mathbf{x}}_{\text{RLS}} = (\mathbf{H}^T\mathbf{H} + \gamma\mathbf{I})^{-1}\mathbf{H}^T\mathbf{y}, \quad (4)$$

  where $\gamma = \eta$.
- Algorithms to find γ ? There are many methods:
  - Generalized cross validation (GCV).
  - L-curve.
  - Quasi-optimal.

3. Constrained Perturbation Regularization Approach (COPRA)

- As a form of regularization, we allow a perturbation $\Delta\mathbf{H}$ into $\mathbf{H}$.
- This perturbation is aimed to improve the eigenvalue/singular-value (SV) structure of the matrix $\mathbf{H}$.
- In order to maintain the balance between improving the SV and maintaining the fidelity of the basic model in (1), we add the constraint $||\Delta\mathbf{H}||_2 \le \lambda$, $\lambda \in \mathbb{R}^+$.
- As a result, the model in (1) is modified to

$$\mathbf{y} \approx (\mathbf{H} + \Delta\mathbf{H})\mathbf{x} + \mathbf{z}. \quad (5)$$

Question How to choose $\Delta\mathbf{H}$ and λ ?

- Assuming that we know the *best* choice of λ, we consider minimizing the worst-case residual function of (5)

$$\min_{\hat{\mathbf{x}}} \max_{\Delta\mathbf{H}} ||\mathbf{y} - (\mathbf{H} + \Delta\mathbf{H})\hat{\mathbf{x}}||_2$$
$$\text{subject to: } ||\Delta\mathbf{H}||_2 \le \lambda. \quad (6)$$

- It can be shown that the solution to (6) is given by the RLS

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

  where γ is obtained by solving the following equation:

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

Problem The solution requires knowledge of λ, which we do not know.

- By taking the expected value of (8) we can manipulate to obtain

$$\lambda_o^2\mathbb{E}\left[\sigma_\mathbf{z}^2\frac{1}{N^2\tilde{\gamma}_o^2}\text{Tr}\left(\left(\frac{1}{N\tilde{\gamma}_o}\mathbf{H}^H\mathbf{H} + \mathbf{I}\right)^{-2}\right)\right] + \lambda_o^2\mathbb{E}\left[\frac{1}{N\tilde{\gamma}_o^2}\text{Tr}\left(\mathbf{R_x}\left(\frac{1}{N\tilde{\gamma}_o}\mathbf{H}^H\mathbf{H} + \mathbf{I}\right)^{-2}\frac{\mathbf{H}^H\mathbf{H}}{N}\right)\right]$$
$$= \mathbb{E}\left[\sigma_\mathbf{z}^2\frac{1}{N\tilde{\gamma}_o^2}\text{Tr}\left(\left(\frac{1}{N\tilde{\gamma}_o}\mathbf{H}^H\mathbf{H} + \mathbf{I}\right)^{-2}\frac{\mathbf{H}^H\mathbf{H}}{N}\right)\right] + \mathbb{E}\left[\text{Tr}\left(\mathbf{H}^H\mathbf{H}\left(\mathbf{H}^H\mathbf{H} + N\tilde{\gamma}_o\mathbf{I}\right)^{-2}\mathbf{H}^H\mathbf{H}\mathbf{R_x}\right)\right], \quad (9)$$

  where Tr$\{.\}$ denotes the trace operator and $\gamma_o \triangleq N\tilde{\gamma}_o$.

Theorem 1 *Under the settings of Assumptions 1 and 2, the optimal perturbation bound λ_o is given by*

$$\lambda_o^2 \approx \frac{\tilde{\delta}_o N\left(\delta_o^2\tilde{\delta}_o(\tilde{\delta}_o Tr(\mathbf{R_x}) + \rho\left(\sigma_\mathbf{z}^2 - \tilde{\gamma}_o Tr(\mathbf{R_x})\right))\right)}{\delta_o\tilde{\delta}_o Tr(\mathbf{R_x})(\delta_o\tilde{\delta}_o - \tilde{\gamma}_o) - \delta_o\tilde{\gamma}_o\rho\sigma_\mathbf{z}^2} + \frac{\tilde{\delta}_o N\left(\delta_o\tilde{\gamma}_o\rho\left(\tilde{\gamma}_o Tr(\mathbf{R_x}) - \sigma_\mathbf{z}^2\right) - \tilde{\gamma}_o^2\rho\, Tr(\mathbf{R_x})\right)}{\delta_o\tilde{\delta}_o Tr(\mathbf{R_x})(\delta_o\tilde{\delta}_o - \tilde{\gamma}_o) - \delta_o\tilde{\gamma}_o\rho\sigma_\mathbf{z}^2}. \quad (10)$$

- where:

$$\delta_o, \tilde{\delta}_o = \frac{1}{2}\left(\tilde{\gamma}_o\left(\sqrt{\tilde{\gamma}_o^{-2}\left((\tilde{\gamma}_o+1)^2 + 2(\tilde{\gamma}_o - 1)\rho + \rho^2\right)} - 1\right) \pm \rho \mp 1\right). \quad (11)$$

Problem λ_o dependents on $\sigma_\mathbf{z}^2$ and $\mathbf{R_x}$ which are not known.

- We propose applying the MSE criterion to eliminate this dependency and to set λ_o that minimizes the MSE approximately.

4. Minimizing the MSE

- The MSE for an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}$ can be defined as

$$\text{MSE} = \text{Tr}\left\{\mathbb{E}\left((\hat{\mathbf{x}} - \mathbf{x})(\hat{\mathbf{x}} - \mathbf{x})^T\right)\right\}. \quad (12)$$

- We can manipulate the MSE to the form:

$$\text{MSE} = \mathbb{E}\left[\sigma_\mathbf{z}^2\,\text{Tr}\left(\mathbf{H}^H\mathbf{H}\left(\mathbf{H}^H\mathbf{H} + \gamma\mathbf{I}\right)^{-2}\right) + \gamma^2\,\text{Tr}\left(\left(\mathbf{H}^H\mathbf{H} + \gamma\mathbf{I}\right)^{-2}\mathbf{R_x}\right)\right]. \quad (13)$$

- **Theorem 2** *Under the settings of Assumptions 1 and 2, and by defining $\gamma \triangleq N\tilde{\gamma}$, the DE of the MSE function in (13) can be obtained as*

$$MSE \approx \frac{\delta^2\left(\tilde{\delta}\rho\left(\tilde{\gamma} - \delta\tilde{\delta}\right)\sigma_\mathbf{z}^2 + \tilde{\gamma}^3\, Tr(\mathbf{R_x})\right)}{\tilde{\gamma}\rho\left(\tilde{\gamma}^2\rho - \delta^2\tilde{\delta}^2\right)}. \quad (14)$$

  where δ and $\tilde{\delta}$ are given by (11) when $\tilde{\gamma}_o = \tilde{\gamma}$.
- By taking the derivative of (14) w.r.t γ, then equating the result to zero we obtain

$$\tilde{\gamma}_o \approx \frac{\rho\,\sigma_\mathbf{z}^2}{\text{Tr}\left(\mathbf{R_x}\right)}. \quad (15)$$

- Substitute this results in (10) and then substitute the result in (8) we obtain COPRA characteristic equation:

$$S\left(\tilde{\gamma}_o\right) = \text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2 + N\tilde{\gamma}_o\mathbf{I}\right)^{-2}\mathbf{b}\mathbf{b}^H\right)\left(\delta_o^2\tilde{\delta}_o^2 - \tilde{\gamma}_o{}^2\delta_o - \tilde{\gamma}_o\delta_o\tilde{\delta}_o\right) + \text{Tr}\left(\left(\mathbf{\Sigma}^2 + N\tilde{\gamma}_o\mathbf{I}\right)^{-2}\mathbf{b}\mathbf{b}^H\right)\times$$
$$\left(N\delta_o\tilde{\delta}_o\left(\tilde{\gamma}_o{}^2 - \tilde{\gamma}_o\delta_o\tilde{\delta}_o - \delta_o\tilde{\delta}_o^2\right) + M\tilde{\delta}_o\tilde{\gamma}_o\left(\tilde{\gamma}_o - \tilde{\gamma}_o\delta_o + \delta_o^2\tilde{\delta}_o\right)\right) = 0. \quad (16)$$

  where $\mathbf{b} \triangleq \mathbf{U}^T\mathbf{y}$ and $\mathbf{H} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$ is the SVD of $\mathbf{H}$.

- Solving (16) provides the near-optimal regularization parameter $\tilde{\gamma}_o$ that minimizes the MSE of the estimator.

5. Simulation Results

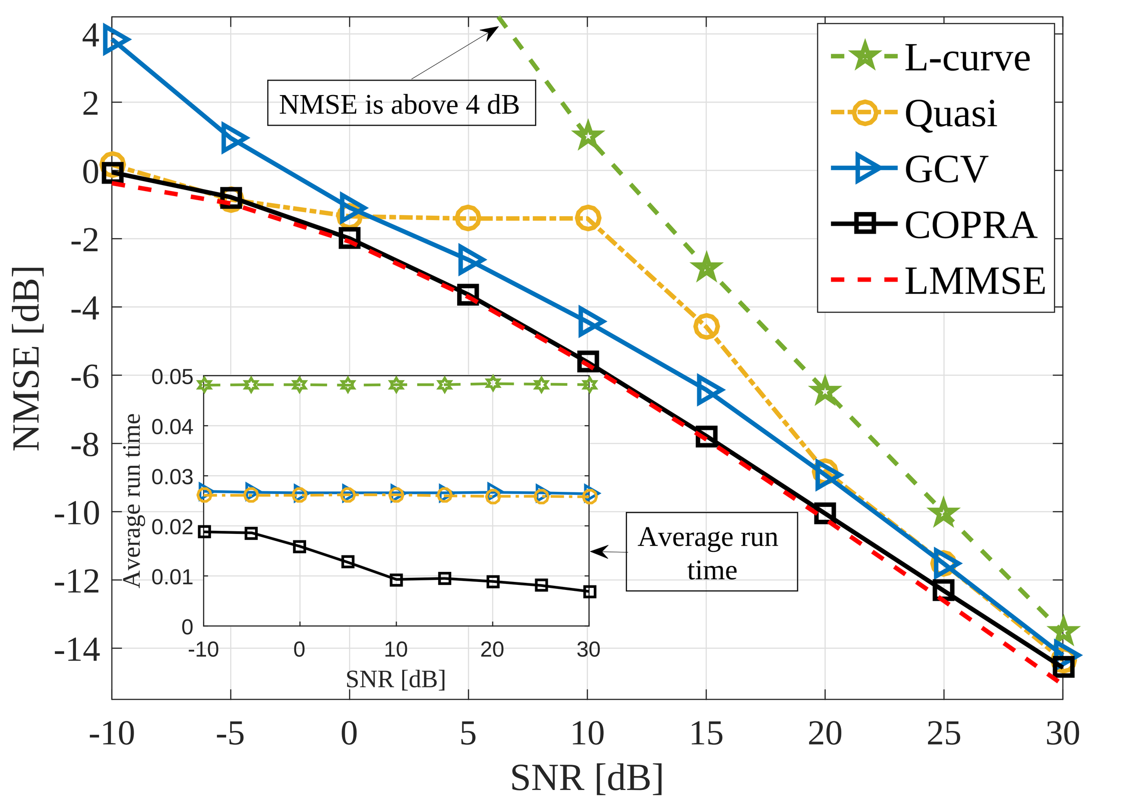

Figure 1: *Performance comparison with perfect $\mathbf{H}$ and $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ with i.i.d. elements.*

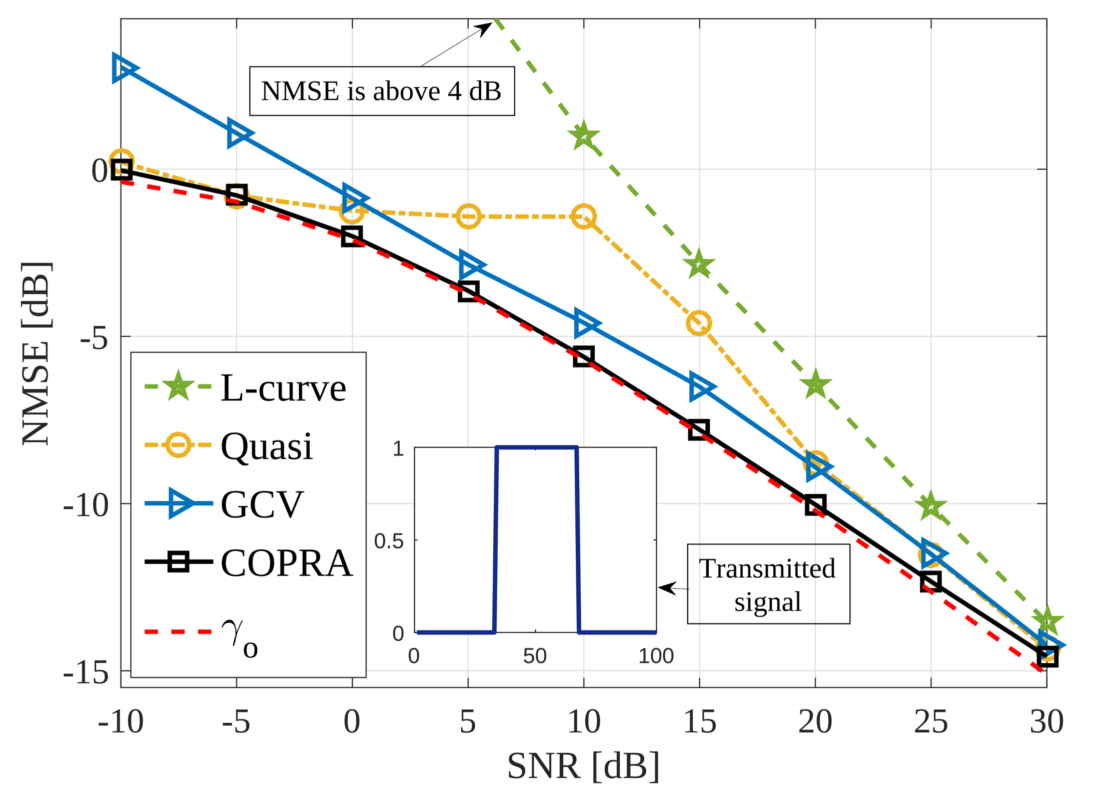

Figure 2: *Performance comparison with perfect $\mathbf{H}$ and $\mathbf{x}$ is a deterministic square pulse signal.*

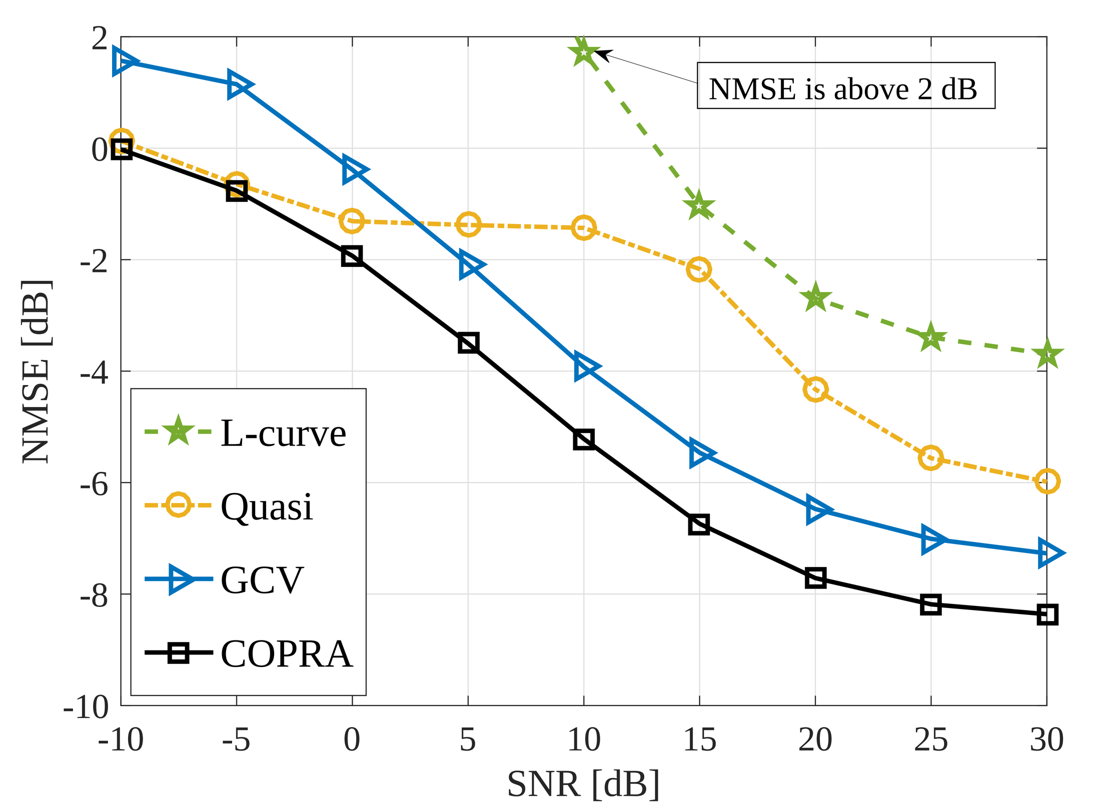

Figure 3: *Performance comparison with uncertainty in $\mathbf{H}$ i.e., $\hat{\mathbf{H}} = \mathbf{H} - \epsilon\mathbf{\Omega}$ where $\mathbf{H}$ is the true unknown model matrix; $\hat{\mathbf{H}}$ is the known estimated matrix; and $\mathbf{\Omega}$ is the model error matrix, which is independent of $\hat{\mathbf{H}}$ and has i.i.d. entries with $\Omega_{ij} \sim \mathcal{CN}(0,1)$.*

6. Conclusions

A new regularization approach for a linear LS estimation is proposed. The algorithm is shown to outperform several benchmark methods with low computational complexity and also to be robust in the presence of model uncertainty.