Accepted Manuscript

Perturbation-Based Regularization for Signal Estimation in Linear Discrete Ill-Posed Problems

Mohamed A. Suliman, Tarig Ballal, Tareq Y. Al-Naffouri

PII: S0165-1684(18)30165-8
DOI: 10.1016/j.sigpro.2018.05.005
Reference: SIGPRO 6813

To appear in: *Signal Processing*

Received date: 12 October 2017
Revised date: 26 March 2018
Accepted date: 6 May 2018

Please cite this article as: Mohamed A. Suliman, Tarig Ballal, Tareq Y. Al-Naffouri, Perturbation-Based Regularization for Signal Estimation in Linear Discrete Ill-Posed Problems, *Signal Processing* (2018), doi: 10.1016/j.sigpro.2018.05.005

This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

ACCEPTED MANUSCRIPT

Highlights

- *New regularization approach*: We proposed a new approach for linear discrete ill-posed problems based on adding an artificial perturbation matrix with a bounded norm to the model matrix A. The objective of this artificial perturbation is to improve the singularvalue structure of A. This perturbation affects the fidelity of the model $\mathrm{y} = \mathrm{A}\mathrm{x}_0 + \mathrm{z}$, and as a result, the equality relation becomes invalid. We show that using such modification provides a solution with better numerical stability.
- *New regularization parameter selection method*: We develop a new regularization parameter selection approach that selects the regularizer in a way that minimizes the mean-squared error (MSE) between x_0 and its estimate $\hat{\mathrm{x}}$, i.e., $\mathbb{E}\|\hat{\mathrm{x}} - \mathrm{x}_0\|_2^2$.
- *Generality*: A key feature of the approach is that it does not impose any prior assumptions on x_0. The vector x_0 can be deterministic or stochastic, and in the later case we do not assume any prior statistical knowledge. In addition, knowledge of the noise variance σ_{z}^2 is not required. This makes the proposed approach applicable to a large number of linear discrete ill-posed problems.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Perturbation-Based Regularization for Signal Estimation in Linear Discrete Ill-Posed Problems

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

Abstract

Estimating the values of unknown parameters in ill-posed problems from corrupted measured data presents formidable challenges in ill-posed problems. In such problems, many of the fundamental estimation methods fail to provide meaningful stabilized solutions. In this work, we propose a new regularization approach combined with a new regularization-parameter selection method for linear least-squares discrete ill-posed problems called constrained perturbation regularization approach (COPRA). The proposed COPRA is based on perturbing the singular-value structure of the linear model matrix to enhance the stability of the problem solution. Unlike many regularization methods that seek to minimize the estimated data error, the proposed approach is developed to minimize the mean-squared error of the estimator, which is the objective in many estimation scenarios. The performance of the proposed approach is demonstrated by applying it to a large set of real-world discrete ill-posed problems. Simulation results show that the proposed approach outperforms a set of benchmark regularization methods in most cases. In addition, the approach enjoys the shortest runtime and offers the highest level of robustness of all the tested benchmark regularization methods.

Index Terms

Linear estimation, ill-posed problems, linear least squares, regularization, perturbed models.

The research reported in this publication was supported by funding from King Abdullah University of Science and Technology (KAUST), under award number# 2221 from the Office of Competitive Research and Grant OSR-2016-KKI-2899.

M. A. Suliman, T. Ballal, and T. Y. Al-Naffouri are with King Abdullah University of Science and Technology (KAUST), Computer, Electrical, and Mathematical Sciences and Engineering (CEMSE) Division, Thuwal, 23955-6900, Saudi Arabia. E-mails:{mohamed.suliman, tarig.ahmed, tareq.alnaffouri}@kaust.edu.sa.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

I. Introduction

We consider the standard problem of recovering an unknown signal $\mathbf{x}_0 \in \mathbb{R}^n$ from a vector $\mathbf{y} \in \mathbb{R}^m$ of noisy, linear observations given by

$$\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ is a known linear-model matrix, and $\mathbf{z} \in \mathbb{R}^{m\times 1}$ is a vector of additive white Gaussian noise (AWGN) with unknown variance $\sigma_{\mathbf{z}}^2$ that is independent of $\mathbf{x}_0$. This problem has been extensively studied because of its practical and theoretical importance in many fields of science and engineering, e.g., communication, signal processing, computer vision, control theory, and economics [1]–[3].

Over the past years, several mathematical tools have been developed for estimating the unknown vector $\mathbf{x}_0$. The most prominent approach is the ordinary least-squares (OLS) estimator [4], which finds an estimate $\hat{\mathbf{x}}_{\text{OLS}}$ of $\mathbf{x}_0$ by minimizing the Euclidean norm of the estimator residual error, i.e.,

$$\hat{\mathbf{x}}_{\text{OLS}} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2. \tag{2}$$

If $\mathbf{A}$ is a full column rank matrix, then (2) has the unique solution

$$\hat{\mathbf{x}}_{\text{OLS}} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{y} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{y}, \tag{3}$$

where $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T = \sum_{i=1}^{n}\sigma_i\mathbf{u}_i\mathbf{v}_i^T$ is the singular value decomposition (SVD) of $\mathbf{A}$; $\mathbf{u}_i$ and $\mathbf{v}_i$ are the left and right orthogonal singular vectors, respectively, and the singular values σ_i are assumed to satisfy $\sigma_1 \geq \sigma_1 \geq \cdots \geq \sigma_n$.

Despite being a popular approach, the OLS estimator suffers when it is applied to discrete ill-posed problems. A problem is considered well-posed when its solution always exists, is unique, and depends continuously on the initial data. Ill-posed problems fail to satisfy at least one of these conditions [5]. The matrix $\mathbf{A}$ of an ill-posed problem is ill-conditioned and the computed OLS solution in (3) is potentially very sensitive to perturbations in the data such as $\mathbf{z}$ [6].

Discrete ill-posed problems arise in a variety of applications in signal processing and computer vision [7]–[10], computerized tomography [11], astronomy [12], image restoration and deblurring [13], [14], and edge detection [15]. Interestingly, in all these applications, the data are gathered by convoluting a noisy signal with a detector [16], [17]. A linear representation of such process

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

is given by

$$\int_{b_1}^{b_2} a(s,t)\,\mathbf{x}_0(t)\,\mathrm{dt} = \mathbf{y}_0(s) + \mathbf{z}(s) = \mathbf{y}(s), \tag{4}$$

where $\mathbf{y}_0(s)$ is the true signal, and the kernel function $a(s,t)$ represents the response. It is shown in [18] how a problem with a formulation similar to (4) fails to satisfy the well-posed conditions introduced above. The discretized version of (4) can be represented by (1).

To solve ill-posed problems, regularization methods, such as truncated SVD [19], hybrid methods [20], the covariance-shaping LS estimator [21], and the weighted LS estimator [22], are commonly used. These methods are based on leveraging additional known information into the solution of the problem and replacing the ill-posed problem with a well-posed one. This replacement should be done after carefully analyzing the ill-posed problem in terms of its physical plausibility and mathematical properties.

The most common and widely used approach is the regularized M-estimator that obtains an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}_0$ by solving the convex problem

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}} \mathcal{L}(\mathbf{y} - \mathbf{A}\mathbf{x}) + \gamma f(\mathbf{x}), \tag{5}$$

where the loss function $\mathcal{L} : \mathbb{R}^m \to \mathbb{R}$ measures the fit of $\mathbf{Ax}$ to the observation vector $\mathbf{y}$, the penalty function $f : \mathbb{R}^m \to \mathbb{R}$ establishes the structure of $\mathbf{x}$, and γ provides a balance between the two functions. Different choices of $\mathcal{L}$ and f distinguish the different estimation techniques. The most popular technique is the Tikhonov regularization [23] which is given in its simplified form by

$$\hat{\mathbf{x}}_{\text{RLS}} := \arg\min_{\mathbf{x}} \ ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 + \gamma \ ||\mathbf{x}||_2^2. \tag{6}$$

The solution to (6) is given by the regularized least-square (RLS) estimator,

$$\hat{\mathbf{x}}_{\text{RLS}} = \left(\mathbf{A}^T\mathbf{A} + \gamma \mathbf{I}_{\text{n}}\right)^{-1} \mathbf{A}^T \mathbf{y}, \tag{7}$$

where $\mathbf{I}_{\text{n}}$ is an $n \times n$ identity matrix. In general, γ is unknown and must be chosen judiciously.

Several methods have been proposed to select the value of the regularization parameter γ. These include the generalized cross validation (GCV) [24], L-curve [25], [26], and quasi-optimal method [27]. A survey of regularization parameter selection methods is given in [28]. The GCV method obtains the regularizer γ by minimizing the GCV function, which suffers from a very flat minimum that is challenging to locate numerically. The L-curve method, on the other hand, is a graphical tool with a very high computational complexity. Finally, the quasi-optimal method

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

does not take noise level into account. In general, the performance of these methods varies significantly depending on the nature of the problem.

A. Paper contributions

The contributions of this paper can be summarized as follows:

1) *New regularization approach*: We propose a new approach for linear discrete ill-posed problems that is based on adding an artificial perturbation matrix with a bounded norm to the model matrix $\mathbf{A}$. The objective of this artificial perturbation is to improve the singular-value structure of $\mathbf{A}$. This perturbation affects the fidelity of the model $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$; as a result, the equality relation becomes invalid. We show that using such a modification provides a solution with better numerical stability.[1]
2) *New regularization-parameter selection method*: We develop a new approach for selecting a regularization parameter that minimizes the mean-squared error (MSE) between $\mathbf{x}_0$ and its estimate $\hat{\mathbf{x}}$, i.e., $\mathbb{E}\ ||\hat{\mathbf{x}} - \mathbf{x}_0||_2^2$. [2]
3) *Generality*: A key feature of the approach is that it does not impose any prior assumptions on $\mathbf{x}_0$. The vector $\mathbf{x}_0$ can be deterministic or stochastic and, in the later case, we do not assume any prior statistical knowledge. In addition, knowledge of the noise variance $\sigma_{\mathbf{z}}^2$ is not required. This makes the proposed approach applicable to a large number of linear discrete ill-posed problems.

B. Paper organization

This paper is organized as follows. In Section II, we present the formulation of the problem and derive the solution. In Section III, we derive the artificial perturbation bound that minimizes the MSE. Further, we derive the characteristic equation of the proposed approach which is used to obtain the regularization parameter. In Section IV, we study the properties of the characteristic equation, and in Section V we present the performance of the proposed approach based on simulation results. Finally, concluding remarks are given in Section VI.

[1]The work presented in this paper is an extended version of [29].

[2]Little work on MSE-based estimators is available in the literature; for example, in [30] the authors derived an estimator for the linear model problem that was based on minimizing the *worst-case* MSE (as opposed to the actual MSE) while imposing a constraint on the unknown vector $\mathbf{x}_0$.

C. Notations

Matrices are denoted by boldface uppercase letters (e.g., $\mathbf{X}$). Column vectors are represented by boldface lowercase letters (e.g., $\mathbf{x}$). The notation $(\cdot)^T$ denotes the transpose operator, $\mathbb{E}(\cdot)$ denotes the expectation operator, while $\mathbf{I}_\mathrm{n}$ and $\mathbf{0}$ denote the $(n \times n)$ identity matrix and the zero matrix, respectively. The notation $||\cdot||_2$ indicates the spectral norm for matrices and the Euclidean norm for vectors. The operator diag $(\cdot)$ returns a vector that contains either the diagonal elements of a matrix, or a diagonal matrix if it operates on a vector where the diagonal entries of the matrix are the elements of that vector.

II. Proposed Regularization Approach

A. Background

We consider the linear discrete ill-posed problem in (1) without imposing any assumptions on $\mathbf{x}_0$. As stated above, matrix $\mathbf{A}$ is ill-conditioned and may have a very fast singular-value decay [31]. In Fig. 1, we observe that the singular values of matrix $\mathbf{A}$ decay very fast, though without a sharp transition, towards markedly small singular values.

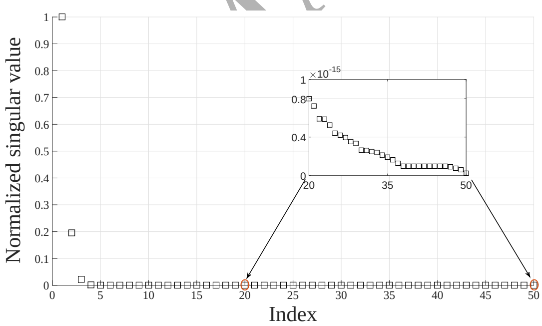


Fig. 1: Singular-value decay pattern of an ill-posed matrix, $\mathbf{A} \in \mathbb{R}^{50\times 50}$.

B. Problem formulation

We start by considering the OLS solution in (3). Due to the singular-value structure of matrices in ill-posed problems, and the interaction that they have with the noise, (3) cannot produce a sensible estimate of $\mathbf{x}_0$. Herein, we propose adding an artificial perturbation $\mathbf{\Delta_A} \in \mathbb{R}^{m\times n}$ to $\mathbf{A}$. We will show later that adding $\mathbf{\Delta_A}$ to $\mathbf{A}$ tends to provide a regularized solution in the form $\hat{\mathbf{x}} = \left(\mathbf{A}^T\mathbf{A} + \rho\left(\delta, \hat{\mathbf{x}}\right)\mathbf{I}_\mathrm{n}\right)^{-1}\mathbf{A}^T\mathbf{y}$ as opposed to the OLS solution $\hat{\mathbf{x}}_\mathrm{LS} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{y}$.

ACCEPTED MANUSCRIPT

Therefore, perturbing the model (1) is equivalent to enhancing the singular values of the ill-posed matrix $\mathbf{A}^T\mathbf{A}$ by adding a regularization matrix given by $\rho\left(\delta, \hat{\mathbf{x}}\right)\mathbf{I}_{\text{n}}$. We assume that this process, which replaces $\mathbf{A}$ with $(\mathbf{A} + \boldsymbol{\Delta}_{\mathbf{A}})$, improves the singular-value structure of $\mathbf{A}$ and the estimate of $\mathbf{x}_0$. In other words, we assume that using $(\mathbf{A} + \boldsymbol{\Delta}_{\mathbf{A}})$ to estimate $\mathbf{x}_0$ from $\mathbf{y}$ can provide more accurate estimation results than using $\mathbf{A}$. To strike a balance between improving the singular-value structure and maintaining the fidelity of the basic linear model, we add the constraint $||\boldsymbol{\Delta}_{\mathbf{A}}||_2 \leq \delta$, $\delta \in \mathbb{R}^+$.

The linear model in (1), modified according to the discussion above, can be written as

$$\mathbf{y} \approx \left(\mathbf{A} + \boldsymbol{\Delta}_{\mathbf{A}}\right)\mathbf{x}_0 + \mathbf{z};\ \ ||\boldsymbol{\Delta}_{\mathbf{A}}||_2 \leq \delta. \tag{8}$$

The model in (8) has been considered for signal estimation in the presence of data errors but with strict equality (e.g., [30], [32], [33]). These studies assumed that $\mathbf{A}$ was not known perfectly due to some error contamination, but that prior knowledge about the real error bound (which corresponds to δ in our case) was available. However, in our case matrix $\mathbf{A}$ is known perfectly, whereas δ is unknown.

The question now is what is the best $\boldsymbol{\Delta}_{\mathbf{A}}$ and the bound on the norm of this perturbation δ. It is clear that these values are important since they affect the model fidelity and dictate the quality of the estimator. This question is addressed further ahead in this section. For now, we start by assuming that δ is known. We use this assumption to obtain and estimate of $\mathbf{x}_0$ based on (8) that is a function of δ; then, we address the problem of obtaining the value of δ.

To obtain an estimate of $\mathbf{x}_0$, we consider minimizing the worst-case residual function of the new perturbed model in (8), which is given by

$$\min_{\mathbf{x}} \max_{||\boldsymbol{\Delta}_{\mathbf{A}}||_2 \leq \delta} Q\left(\mathbf{x}, \boldsymbol{\Delta}_{\mathbf{A}}\right) :=\ ||\mathbf{y} - \left(\mathbf{A} + \boldsymbol{\Delta}_{\mathbf{A}}\right)\mathbf{x}||_2. \tag{9}$$

For each choice of $\mathbf{x}$ in (9), there are infinite choices of the perturbation $\boldsymbol{\Delta}_{\mathbf{A}}$ that satisfy $||\boldsymbol{\Delta}_{\mathbf{A}}||_2 \leq \delta$. For example, for $\mathbf{x} = \mathbf{x}_1$, the residual error as a function of the perturbation is given by $Q_1(\boldsymbol{\Delta}_{\mathbf{A}}) = Q(\mathbf{x}_1, \boldsymbol{\Delta}_{\mathbf{A}})$. For another choice $\mathbf{x} = \mathbf{x}_2$, we have $Q_2(\boldsymbol{\Delta}_{\mathbf{A}}) = Q(\mathbf{x}_2, \boldsymbol{\Delta}_{\mathbf{A}})$, and so on. Each $Q_i(\boldsymbol{\Delta}_{\mathbf{A}})$ has a maximum value $Q_{i,\text{max}}$ at a certain choice of $\boldsymbol{\Delta}_{\mathbf{A}}$ (possibly for multiple choices of $\boldsymbol{\Delta}_{\mathbf{A}}$). In (9), we choose an estimate $\mathbf{x}_i$ that corresponds to the smallest $Q_{i,\text{max}}$ because it explains the data best when a worst-case bounded perturbation is applied.

ACCEPTED MANUSCRIPT

Theorem 1. *The unique minimizer* $\hat{\mathbf{x}}$ *for (9) when* $\delta > 0$ *is given by*

$$\hat{\mathbf{x}} = \left(\mathbf{A}^T\mathbf{A} + \rho\left(\delta, \hat{\mathbf{x}}\right)\mathbf{I}_n\right)^{-1}\mathbf{A}^T\mathbf{y}, \tag{10}$$

where $\rho\left(\delta, \hat{\mathbf{x}}\right)$ *is a regularization parameter related to the perturbation bound* δ *by*

$$\rho\left(\delta, \hat{\mathbf{x}}\right) = \delta\ \frac{||\mathbf{y} - \mathbf{A}\hat{\mathbf{x}}||_2}{||\hat{\mathbf{x}}||_2}. \tag{11}$$

Proof: By using the Minkowski inequality [34], we find the upper bound of the cost function $Q\left(\mathbf{x}, \mathbf{\Delta_A}\right)$ in (9) as

$$\begin{aligned} ||\mathbf{y} - \left(\mathbf{A} + \mathbf{\Delta_A}\right)\mathbf{x}||_2 &\le ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2 + ||\mathbf{\Delta_A}\ \mathbf{x}||_2 \le ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2 + ||\mathbf{\Delta_A}||_2||\mathbf{x}||_2 \\ &\le ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2 + \delta\ ||\mathbf{x}||_2. \end{aligned} \tag{12}$$

However, upon setting $\mathbf{\Delta_A}$ to be the rank-one matrix

$$\mathbf{\Delta_A} = \frac{\left(\mathbf{A}\mathbf{x} - \mathbf{y}\right)}{||\mathbf{y} - \mathbf{A}\mathbf{x}||_2}\frac{\mathbf{x}^T}{||\mathbf{x}||_2}\delta, \tag{13}$$

we show that the bound in (12) is achievable by

$$\begin{aligned} ||\mathbf{y} - \left(\mathbf{A} + \mathbf{\Delta_A}\right)\mathbf{x}||_2 &= ||\left(\mathbf{y} - \mathbf{A}\mathbf{x}\right) + \frac{\left(\mathbf{y} - \mathbf{A}\mathbf{x}\right)}{||\mathbf{y} - \mathbf{A}\mathbf{x}||_2}\frac{\mathbf{x}^T}{||\mathbf{x}||_2}\mathbf{x}\delta||_2 \\ &= ||\left(\mathbf{y} - \mathbf{A}\mathbf{x}\right) + \frac{\left(\mathbf{y} - \mathbf{A}\mathbf{x}\right)}{||\mathbf{y} - \mathbf{A}\mathbf{x}||_2}||\mathbf{x}||_2\delta||_2. \end{aligned} \tag{14}$$

Since the two added vectors $\left(\mathbf{y} - \mathbf{A}\mathbf{x}\right)$ and $\frac{\left(\mathbf{y} - \mathbf{A}\mathbf{x}\right)}{||\mathbf{y} - \mathbf{A}\mathbf{x}||_2}||\mathbf{x}||_2\delta$ in (14) are positively linearly dependent (i.e., pointing in the same direction), we conclude that

$$||\left(\mathbf{y} - \mathbf{A}\mathbf{x}\right) + \frac{\left(\mathbf{y} - \mathbf{A}\mathbf{x}\right)}{||\mathbf{y} - \mathbf{A}\mathbf{x}||_2}||\mathbf{x}||_2\ \delta||_2 = \underbrace{||\mathbf{y} - \mathbf{A}\mathbf{x}||_2 + \delta||\mathbf{x}||_2}_{W(\mathbf{x})}. \tag{15}$$

As a result, (9) can be expressed equivalently by

$$\min_{\mathbf{x}}\ \max_{||\mathbf{\Delta_A}||_2 \le \delta}\ Q\left(\mathbf{x}, \mathbf{\Delta_A}\right) \equiv \min_{\mathbf{x}}\ W\left(\mathbf{x}\right). \tag{16}$$

It is easy to check that the solution space for $W\left(\mathbf{x}\right)$ is convex in $\mathbf{x}$, and hence, any local minimum is also a global minimum. But at any local minimum, it either holds that the gradient of $W\left(\mathbf{x}\right)$ is zero, or $W\left(\mathbf{x}\right)$ is not differentiable. More precisely, $W\left(\mathbf{x}\right)$ is not differentiable only when $\mathbf{x} = \mathbf{0}$ or $\mathbf{y} - \mathbf{A}\mathbf{x} = \mathbf{0}$. However, we do not consider the trivial case of $\mathbf{x} = \mathbf{0}$, and $\mathbf{y} - \mathbf{A}\mathbf{x} = \mathbf{0}$ is impossible by definition. Therefore, we can obtain the gradient of $W\left(\mathbf{x}\right)$ as

$$\nabla_{\mathbf{x}}W\left(\mathbf{x}\right) = \frac{\mathbf{A}^T\left(\mathbf{A}\mathbf{x} - \mathbf{y}\right)}{||\mathbf{y} - \mathbf{A}\mathbf{x}||_2} + \frac{\delta\ \mathbf{x}}{||\mathbf{x}||_2} = \frac{1}{||\mathbf{y} - \mathbf{A}\mathbf{x}||_2}\left(\mathbf{A}^T\mathbf{A}\mathbf{x} + \frac{\delta\ ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2\ \mathbf{x}}{||\mathbf{x}||_2} - \mathbf{A}^T\mathbf{y}\right). \tag{17}$$

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Defining $\rho(\delta, \hat{\mathbf{x}})$ as in (11), we solve for $\nabla_{\mathbf{x}} W(\hat{\mathbf{x}}) = 0$ to obtain (10). ■

Remark 1. It can be said that perturbing $\mathbf{A}$ allows us to move from the LS estimator that has zero residual norm (fits the observations perfectly), to a class of estimators that fit the observations loosely (have less respect for the observations), which is the basic idea of regularization. On the other hand, the linear minimum mean-squared error (LMMSE) [4] does not try to fit the observations model by definition. Instead, it tries to minimize the difference between $\hat{\mathbf{x}}$ and $\mathbf{x}_0$.

Remark 2. The regularization parameter ρ in (11) is a function of the unknown estimate $\hat{\mathbf{x}}$ and of the perturbation upper bound δ (we drop the dependence of ρ on δ and $\hat{\mathbf{x}}$ in the notation to simplify it). In addition, it is clear from (15) that δ controls the weight given to the side-constraint minimization relative to the residual-norm minimization. We have assumed that δ is known in order to obtain the min-max optimization solution in (9). However, this assumption is not valid in reality. Thus, is it impossible to obtain ρ directly from (11) given that both δ and $\hat{\mathbf{x}}$ are unknown.

Now, it is obvious with (10) and (11) in hand, we can eliminate the dependency of ρ on $\hat{\mathbf{x}}$. By substituting (10) into (11) and performing some manipulations, we obtain

$$\delta^2 \Big[\mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\mathbf{A}\left(\mathbf{A}^T\mathbf{A} + \rho\mathbf{I}_{\mathrm{n}}\right)^{-1}\mathbf{A}^T\mathbf{y} + ||\mathbf{A}\left(\mathbf{A}^T\mathbf{A} + \rho\mathbf{I}_{\mathrm{n}}\right)^{-1}\mathbf{A}^T\mathbf{y}||^2\Big]$$
$$= \rho^2\mathbf{y}^T\mathbf{A}\left(\mathbf{A}^T\mathbf{A} + \rho\mathbf{I}_{\mathrm{n}}\right)^{-2}\mathbf{A}^T\mathbf{y}. \tag{18}$$

In the next subsection, we utilize (18) to obtain the δ that corresponds to an optimal choice of ρ.

C. Finding the optimal perturbation bound

The optimal ρ and δ that minimize the MSE are denoted by ρ_{o} and δ_{o}, respectively. To simplify (18), we substitute the SVD of $\mathbf{A}$ and solve for δ^2. Next, we take the trace $\mathrm{Tr}(\cdot)$ of the two sides considering the evaluation point to be $(\delta_{\mathrm{o}}, \rho_{\mathrm{o}})$. This results in

$$\underbrace{\delta_{\mathrm{o}}^2\ \mathrm{Tr}\left(\left(\mathbf{\Sigma}^2 + \rho_{\mathrm{o}}\mathbf{I}_{\mathrm{n}}\right)^{-2}\mathbf{U}^T\left(\mathbf{y}\mathbf{y}^T\right)\mathbf{U}\right)}_{D(\rho_{\mathrm{o}})} = \underbrace{\mathrm{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2 + \rho_{\mathrm{o}}\mathbf{I}_{\mathrm{n}}\right)^{-2}\mathbf{U}^T\left(\mathbf{y}\mathbf{y}^T\right)\mathbf{U}\right)}_{N(\rho_{\mathrm{o}})}. \tag{19}$$

In order to obtain a useful expression, we think of δ_{o} as a single universal value that is computed over many realizations of the observation vector $\mathbf{y}$. Based on this perception, $\mathbf{y}\mathbf{y}^T$ can be replaced

ACCEPTED MANUSCRIPT

by its expected value, $\mathbb{E}(\mathbf{y}\mathbf{y}^T)$. In other words, we look for δ_o that is optimal (in the MSE sense) for all realizations of $\mathbf{y}$. We assume that such a value exists. Then, the parameter δ_o is clearly of a deterministic nature. Taking the expected value of both sides of (19) for a fixed δ_o is equivalent to replacing $\mathbf{y}\mathbf{y}^T$ with $\mathbb{E}(\mathbf{y}\mathbf{y}^T)$, which can be expressed using (1) as

$$\mathbb{E}\left(\mathbf{y}\mathbf{y}^T\right) = \mathbf{A}\mathbf{R}_{\mathbf{x}_0}\mathbf{A}^T + \sigma_\mathbf{z}^2\mathbf{I}_\mathrm{m} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}\boldsymbol{\Sigma}\mathbf{U}^T + \sigma_\mathbf{z}^2\mathbf{I}_\mathrm{m}, \tag{20}$$

where $\mathbf{R}_{\mathbf{x}_0} \triangleq \mathbb{E}\left(\mathbf{x}_0\mathbf{x}_0^T\right)$. For a deterministic $\mathbf{x}_0$, $\mathbf{R}_{\mathbf{x}_0} = \mathbf{x}_0\mathbf{x}_0^T$ (for simplicity, we use $\mathbf{R}_{\mathbf{x}_0}$ in both the deterministic and stochastic cases). Substituting (20) for $\mathbf{y}\mathbf{y}^T$ in (19) results in

$$N\left(\rho_\mathrm{o}\right) = \mathrm{Tr}\left(\boldsymbol{\Sigma}^2\left(\boldsymbol{\Sigma}^2 + \rho_\mathrm{o}\mathbf{I}_\mathrm{n}\right)^{-2}\boldsymbol{\Sigma}^2\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}\right) + \sigma_\mathbf{z}^2\mathrm{Tr}\left(\boldsymbol{\Sigma}^2\left(\boldsymbol{\Sigma}^2 + \rho_\mathrm{o}\mathbf{I}_\mathrm{n}\right)^{-2}\right), \tag{21}$$

and

$$D\left(\rho_\mathrm{o}\right) = \delta_\mathrm{o}^2\Big[\mathrm{Tr}\left(\left(\boldsymbol{\Sigma}^2 + \rho_\mathrm{o}\mathbf{I}_\mathrm{n}\right)^{-2}\boldsymbol{\Sigma}^2\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}\right) + \sigma_\mathbf{z}^2\mathrm{Tr}\left(\left(\boldsymbol{\Sigma}^2 + \rho_\mathrm{o}\mathbf{I}_\mathrm{n}\right)^{-2}\right)\Big]. \tag{22}$$

Considering the singular-value structure of the ill-posed problems in the general case, we can divide the singular values into two groups of *significant*, or relatively large, and *trivial*, or nearly zero, singular values.[3] For example, we see in Fig. 1 that the singular values of the ill-posed model matrix $\mathbf{A}$ decay very fast, making it possible to distinguish the significant and trivial groups. Thus, the matrix $\boldsymbol{\Sigma}$ can be divided into two diagonal sub-matrices: $\boldsymbol{\Sigma}_{n1}$, which contains the significant n_1 diagonal entries; and $\boldsymbol{\Sigma}_{n2}$, which contains the trivial $n_2 = n - n_1$ diagonal entries.[4] Therefore, $\boldsymbol{\Sigma}$ can be written as

$$\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{n1} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Sigma}_{n2} \end{bmatrix}. \tag{23}$$

Similarly, we partition $\mathbf{V}$ as $\mathbf{V} = [\mathbf{V}_{n1}\ \mathbf{V}_{n2}]$, where $\mathbf{V}_{n1} \in \mathbb{R}^{n\times n_1}$ and $\mathbf{V}_{n2} \in \mathbb{R}^{n\times n_2}$. Now, we can rewrite (21) in terms of the partitioned $\boldsymbol{\Sigma}$ and $\mathbf{V}$ as

$$\begin{aligned} N\left(\rho_\mathrm{o}\right) = {} & \mathrm{Tr}\left(\boldsymbol{\Sigma}_{n1}^2\left(\boldsymbol{\Sigma}_{n1}^2 + \rho_\mathrm{o}\mathbf{I}_{n1}\right)^{-2}\boldsymbol{\Sigma}_{n1}^2\mathbf{V}_{n1}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}_{n1}\right) \\ & + \mathrm{Tr}\left(\boldsymbol{\Sigma}_{n2}^2\left(\boldsymbol{\Sigma}_{n2}^2 + \rho_\mathrm{o}\mathbf{I}_{n2}\right)^{-2}\boldsymbol{\Sigma}_{n2}^2\mathbf{V}_{n2}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}_{n2}\right) \\ & + \sigma_\mathbf{z}^2\mathrm{Tr}\left(\boldsymbol{\Sigma}_{n1}^2\left(\boldsymbol{\Sigma}_{n1}^2 + \rho_\mathrm{o}\mathbf{I}_{n1}\right)^{-2}\right) + \sigma_\mathbf{z}^2\mathrm{Tr}\left(\boldsymbol{\Sigma}_{n2}^2\left(\boldsymbol{\Sigma}_{n2}^2 + \rho_\mathrm{o}\mathbf{I}_{n2}\right)^{-2}\right). \end{aligned} \tag{24}$$

[3] This includes the special case when all the singular values are significant, and so all are considered.

[4] To identify the two singular-value clusters, simple thresholding can be applied, e.g., using a threshold obtained by multiplying the average of the singular values by a constant c, where $\mathrm{c} \in (0, 1)$.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Given that $\mathbf{\Sigma}_{n1}$ contains the significant singular values and $\mathbf{\Sigma}_{n2}$ contains the nearly zero singular values, we have $\|\mathbf{\Sigma}_{n2}\| \approx 0$. So, we can approximate $N(\rho_{\rm o})$ by

$$N(\rho_{\rm o}) \approx \text{Tr}\left(\mathbf{\Sigma}_{n1}^2 \left(\mathbf{\Sigma}_{n1}^2 + \rho_{\rm o}\mathbf{I}_{n1}\right)^{-2} \mathbf{\Sigma}_{n1}^2 \mathbf{V}_{n1}^T \mathbf{R}_{\mathbf{x}_0} \mathbf{V}_{n1}\right) + \sigma_{\mathbf{z}}^2 \text{Tr}\left(\mathbf{\Sigma}_{n1}^2 \left(\mathbf{\Sigma}_{n1}^2 + \rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\right). \quad (25)$$

Similarly, $D(\rho_{\rm o})$ in (22) can be approximated as

$$D(\rho_{\rm o}) \approx \sigma_{\mathbf{z}}^2 \ \text{Tr}\left(\left(\mathbf{\Sigma}_{n1}^2 + \rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\right) + \frac{n_2\sigma_{\mathbf{z}}^2}{\rho_{\rm o}^2} + \text{Tr}\left(\left(\mathbf{\Sigma}_{n1}^2 + \rho_{\rm o}\mathbf{I}_{n1}\right)^{-2} \mathbf{\Sigma}_{n1}^2 \mathbf{V}_{n1}^T \mathbf{R}_{\mathbf{x}_0} \mathbf{V}_{n1}\right). \quad (26)$$

By substituting (25) and (26) into (19) and performing manipulations, we obtain

$$\delta_{\rm o}^2 \approx \left[\sigma_{\mathbf{z}}^2 \text{Tr}\left(\mathbf{\Sigma}_{n1}^2 \left(\mathbf{\Sigma}_{n1}^2 + \rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\right) + \text{Tr}\left(\mathbf{\Sigma}_{n1}^2 \left(\mathbf{\Sigma}_{n1}^2 + \rho_{\rm o}\mathbf{I}_{n1}\right)^{-2} \mathbf{\Sigma}_{n1}^2 \mathbf{V}_{n1}^T \mathbf{R}_{\mathbf{x}_0} \mathbf{V}_{n1}\right)\right] \Big/$$
$$\left[\sigma_{\mathbf{z}}^2 \text{Tr}\left(\left(\mathbf{\Sigma}_{n1}^2 + \rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\right) + \frac{n_2\sigma_{\mathbf{z}}^2}{\rho_{\rm o}^2} + \text{Tr}\left(\left(\mathbf{\Sigma}_{n1}^2 + \rho_{\rm o}\mathbf{I}_{n1}\right)^{-2} \mathbf{\Sigma}_{n1}^2 \mathbf{V}_{n1}^T \mathbf{R}_{\mathbf{x}_0} \mathbf{V}_{n1}\right)\right]. \quad (27)$$

The bound $\delta_{\rm o}$ given by (27) is a function of the unknown quantities $\rho_{\rm o}$, $\sigma_{\mathbf{z}}^2$, and $\mathbf{R}_{\mathbf{x}_0}$. Estimating $\sigma_{\mathbf{z}}^2$ and $\mathbf{R}_{\mathbf{x}_0}$ without any prior knowledge is a very tedious process. The problem is worse when $\mathbf{x}_0$ is deterministic. In such case, the exact value of $\mathbf{x}_0$ is required to obtain $\mathbf{R}_{\mathbf{x}_0} = \mathbf{x}_0\mathbf{x}_0^T$. In the following section, we use the MSE criterion to eliminate the dependence of $\delta_{\rm o}$ on these unknown quantities, and obtain the final expression of the optimal perturbation bound $\delta_{\rm o}$.

III. Minimizing the MSE for the solution of the proposed approach

The MSE for an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}_0$ is given by

$$\text{MSE} = \mathbb{E}\left[\|\hat{\mathbf{x}} - \mathbf{x}_0\|^2\right] = \text{Tr}\left(\mathbb{E}\left((\hat{\mathbf{x}} - \mathbf{x}_0)(\hat{\mathbf{x}} - \mathbf{x}_0)^T\right)\right). \quad (28)$$

For the proposed approach, the signal estimation is given by (10). Hence, we substitute (10) for $\hat{\mathbf{x}}$ in (28) and use the SVD of $\mathbf{A}$ to obtain

$$\text{MSE}(\rho) = \sigma_{\mathbf{z}}^2 \text{Tr}\left(\mathbf{\Sigma}^2 \left(\mathbf{\Sigma}^2 + \rho\mathbf{I}_{\rm n}\right)^{-2}\right) + \rho^2 \text{Tr}\left(\left(\mathbf{\Sigma}^2 + \rho\mathbf{I}_{\rm n}\right)^{-2} \mathbf{V}^T \mathbf{R}_{\mathbf{x}_0} \mathbf{V}\right). \quad (29)$$

Theorem 2. *For $\sigma_{\mathbf{z}}^2 > 0$, an approximate value of the optimal regularizer ρ_o that approximately minimizes the MSE in (29) is given by*

$$\rho_o \approx \frac{\sigma_{\mathbf{z}}^2}{Tr\left(\mathbf{R}_{\mathbf{x}_0}\right)/n}. \quad (30)$$

Proof: The global minimizer of the function in (29) (i.e., $\rho_{\rm o}$) can be obtained by differentiating (29) with respect to ρ and setting the result to equal zero, i.e.,

$$\nabla_\rho \ \text{MSE}(\rho) = -2\sigma_{\mathbf{z}}^2 \text{Tr}\left(\mathbf{\Sigma}^2 \left(\mathbf{\Sigma}^2 + \rho\mathbf{I}_{\rm n}\right)^{-3}\right) + 2\rho \underbrace{\text{Tr}\left(\mathbf{\Sigma}^2 \left(\mathbf{\Sigma}^2 + \rho\mathbf{I}_{\rm n}\right)^{-3} \mathbf{V}^T \mathbf{R}_{\mathbf{x}_0} \mathbf{V}\right)}_{S} = 0. \quad (31)$$

ACCEPTED MANUSCRIPT

Equation (31) dictates the relationship between the optimal regularization parameter ρ_{o} and the parameters of (1). By solving (31), we obtain the optimal regularizer ρ_{o}. However, with the lacking of knowledge on $\mathbf{R}_{\mathbf{x}_0}$, we cannot obtain a closed-form expression for ρ_{o} in the general case. Instead, we seek to obtain a suboptimal regularizer that approximately minimizes (29). In what follows, we show how through some bounds and approximations, we can obtain this suboptimal regularizer.

By using the trace inequalities in [35, Eq.(5)], we bound the second term in (31) by

$$\lambda_{\text{min}}\left(\mathbf{R}_{\mathbf{x}_0}\right)\text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho\mathbf{I}_{\text{n}}\right)^{-3}\right)\leq S=\text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho\mathbf{I}_{\text{n}}\right)^{-3}\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}\right)$$
$$\leq\lambda_{\text{max}}\left(\mathbf{R}_{\mathbf{x}_0}\right)\text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho\mathbf{I}_{\text{n}}\right)^{-3}\right),\tag{32}$$

where λ_{min} and λ_{max} are the smallest and the largest singular values, respectively. Our main goal in this paper is to find a solution that is approximately feasible for all discrete ill-posed problems and also suboptimal in some sense. In other words, we would like to find a ρ_{o} for all (or almost all) possible $\mathbf{A}$ that approximately minimizes the MSE. To achieve this, we consider the *average* value of S in (31) based on the inequalities in (32) as our evaluation point:

$$S\approx\text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho\mathbf{I}_{\text{n}}\right)^{-3}\right)\frac{\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)}{n}.\tag{33}$$

Substituting (33) in (31) yields

$$\nabla_{\rho}\ \text{MSE}\left(\rho\right)\approx-2\sigma_{\mathbf{z}}^2\text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho\mathbf{I}_{\text{n}}\right)^{-3}\right)+2\rho\frac{\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)}{n}\text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho\mathbf{I}_{\text{n}}\right)^{-3}\right)=0.\tag{34}$$

Note that the same approximation can be applied from the beginning to the second term in (29) and the same result in (34) will be obtained after taking the derivative of the new approximated MSE function. In Appendix A, we provide an error analysis for this approximation and show that, in most cases, the error bound is sufficiently small to consider the approximation feasible. In the case where the elements of $\mathbf{x}_0$ are independent and identically distributed (i.i.d.), it can be shown that (34) and (31) are exactly equivalent to each other (see Appendix A). By solving (34), we obtain ρ_{o} as in (30). ■

Remark 3. The result in (30) shows that there always exists a positive ρ_{o} for $\sigma_{\mathbf{z}}^2\neq 0$ that approximately minimizes the MSE. The fact that the regularization parameter ρ_{o} is generally dependent on the noise variance has been shown before in different contexts (e.g., [36], [37]).

ACCEPTED MANUSCRIPT

For the special case where the entries of $\mathbf{x}_0$ are i.i.d. with zero mean, we have $\mathbf{R}_{\mathbf{x}_0} = \sigma^2_{\mathbf{x}_0}\mathbf{I}_{\text{n}}$. Since the optimal LMMSE estimator of $\mathbf{x}_0$ is defined as [4]

$$\hat{\mathbf{x}}_{\text{LMMSE}} = \left(\mathbf{A}^T\mathbf{A} + \sigma^2_{\mathbf{z}}\mathbf{R}^{-1}_{\mathbf{x}_0}\mathbf{I}_{\text{n}}\right)^{-1}\mathbf{A}^T\mathbf{y}, \tag{35}$$

substituting $\mathbf{R}_{\mathbf{x}_0} = \sigma^2_{\mathbf{x}_0}\mathbf{I}$ makes the value of the parameter multiplying the identity matrix in the LMMSE expression exactly equivalent to ρ_{o} in (30), since $\rho_{\text{o}} = \frac{\sigma^2_{\mathbf{z}}}{\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)/n} = \frac{\sigma^2_{\mathbf{z}}}{\sigma^2_{\mathbf{x}_0}}$. In this case, and with the presence of prior information about the distributions of $\mathbf{x}_0$ and $\mathbf{z}$, the value of ρ_{o} can be estimated as in [38] under certain assumptions about the distribution of $\mathbf{A}$. This shows that (30) is exact when the input is white; while for a general input $\mathbf{x}_0$, the optimum matrix regularizer is given in (35). In other words, the result in (30) provides an approximate optimum scalar regularizer for a general colored input. Note that since $\sigma^2_{\mathbf{z}}$ and $\mathbf{R}_{\mathbf{x}_0}$ are both unknowns, ρ_{o} cannot be obtained directly from (30).

We are now ready to eliminate the dependency of δ_{o} (27) on the unknowns $\sigma^2_{\mathbf{z}}$ and $\mathbf{R}_{\mathbf{x}_0}$ using the result in (30) and the perturbation-bound expression in (27).

A. Setting the optimal perturbation bound that minimizes the MSE

By applying the same reasoning used to obtain (33) to both the numerator and the denominator of (27) and performing some manipulations, we obtain

$$\begin{aligned}&\text{Tr}\left(\mathbf{\Sigma}^2_{n1}\left(\mathbf{\Sigma}^2_{n1} + \rho_{\text{o}}\mathbf{I}_{n1}\right)^{-2}\left(\mathbf{\Sigma}^2_{n1} + \frac{n_1\sigma^2_{\mathbf{z}}}{\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)}\mathbf{I}_{n1}\right)\right)\\ &\approx \delta^2_{\text{o}}\left[\text{Tr}\left(\left(\mathbf{\Sigma}^2_{n1} + \rho_{\text{o}}\mathbf{I}_{n1}\right)^{-2}\left(\mathbf{\Sigma}^2_{n1} + \frac{n_1\sigma^2_{\mathbf{z}}}{\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)}\mathbf{I}_{n1}\right)\right) + \frac{n_2 n_1\sigma^2_{\mathbf{z}}}{\rho^2_{\text{o}}\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)}\right].\end{aligned} \tag{36}$$

We evaluate the accuracy of this approximation using simulations in Section V.

Now, we use the relationship of $\sigma^2_{\mathbf{z}}$ and $\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)$ to the suboptimal regularizer, $\frac{n_1\sigma^2_{\mathbf{z}}}{\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)} \approx \frac{n_1}{n}\rho_{\text{o}}$, obtained from (30) to impose a constraint on (36) that makes the selected perturbation bound minimize the MSE and also makes (36) an implicit equation in δ_{o} and ρ_{o} only. By substituting $\frac{n_1\sigma^2_{\mathbf{z}}}{\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)} = \frac{n_1}{n}\rho_{\text{o}}$ in (36) and performing manipulations, we obtain

$$\delta^2_{\text{o}} \approx \frac{\text{Tr}\left(\mathbf{\Sigma}^2_{n1}\left(\mathbf{\Sigma}^2_{n1} + \rho_{\text{o}}\mathbf{I}_{n1}\right)^{-2}\left(\frac{n}{n_1}\mathbf{\Sigma}^2_{n1} + \rho_{\text{o}}\mathbf{I}_{n1}\right)\right)}{\text{Tr}\left(\left(\mathbf{\Sigma}^2_{n1} + \rho_{\text{o}}\mathbf{I}_{n1}\right)^{-2}\left(\frac{n}{n_1}\mathbf{\Sigma}^2_{n1} + \rho_{\text{o}}\mathbf{I}_{n1}\right)\right) + \frac{n_2}{\rho_{\text{o}}}}. \tag{37}$$

ACCEPTED MANUSCRIPT

The expression in (37) represents the δ_o that approximately minimizes the MSE. Now, we have two equations, (18) (evaluated at δ_0 and ρ_0) and (37), and two unknowns δ_0 and ρ_0. By substituting the SVD of $\mathbf{A}$ in (18) and solving simultaneously with (37), we reach the characteristic equation of our proposed constrained perturbation regularization approach (COPRA) as

$$\begin{aligned} G(\rho_\mathrm{o}) = &\operatorname{Tr}\left(\boldsymbol{\Sigma}^2\left(\boldsymbol{\Sigma}^2+\rho_\mathrm{o}\mathbf{I}_\mathrm{n}\right)^{-2}\mathbf{U}^T\mathbf{y}\mathbf{y}^T\mathbf{U}\right)\operatorname{Tr}\left(\left(\boldsymbol{\Sigma}_{n1}^2+\rho_\mathrm{o}\mathbf{I}_{n1}\right)^{-2}\left(\beta\boldsymbol{\Sigma}_{n1}^2+\rho_\mathrm{o}\mathbf{I}_{n1}\right)\right) \\ &+\frac{n_2}{\rho_\mathrm{o}}\operatorname{Tr}\left(\boldsymbol{\Sigma}^2\left(\boldsymbol{\Sigma}^2+\rho_\mathrm{o}\mathbf{I}_\mathrm{n}\right)^{-2}\mathbf{U}^T\mathbf{y}\mathbf{y}^T\mathbf{U}\right) \\ &-\operatorname{Tr}\left(\left(\boldsymbol{\Sigma}^2+\rho_\mathrm{o}\mathbf{I}_\mathrm{n}\right)^{-2}\mathbf{U}^T\mathbf{y}\mathbf{y}^T\mathbf{U}\right)\operatorname{Tr}\left(\boldsymbol{\Sigma}_{n1}^2\left(\boldsymbol{\Sigma}_{n1}^2+\rho_\mathrm{o}\mathbf{I}_{n1}\right)^{-2}\left(\beta\boldsymbol{\Sigma}_{n1}^2+\rho_\mathrm{o}\mathbf{I}_{n1}\right)\right)=0, \quad (38) \end{aligned}$$

where $\beta = \frac{n}{n_1}$.

For simplicity, the first two terms in (38) are denoted by $G_1(\rho_\mathrm{o})$ and the last term in (38) is denoted by $G_2(\rho_\mathrm{o})$. The COPRA equation (38) is a function of the model matrix $\mathbf{A}$, the received signal $\mathbf{y}$, and the regularization parameter ρ_o, which is the only unknown in (38). Solving for $G(\rho_\mathrm{o}) = 0$ should lead to the regularization parameter ρ_o that approximately minimizes the MSE of the RLS estimator. Our main interest is to find a positive root $\rho_\mathrm{o}^* > 0$ for (38). In the following section, we study the main properties of this equation and investigate the existence and uniqueness of its positive root.

IV. Analysis of the function $G(\rho_\mathrm{o})$

In this section, we analyze the COPRA function $G(\rho_\mathrm{o})$ in (38) in details. We start by examining some main properties of $G(\rho_\mathrm{o})$ that are straightforward to prove.

Property 1. *$G(\rho_o)$ is continuous over the interval $(0,+\infty)$.*

Property 2. *$G(\rho_o)$ has n different discontinuities at $\rho_o = -\sigma_i^2, \forall i \in [1,n]$. However, these discontinuities are of no interest as far as COPRA is considered.*

Property 3. $\lim_{\rho_o \to 0^+} G(\rho_o) = +\infty$.

Property 4. $\lim_{\rho_o \to 0^-} G(\rho_o) = -\infty$.

Property 5. $\lim_{\rho_o \to +\infty} G(\rho_o) = 0$.

Properties 3 and 4 show clearly that $G(\rho_\mathrm{o})$ has a discontinuity at $\rho_\mathrm{o} = 0$.

Property 6. *Both functions $G_1(\rho_o)$ and $G_2(\rho_o)$ in (38) are completely monotonic in the interval* $(0, +\infty)$.

Proof: According to [39] and [40], a function $F(\rho_\text{o})$ is completely monotonic if it satisfies

$$(-1)^n F^{(n)}(\rho_\text{o}) \geq 0,\ 0 < \rho_\text{o} < \infty, \forall n \in \mathbb{N}, \tag{39}$$

where $F^{(n)}(\rho_\text{o})$ is the n-th derivative of $F(\rho_\text{o})$.

By continuously differentiating $G_1(\rho_\text{o})$ and $G_2(\rho_\text{o})$, we see that both functions satisfy the monotonic condition in (39). ■

Theorem 3. *The COPRA function $G(\rho_o)$ in (38) has at most two roots in the interval* $(0, +\infty)$.

Proof: The proof of Theorem 3 will be conducted in two steps. Firstly, [41], [42] proved that any completely monotonic function can be approximated by a sum of exponential functions. That is, if $F(\rho_\text{o})$ is a completely monotonic, then it can be approximated by

$$F(\rho_\text{o}) \approx \sum_{i=1}^{l} l_i e^{-k_i \rho_\text{o}}, \tag{40}$$

where l is the number of the terms in the sum, and l_i and k_i are two constants. There always exists a best uniform approximation of $F(\rho_\text{o})$ and the error in this approximation gets smaller as we increase the number of the terms l. However, rather than finding the best number of the terms or the unknown parameters l_i and k_i, our main concern here is the relation given by (40). To conclude, both functions $G_1(\rho_\text{o})$ and $G_2(\rho_\text{o})$ in (38) can be approximated by a sum of exponential functions.

Secondly, [43] showed that a sum of exponential functions has at most two intersections with the abscissa. Consequently, noting that the relation in (38) can be expressed as a sum of exponential functions, the function $G(\rho_\text{o})$ has at most two roots in the interval $(0, +\infty)$. ■

Theorem 4. *There exists a sufficiently small positive value ϵ, such that $\epsilon \ll \sigma_i^2$, $\forall i \in [1, n]$ where $G(\epsilon) = 0$ (i.e., ϵ is a positive root for (38)). However, we are not interested in this root in the proposed COPRA.*

Proof: The proof of Theorem 4 is in Appendix B. ■

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Theorem 5. *A sufficient condition for the function $G(\rho_o)$ to approach zero from a positive direction at $\rho_o = +\infty$ is given by*

$$nTr\left(\mathbf{\Sigma}^2\mathbf{b}\mathbf{b}^T\right) > Tr\left(\mathbf{\Sigma}_{n1}^2\right) Tr\left(\mathbf{b}\mathbf{b}^T\right), \tag{41}$$

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

Proof: We let $\mathbf{b} = \mathbf{U}^T\mathbf{y}$, as in (38). Given that $\mathbf{\Sigma}^2$ is a diagonal matrix, $\mathbf{\Sigma}^2 = \text{diag}\left(\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2\right)$, and from the trace function property, we can replace $\mathbf{b}\mathbf{b}^T = \mathbf{U}^T\mathbf{y}\mathbf{y}^T\mathbf{U}$ in (38) with a diagonal matrix $\mathbf{b}\mathbf{b}_d^T$ that contains $\mathbf{b}\mathbf{b}^T$ diagonal entries without affecting the result. By defining $\mathbf{b}\mathbf{b}_d^T = \text{diag}\left(b_1^2, b_2^2, \cdots, b_n^2\right)$, we write (38) as

$$G(\rho_\text{o}) = \frac{\beta}{\rho_\text{o}^4}\sum_{j=1}^{n}\frac{\sigma_j^2 b_j^2}{\left(\frac{\sigma_j^2}{\rho_\text{o}}+1\right)^2}\sum_{i=1}^{n_1}\frac{\sigma_i^2}{\left(\frac{\sigma_i^2}{\rho_\text{o}}+1\right)^2} + \frac{1}{\rho_\text{o}^3}\sum_{j=1}^{n}\frac{\sigma_j^2 b_j^2}{\left(\frac{\sigma_j^2}{\rho_\text{o}}+1\right)^2}\sum_{i=1}^{n_1}\frac{1}{\left(\frac{\sigma_i^2}{\rho_\text{o}}+1\right)^2}$$
$$-\frac{\beta}{\rho_\text{o}^4}\sum_{j=1}^{n}\frac{b_j^2}{\left(\frac{\sigma_j^2}{\rho_\text{o}}+1\right)^2}\sum_{i=1}^{n_1}\frac{\sigma_i^4}{\left(\frac{\sigma_i^2}{\rho_\text{o}}+1\right)^2} - \frac{1}{\rho_\text{o}^3}\sum_{j=1}^{n}\frac{b_j^2}{\left(\frac{\sigma_j^2}{\rho_\text{o}}+1\right)^2}\sum_{i=1}^{n_1}\frac{\sigma_i^2}{\left(\frac{\sigma_i^2}{\rho_\text{o}}+1\right)^2} + \frac{n_2}{\rho_\text{o}^3}\sum_{j=1}^{n}\frac{\sigma_j^2 b_j^2}{\left(\frac{\sigma_j^2}{\rho_\text{o}}+1\right)^2}. \tag{42}$$

Then, we use some algebraic manipulations to obtain

$$G(\rho_\text{o}) = \frac{1}{\rho_\text{o}^3}\sum_{j=1}^{n}\frac{\sigma_j^2 b_j^2}{\left(\frac{\sigma_j^2}{\rho_\text{o}}+1\right)^2}\left[\frac{\beta}{\rho_\text{o}}\sum_{i=1}^{n_1}\frac{\sigma_i^2}{\left(\frac{\sigma_i^2}{\rho_\text{o}}+1\right)^2} + \sum_{i=1}^{n_1}\frac{1}{\left(\frac{\sigma_i^2}{\rho_\text{o}}+1\right)^2} + n_2\right] - \frac{1}{\rho_\text{o}^3}\sum_{j=1}^{n}\frac{b_j^2}{\left(\frac{\sigma_j^2}{\rho_\text{o}}+1\right)^2}\times$$
$$\left[\frac{\beta}{\rho_\text{o}}\sum_{i=1}^{n_1}\frac{\sigma_i^4}{\left(\frac{\sigma_i^2}{\rho_\text{o}}+1\right)^2} + \sum_{i=1}^{n_1}\frac{\sigma_i^2}{\left(\frac{\sigma_i^2}{\rho_\text{o}}+1\right)^2}\right]. \tag{43}$$

Now, evaluating the limit of (43) as $\rho_\text{o} \to +\infty$ we obtain

$$\lim_{\rho_\text{o}\to+\infty} G(\rho_\text{o}) = \left(\lim_{\rho_\text{o}\to+\infty}\frac{1}{\rho_\text{o}^3}\right)\left[\sum_{j=1}^{n}\sigma_j^2 b_j^2\left(\tau\beta\sum_{i=1}^{n_1}\sigma_i^2 + \sum_{i=1}^{n_1}1 + n_2\right) - \sum_{j=1}^{n}b_j^2\left(\tau\beta\sum_{i=1}^{n_1}\sigma_i^4 + \sum_{i=1}^{n_1}\sigma_i^2\right)\right], \tag{44}$$

where $\tau = \lim_{\rho_\text{o}\to+\infty}\frac{1}{\rho_\text{o}}$. The relation in (44) can be simplified to

$$\lim_{\rho_\text{o}\to+\infty} G(\rho_\text{o}) = \left(\lim_{\rho_\text{o}\to+\infty}\frac{1}{\rho_\text{o}^3}\right)\left[\sum_{j=1}^{n}\sigma_j^2 b_j^2\left(\tau\beta\sum_{i=1}^{n_1}\sigma_i^2 + n\right) - \sum_{j=1}^{n}b_j^2\left(\tau\beta\sum_{i=1}^{n_1}\sigma_i^4 + \sum_{i=1}^{n_1}\sigma_i^2\right)\right] \tag{45}$$

It is clear that (45) is zero; however, the direction from which (45) approaches zero depends on the sign of the term between the square brackets. For $G(\rho_\text{o})$ to approach zero from a positive direction, knowing that the terms independent of τ are the dominants, (41) must hold. ■

ACCEPTED MANUSCRIPT

Theorem 6. *If (41) is satisfied, then $G(\rho_o)$ has a unique positive root in the interval $(\epsilon, +\infty)$.*

Proof: According to Theorem 3, the function $G(\rho_{\text{o}})$ can have no root, one root, or two roots. We already proved in Theorem 4 (Appendix B) that there exists a significantly small positive root for the COPRA function at $\rho_{\text{o},1} = \epsilon$, but we are not interested in this root. Therefore, we would like to see if there exists a second root for $G(\rho_{\text{o}})$ in the interval $(\epsilon, +\infty)$.

From Property 3 and Theorem 4, we conclude that the COPRA function has a positive value before ϵ, then it switches to negative. The condition in (41) guarantees that $G(\rho_{\text{o}})$ approaches zero from a positive direction as ρ_{o} approaches $+\infty$. This means that $G(\rho_{\text{o}})$ has an extremum in the interval $(\epsilon, +\infty)$, and this extremum is actually a minimum point. If the point of the extremum is considered to be $\rho_{\text{o,m}}$, then the function increases for $\rho_{\text{o}} > \rho_{\text{o,m}}$ until it approaches the second zero crossing at $\rho_{\text{o},2}$. Since Theorem 3 states clearly that we cannot have more than two roots, we conclude that we have only one unique positive root over the interval $(\epsilon, +\infty)$ when (41) holds. ■

A. Finding the root of $G(\rho_o)$

To find the positive root of the COPRA function $G(\rho_{\text{o}})$ in (38), we use Newton's method [44]. The function $G(\rho_{\text{o}})$ is differentiable in the interval $(\epsilon, +\infty)$, and we can obtain the expression of the first derivative $G'(\rho_{\text{o}})$ easily. Newton's method can then be applied in a straightforward manner to find this root. Starting from an initial value $\rho_{\text{o}}^{n=0} > \epsilon$ that is sufficiently small, we perform the following iterations:

$$\rho_{\text{o}}^{n+1} = \rho_{\text{o}}^{n} - \frac{G(\rho_{\text{o}}^{n})}{G'(\rho_{\text{o}}^{n})}. \tag{46}$$

The iterations stop when $|G(\rho_{\text{o}}^{n+1})| < \bar{\xi}$, i.e., where $\bar{\xi}$ is a sufficiently small positive quantity.

B. Convergence

The convergence of Newton's method can be easily proved when condition (41) is satisfied. From Theorem 5, the function $G(\rho_{\text{o}})$ always has a negative value in the interval $(\epsilon, \rho_{o,2})$. It is also clear that $G(\rho_{\text{o}})$ is an increasing function in the interval $[\rho_{\text{o}}^{n=0}, \rho_{\text{o},2}]$. Thus, starting from $\rho_{\text{o}}^{n=0}$, (46) produces a consecutively increasing estimate for ρ_{o}. Convergence occurs when $G(\rho_{\text{o}}^{n}) \to 0$ and $\rho_{\text{o}}^{n+1} \to \rho_{\text{o}}^{n}$. When the condition in (41) is not satisfied, the regularization parameter ρ_{o} should be set to ϵ.

ACCEPTED MANUSCRIPT

C. *COPRA summary*

Our proposed COPRA discussed is summarized in Algorithm 1.

Algorithm 1: COPRA summary

Input : $\mathbf{y}, \mathbf{A} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T, \tilde{\xi}, \rho_{\text{o}}^{n=0}$, splitting threshold c, ϵ.

Output: Signal estimate $\hat{\mathbf{x}}$.

1 **if** *(41) is not satisfied* **then**

2 $\quad\rho_{\text{o}} = \epsilon$ and go to step 11;

3 **end**

4 Obtain n_1 as $n_1 = \max_i[\sigma_i \geq c \ \text{avg}\left(\text{diag}\left(\boldsymbol{\Sigma}\right)\right)]$, and $n_2 = n - n_1$;

5 $\boldsymbol{\Sigma}_{\text{n1}} = \boldsymbol{\Sigma}\left(1 : n_1\right), \boldsymbol{\Sigma}_{\text{n2}} = \boldsymbol{\Sigma}\left(n_1 + 1 : n\right)$ as in (23);

6 Evaluate $G\left(\rho_{\text{o}}^{n=0}\right)$ and its derivative $G^{'}\left(\rho_{\text{o}}^{n=0}\right)$ using (38);

7 **while** $|G\left(\rho_o^n\right)| > \tilde{\xi}$ **do**

8 $\quad$Solve (46) to obtain ρ_{o}^{n+1};

9 $\quad\rho_{\text{o}}^{n} = \rho_{\text{o}}^{n+1}$;

10 **end**

11 Find $\hat{\mathbf{x}}$ using (10);

V. SIMULATION RESULTS

In this section, we perform a comprehensive set of simulations to examine the performance of the proposed COPRA and compare it with other benchmark regularization methods.

We simulate three different scenarios with an additional fourth scenario presented in [45]. First, we apply the proposed COPRA to a set of nine discrete an ill-posed real-world problems that are commonly used to test the performance of regularization methods in discrete ill-posed problems. Second, we illustrate the robustness and broad applicability of COPRA by using it to estimate the signal $\mathbf{x}_0$ when $\mathbf{A}$ is a random rank deficient matrix generated as $\mathbf{A} = \frac{1}{n}\mathbf{B}\mathbf{B}^T$, where $\mathbf{B}\,(m \times n, m > n)$ is a random matrix with i.i.d. zero-mean unit variance and Gaussian random entries. Finally, we restore an image in an ill-posed image tomography problem.[5,6]

[5] The MATLAB code of the COPRA is provided at http://www.mohamedasuliman.com/research.html and http://faculty.kfupm.edu.sa/ee/naffouri/publications.html and

[6] Another versions of the proposed COPRA are presented in [46], [47] using different assumptions.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

A. Real-world discrete ill-posed problems

We use the Matlab regularization toolbox [48] to generate pairs of a matrix $\mathbf{A} \in \mathbb{R}^{50\times 50}$ and a signal $\mathbf{x}_0$. The discrete ill-posed test problems provided in the toolbox are derived from discretizations of Fredholm integral equations as in (4) and they arise in many signal processing applications.[7]

Experiment setup: The performance of COPRA is compared with the regularization algorithms that provide the best performance amogst all the methods discussed in [28]. Precisely, we compare with quasi-optimal, GCV, L-curve, and LS. The performances are evaluated in terms of the normalized MSE (NMSE); that is, the MSE normalized by $\|\mathbf{x}_0\|_2^2$. Noise is added to the vector $\mathbf{A}\mathbf{x}_0$ according to the signal-to-noise-ratio (SNR) SNR $\triangleq \|\mathbf{A}\mathbf{x}_0\|_2^2/n\sigma_{\mathbf{z}}^2$ in order to generate $\mathbf{y}$, and we set $c = 1$. The performance is presented in dB as the NMSE (NMSE in dB = $10\log_{10}$ (NMSE)) versus SNR and is evaluated over 10^5 different noise realizations at each SNR value. Since some regularization methods provide a high unreliable NMSE results that hinder the good visualization of the NMSE, we set different upper thresholds for the vertical axis in the results sub-figures.

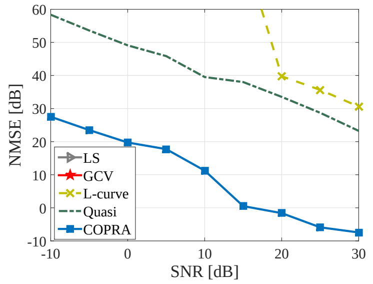
Fig. 2 shows the results from all the nine selected problems. Each sub-figure specifies the condition number (CN) of the problem's matrix. The NMSE curve disappears in certain cases indicating that these methods performed so poorly that they are out of scale. For example, the LS NMSE does not show up in any of the tested scenarios, and the NMSE curves of the quasi-optimal, GCV, and L-curve methods disappear in quite a few cases.

Generally speaking, an estimator offering NMSE > 0 dB is not sufficiently robust. From Fig. 2, it is clear that COPRA is the most robust estimator, as it is the only approach whose NMSE performance remains below 0 dB in almost all cases. Comparing NMSE over the whole range of SNR values, we find that COPRA exhibits the lowest NMSE on average in eight of the nine test problems. The closest contender to COPRA is the quasi-optimal method. However, all the methods except CPRA show lack of robustness in certain situations (i.e., high NMSE).

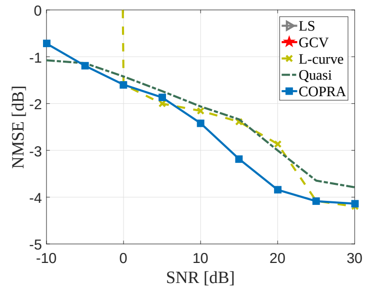
In Fig. 3, we provide the NMSE for the approximation of the perturbation-bound expression in (27) by (36) for the selected ill-posed problem matrices. The two expressions are evaluated at each SNR using the suboptimal regularizer in (30). The sub-figures show that the NMSE of the

[7]For more details regarding the test problems, we refer the reader to [48].

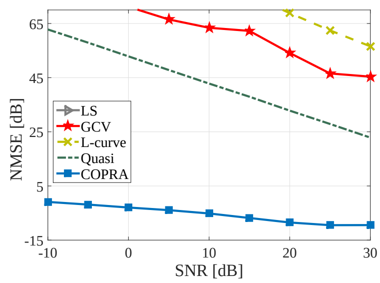
ACCEPTED MANUSCRIPT



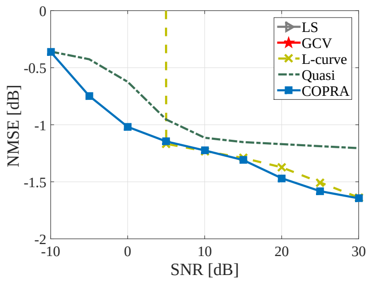
(a) Tomo (CN = 3.2 $\times 10^{16}$).



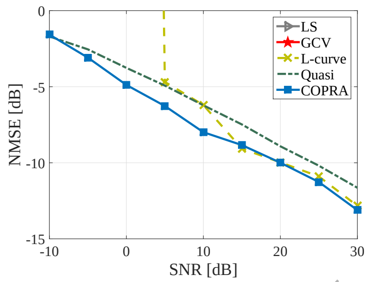
(b) Wing (CN = 1.6 $\times 10^{18}$).



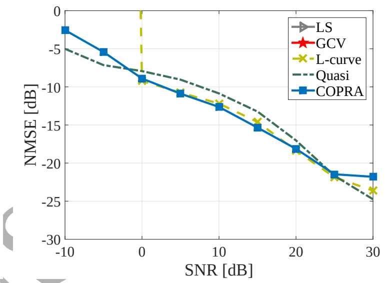
(c) Heat (CN = 2.4 $\times 10^{26}$).



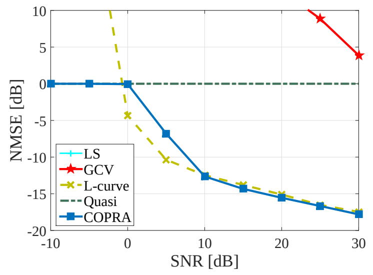
(d) Spikes (CN = 4.6 $\times 10^{18}$).



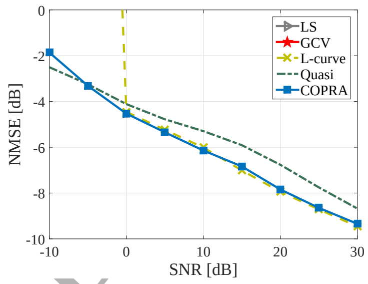
(e) Baart (CN = 4 $\times 10^{17}$).



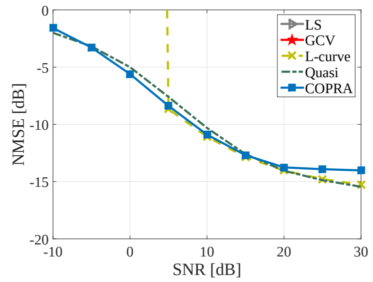
(f) Foxgood (CN = 2.4 $\times 10^{18}$).



(g) I-Laplace (CN = 3.4 $\times 10^{33}$).



(h) Deriv2 (CN = 3 $\times 10^{3}$).



(i) Shaw (CN = 2 $\times 10^{18}$).

Fig. 2: Normalized mean-squared error (NMSE) [dB] versus signal-to-noise ratio (SNR) [dB] (CN $\equiv$ condition number).

approximation is extremely small (less than -20 dB in most cases) and that the error increases as the SNR increases. The increase of the error with the SNR is discussed in Appendix A.

B. *Rank-deficient matrices*

In this scenario, a rank-deficient random matrix $\mathbf{A}$ is considered. This is the case where $\|\boldsymbol{\Sigma}_{n2}\| = 0$ in (23). This theoretical test is meant to illustrate the robustness of COPRA.

Experiment setup: A random matrix matrix $\mathbf{A}$ that satisfies $\mathbf{A} = \frac{1}{50}\mathbf{B}\mathbf{B}^T$ is generated, where

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

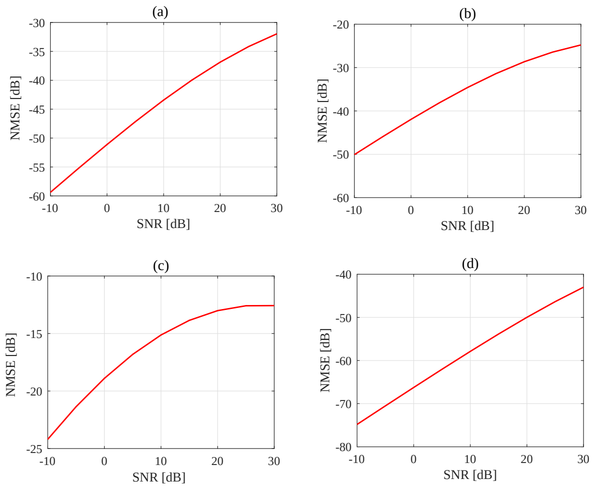
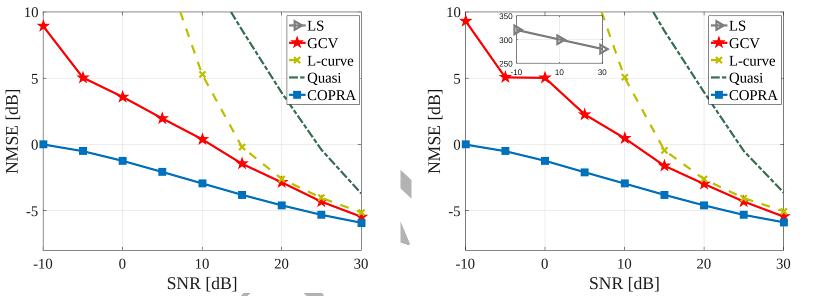


Fig. 3: NMSE [dB] versus SNR [dB] between Eq. (27) and Eq. (36) for various matrices $\mathbf{A}$. (a) Wing problem. (b) Heat problem. (c) Foxgood problem. (d) Deriv2 problem [48].



(a) $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ with i.i.d. elements.

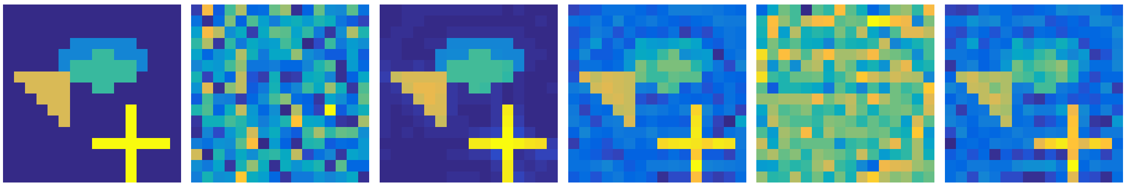
(b) $\mathbf{x}_0$ with i.i.d. elements distributed uniformly in the interval $(0, 1)$.

Fig. 4: Performance comparison when $\mathbf{A} = \frac{1}{50}\mathbf{B}\mathbf{B}^T$, where $\mathbf{B} \in \mathbb{R}^{50\times 45}$, $B_{ij} \sim \mathcal{N}(0, 1)$.

$\mathbf{B} \in \mathbb{R}^{50\times 45}$, $B_{ij} \sim \mathcal{N}(0, 1)$. The elements of $\mathbf{x}_0$ are chosen to be Gaussian i.i.d. with zero-mean unit variance, and are i.i.d. with uniform distribution within the interval $(0, 1)$. Results are obtained as an average over 10^5 different realizations of $\mathbf{A}$, $\mathbf{x}_0$, and $\mathbf{z}$.

When the elements of $\mathbf{x}_0$ are Gaussian i.i.d., COPRA outperforms the other regularization methods (Fig. 4). In fact, COPRA is the only approach that provides a NMSE below 0 dB across the entire SNR range. The other algorithms produce a very high NMSE. The same behavior can be observed when the elements of $\mathbf{x}_0$ are uniformly distributed (Fig. 4(b)). The NMSE of the LS method (not shown) is above 250 dB in both cases.

ACCEPTED MANUSCRIPT



(a) Original. (b) Recieved. (c) COPRA. (d) L-curve. (e) GCV. (f) Quasi.

Fig. 5: Tomography image restoration.

C. *Image restoration*

In this subsection, we present the visual results of the tomography image restoration.

Experiment setup: The elements of $\mathbf{A}\mathbf{x}_0$ represent a line of integrals along direct rays that penetrate a rectangular field. This field is discretized into n^2 cells, and each cell is stored with its own intensity as an element in the image matrix $\mathbf{M}$. Then, the columns of $\mathbf{M}$ are stacked into $\mathbf{x}_0$. The entries of $\mathbf{A}$ are generated as

$$a_{ij} = \begin{cases} l_{ij}, & \text{pixel}_j \in \text{ray}_i \\ 0 & \text{else}, \end{cases}$$

where l_{ij} is the length of the ith ray in pixel j. Finally, the rays are placed randomly. A noise with an SNR of 30 dB is added to the 16×16 image and the performance is evaluated as an average over 10^6 realizations of the noise and $\mathbf{A}$.

In Fig. 5, we present the original image, the received image, and the images reproduced by each of the regularization methods. Fig. 5 demonstrates that COPRA outperforms the other methods by providing a clear image that is very close to the original. A comparison of the peak signal-to-noise ratios (PSNR) of the algorithms (as in Table I) shows similar results, with COPRA having the largest PSNR. Moreover, the GCV algorithm provides an inaccurate result, while the L-curve and quasi-optimal algorithms fail to restore the internal parts clearly, especially those the parts with colors close to the background color.

TABLE I: PSNR of the algorithms with SNR=30

Method	COPRA	L-curve	GCV	Quasi-optimal
PSNR	**29.8331**	13.6469	10.5410	15.9080

ACCEPTED MANUSCRIPT

D. Average runtime

In Fig. 6, we plot the average runtime of each method against the calculated SNR from the simulation. The figure is a good representation for the runtime of all the problems (no significant runtime variation between problems has been seen). The figure shows that COPRA is the fastest algorithm as it has the shortest runtime in compare to all benchmark methods.

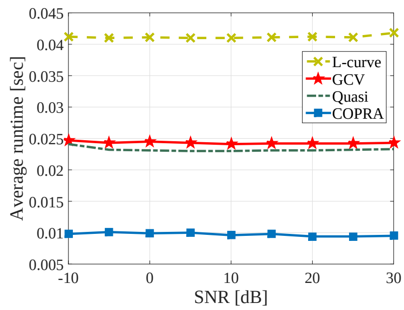


Fig. 6: Average runtime.

VI. Conclusions

In this paper, we developed a new regularization approach and a new regularization parameter selection method for linear discrete ill-posed problems. Due to the challenging singular-value structure of such problems, many regularization approaches fail to provide good stabilized solutions. In the proposed approach, the singular-value structure of the model matrix is modified by introducing an artificial perturbation into it. To maintain the fidelity of the model, an upper-bound constraint is added on the perturbation. The proposed approach minimizes the worst-case residual error of the estimator and selects a perturbation bound that approximately minimizes the MSE. Thus, the approach combines the simplicity of the ordinary least-squares criterion with the robustness of MSE-based estimation. The regularization parameter is obtained as the solution to a non-linear equation. Simulation results demonstrate that the proposed approach outperforms a set of benchmark regularization methods over a host of test problems.

Appendix A
Error Analysis

In this appendix, we analyze the error of the approximation used to obtain (30). To simplify the analysis, we consider the case when the approximation is applied directly to the MSE function

ACCEPTED MANUSCRIPT

given in (29). We start by defining $\mathbf{H} \triangleq \mathbf{\Sigma}^{2k}\left(\mathbf{\Sigma}^2+\rho \boldsymbol{I}_n\right)^{-p}$, whose diagonal entries are written as

$$h_{ii}=\frac{\sigma_i^{2k}}{\left(\sigma_i^2+\rho\right)^p};\ i=1,2,\cdots,n. \tag{A.1}$$

Note that in our case $k=0$ and $p=2$ for the diagonal matrix inside the trace function of the second term in (29). However, we use these two variables to obtain the error expression for the general case, then we substitute for k and p. By using the inequalities from [35, Eq.(5)], we obtain

$$\lambda_{\text{min}}(\mathbf{R}_{\mathbf{x}_0})\text{Tr}\left(\mathbf{H}\right)\leq \text{Tr}\left(\mathbf{H}\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}\right)\leq \lambda_{\text{max}}(\mathbf{R}_{\mathbf{x}_0})\text{Tr}\left(\mathbf{H}\right). \tag{A.2}$$

Similarly, we write

$$\lambda_{\text{min}}\left(\mathbf{H}\right)\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)\leq \text{Tr}\left(\mathbf{H}\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}\right)\leq \lambda_{\text{max}}\left(\mathbf{H}\right)\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right). \tag{A.3}$$

Since $\mathbf{H}$ is diagonal, $\lambda_{\text{min}}\left(\mathbf{H}\right)=\min\left(\text{diag}(\mathbf{H})\right)$ and $\lambda_{\text{max}}\left(\mathbf{H}\right)=\max\left(\text{diag}(\mathbf{H})\right)$. Now, we define the normalized error of the approximation as

$$\varepsilon=\frac{\text{Tr}\left(\mathbf{H}\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}\right)-\frac{1}{n}\text{Tr}\left(\mathbf{H}\right)\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)}{\frac{1}{n}\text{Tr}\left(\mathbf{H}\right)\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)}. \tag{A.4}$$

Note that this is not the standard way of defining the normalized error. Typically, the error ε is normalized by the true quantity, i.e., $\text{Tr}\left(\mathbf{H}\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}\right)$. However, this way of defining the error is more useful for carrying out the following error analysis. Based on (A.4), we see that $|\varepsilon|\geq 1$ indicates an inaccurate approximation. Although it depends totally on the application, we adopt $|\varepsilon|<1$ as the reference for evaluating the accuracy of the approximation. In fact, we can observe from (A.4) that $|\varepsilon|=1$ indicates that $\frac{1}{n}\text{Tr}\left(\mathbf{H}\right)\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)=0.5\ \text{Tr}\left(\mathbf{H}\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}\right)$. To this end, we derive two bounds based on (A.2) and (A.3). Then, we combine them to obtain the final error bound.

Absolute error bound based on (A.2): Subtracting $\frac{1}{n}\text{Tr}\left(\mathbf{H}\right)\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)$ from (A.2) and dividing by the same quantity, we obtain

$$\frac{\lambda_{\text{min}}(\mathbf{R}_{\mathbf{x}_0})}{\lambda_{\text{avg}}(\mathbf{R}_{\mathbf{x}_0})}-1\leq\varepsilon\leq\frac{\lambda_{\text{max}}(\mathbf{R}_{\mathbf{x}_0})}{\lambda_{\text{avg}}(\mathbf{R}_{\mathbf{x}_0})}-1, \tag{A.5}$$

where $\lambda_{\text{avg}}(\mathbf{R}_{\mathbf{x}_0})\triangleq\frac{1}{n}\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)$. Thus, $|\varepsilon|$ can be bounded by a positive quantity:

$$|\varepsilon_x|\leq\mu_x=\max\left[1-\frac{\lambda_{\text{min}}(\mathbf{R}_{\mathbf{x}_0})}{\lambda_{\text{avg}}(\mathbf{R}_{\mathbf{x}_0})},\frac{\lambda_{\text{max}}(\mathbf{R}_{\mathbf{x}_0})}{\lambda_{\text{avg}}(\mathbf{R}_{\mathbf{x}_0})}-1\right]. \tag{A.6}$$

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Absolute error bound based on (A.3): Starting from (A.3), and by applying the same approach used to obtain (A.6), we derive the second bound as

$$|\varepsilon_a| \leq \mu_a = \max\left[1 - \frac{\lambda_{\text{min}}(\mathbf{H})}{\lambda_{\text{avg}}(\mathbf{H})}, \frac{\lambda_{\text{max}}(\mathbf{H})}{\lambda_{\text{avg}}(\mathbf{H})} - 1\right]. \tag{A.7}$$

Using (A.1), wE transform (A.7) into

$$|\varepsilon_a| \leq \mu_a = \max\left[1 - \frac{\min\limits_i \left[\frac{\sigma_i^{2k}}{(\sigma_i^2+\rho)^p}\right]}{\frac{1}{n}\sum_{i=1}^n \frac{\sigma_i^{2k}}{(\sigma_i^2+\rho)^p}}, \frac{\max\limits_i \left[\frac{\sigma_i^{2k}}{(\sigma_i^2+\rho)^p}\right]}{\frac{1}{n}\sum_{i=1}^n \frac{\sigma_i^{2k}}{(\sigma_i^2+\rho)^p}} - 1\right] \quad i = 1, 2, \cdots, n. \tag{A.8}$$

The bound μ_x depends only on $\mathbf{R}_{\mathbf{x}_0}$, while μ_a depends on both the singular values of $\mathbf{A}$ and the unknown regularizer ρ. As in our case, $k = 0$ and $p = 2$, and therefore, (A.8) can be simplified to

$$|\varepsilon_a| \leq \mu_a = \max\left[1 - \frac{\frac{1}{(\sigma_1^2+\rho)^2}}{\frac{1}{n}\sum_{i=1}^n \frac{1}{(\sigma_i^2+\rho)^2}}, \frac{\frac{1}{(\sigma_n^2+\rho)^2}}{\frac{1}{n}\sum_{i=1}^n \frac{1}{(\sigma_i^2+\rho)^2}} - 1\right] \tag{A.9}$$

Combined bound: By combining (A.6) and (A.9), we obtain the final bound on the absolute error as

$$|\varepsilon| \leq \mu = \min\left(\mu_x, \mu_a\right). \tag{A.10}$$

From (A.6), (A.9), and (A.10), we notice that the bound is the minimum of two independent bounds. Below, we analyze each bound separately and then derive a conclusion concerning the overall error bound.

A. Analysis of μ_x

When $\mathbf{x}_0$ is deterministic, $\lambda_{\text{min}}(\mathbf{R}_{\mathbf{x}_0}) = 0$, $\lambda_{\text{max}}(\mathbf{R}_{\mathbf{x}_0}) = ||\mathbf{x}_0||_2^2$ and $\lambda_{\text{avg}}(\mathbf{R}_{\mathbf{x}_0}) = \frac{1}{n}||\mathbf{x}_0||_2^2$. By substituting in (A.6), we obtain

$$\mu_x = \max\left[1, n-1\right] = n - 1. \tag{A.11}$$

On the other hand, when $\mathbf{x}_0$ is stochastic with i.i.d. elements, $\lambda_{\text{min}}(\mathbf{R}_{\mathbf{x}_0}) = \lambda_{\text{avg}}(\mathbf{R}_{\mathbf{x}_0}) = \lambda_{\text{min}}(\mathbf{R}_{\mathbf{x}_0}) = \sigma_{\mathbf{x}_0}^2$. As a result,

$$\mu_x = \max\left[0, 0\right] = 0, \tag{A.12}$$

which means that, based on (A.10), the approximation is exact regardless of the contribution of the error from μ_a. When the entries of $\mathbf{x}_0$ are not i.i.d., μ_x is very difficult to obtain. Since no previous knowledge about $\mathbf{x}_0$ is assumed in this paper, this bound seems to be very loose for

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

a general $\mathbf{x}_0$, and we should rely on μ_a instead to tighten and evaluate the bound of the error. Thus, the modified bound, which can be larger than the actual bound, is given by

$$|\varepsilon| \leq \mu = \mu_a. \tag{A.13}$$

B. Analysis of μ_a

By taking the derivative of each of the two terms inside (A.9) w.r.t. ρ, we can prove that the two functions are decreasing in ρ. In other words, we obtain the two extreme error bounds (the largest and the smallest possible value of the absolute error) by analyzing the two extreme SNR scenarios, i.e., the high SNR regime and the low SNR regime.

1) Analysis of the low SNR regime: In the extreme low SNR regime, we have $\rho \to \infty$. Therefore, we can obtain the minimum bound on the absolute error. Based on (A.1), we write

$$h_{ii} = \frac{1}{(\sigma_i^2 + \rho)^2} \to \frac{1}{\rho^2};\ i = 1, 2, \cdots, n. \tag{A.14}$$

Consequently, (A.9) boils down after some manipulations to

$$\begin{aligned} |\varepsilon_a^l| \leq \mu_a^l &= \max\left[1 - \frac{\frac{1}{\rho^2}}{\frac{1}{n}\sum_{i=1}^n \frac{1}{\rho^2}}, \frac{\frac{1}{\rho^2}}{\frac{1}{n}\sum_{i=1}^n \frac{1}{\rho^2}} - 1\right] \\ &= \max\left[1 - \frac{n}{\sum_{i=1}^n 1}, \frac{n}{\sum_{i=1}^n 1} - 1\right] = 0. \end{aligned} \tag{A.15}$$

The result in (A.15) indicates that the approximation becomes more accurate as the SNR decreases. In the extreme low SNR regime (i.e., $\rho \to \infty$), the absolute error is zero and the approximated term is exactly equal to the original one.

2) Analysis of the high SNR regime: With extremely high SNR, ρ is sufficiently small. Therefore, doing an analysis allows us to obtain the upper worst-case value possible for ε.

Based on [49], there always exists a positive regularizer $\rho > 0$ such that the regularized estimator offers a lower MSE than the OLS estimator. This also applies to well-conditioned problems. However, if the condition number is too small, both the regularization parameter and the corresponding improvement in the MSE performance is to small compared to that of the OLS estimator. Therefore, we conclude that, ρ converges to a minimum value ρ_{min} for extremely high SNRs. In what follows, we find a lower-bound expression for ρ_{min} and examine the absolute error in this value.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

Starting from the definition of SNR, we can write

$$\text{SNR} = \frac{||\mathbf{A}\mathbf{x}_0||_2^2}{||\mathbf{z}||_2^2}. \tag{A.16}$$

Applying the SVD of $\mathbf{A}$ to (A.16) and then doing some algebraic manipulations, we obtain

$$\text{SNR} = \frac{\text{Tr}\left(\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\right)}{n\sigma_{\mathbf{z}}^2}, \tag{A.17}$$

where $\mathbf{R}_{\mathbf{x}_0} = \mathbf{x}_0\mathbf{x}_0^T$. Now, using (A.17) with the suboptimal regularizer ρ_{o} expression from (30), we write

$$\rho_{\text{o}} = \frac{1}{\text{SNR}}\frac{\text{Tr}\left(\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\right)}{\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)}. \tag{A.18}$$

From (A.18), we deduce the minimum achievable suboptimal regularizer ρ_{min} depends on the maximum SNR (i.e., SNR_{max}) for a given $\mathbf{A}$ and $\mathbf{x}_0$. That is

$$\rho_{\text{min}} = \frac{1}{\text{SNR}_{\text{max}}}\frac{\text{Tr}\left(\mathbf{V}\mathbf{\Sigma}^2\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\right)}{\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)}. \tag{A.19}$$

Given the nature of ill-posed problems and their singular values behavior, we partition $\mathbf{\Sigma}$ and $\mathbf{V}$ into two sub-matrices (as in Section II-C), and then approximate (A.19) as

$$\rho_{\text{min}} \approx \frac{1}{\text{SNR}_{\text{max}}}\frac{\text{Tr}\left(\mathbf{V}_{n1}\mathbf{\Sigma}_{n1}^2\mathbf{V}_{n1}^T\mathbf{R}_{\mathbf{x}_0}\right)}{\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)}, \tag{A.20}$$

where $\mathbf{\Sigma}_{n1}^2 = \text{diag}\left(\sigma_1^2, \ldots, \sigma_{n1}^2\right)$. The value of ρ_{min} in (A.20) can be bounded by

$$\frac{\sigma_{n1}^2}{\text{SNR}_{\text{max}}}\frac{\text{Tr}\left(\mathbf{V}_{n1}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}_{n1}\right)}{\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)} \leq \rho_{\text{min}} \leq \frac{\sigma_1^2}{\text{SNR}_{\text{max}}}\frac{\text{Tr}\left(\mathbf{V}_{n1}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}_{n1}\right)}{\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)} \tag{A.21}$$

Since we are considering the worst-case upper bound for the absolute error, and given that this absolute error increases as ρ decreases, we consider the lower bound of ρ_{min} as in (A.21). Moreover, based on the unitary matrix property and the partitioning of $\mathbf{V}$, we obtain a lower bound for the lower bound in (A.21) as

$$\rho_{\text{min}} \geq \frac{\sigma_{n1}^2}{\text{SNR}_{\text{max}}}\frac{\text{Tr}\left(\mathbf{V}_{n1}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}_{n1}\right)}{\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)} \geq \frac{\sigma_{n1}^2}{\text{SNR}_{\text{max}}}\frac{\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)}{\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right)}. \tag{A.22}$$

Thus, a lower bound for ρ_{min} (i.e., ρ_{min}^l) can be written as

$$\rho_{\text{min}}^l = \frac{\sigma_{n1}^2}{\text{SNR}_{\text{max}}}. \tag{A.23}$$

Now we are ready to study the behavior of the error in the high SNR regime. When $\rho \to \rho_{\text{min}}^l$, (A.9) can be written as

$$|\varepsilon_a^h| \leq \max\left[1 - \frac{\frac{1}{(\sigma_1^2+\rho_{\text{min}}^l)^2}}{\frac{1}{n}\sum_{i=0}^{n-1}\frac{1}{(\sigma_i^2+\rho_{\text{min}}^l)^2}}, \frac{\frac{1}{(\sigma_n^2+\rho_{\text{min}}^l)^2}}{\frac{1}{n}\sum_{i=0}^{n-1}\frac{1}{(\sigma_i^2+\rho_{\text{min}}^l)^2}} - 1\right]. \tag{A.24}$$

ACCEPTED MANUSCRIPT

To evaluate (A.24), we rely on numerical results. First, we consider $\text{SNR}_{\text{max}} = 40$ dB, which is a realistic upper value in many signal processing and communication applications. Substituting this value in (A.23), we find that $\rho^l_{\text{min}} = 0.018\sigma^2_{n1}$. By substituting the result in (A.24), and evaluating the expression for the nine ill-posed problems described in Section V, we find that $|\varepsilon^h_a| \leq \mu^h_a \approx 1$. As this represents the worst-case upper value for the absolute error, we find

$$|\varepsilon^h_a| \leq \varrho; \quad \varrho < 1. \tag{A.25}$$

Finally, based on (A.10) and (A.13), and by combining (A.15) and (A.25), we conclude that

$$|\varepsilon| \in [0, \varrho]; \quad \varrho < 1. \tag{A.26}$$

which indicates that $\frac{1}{n}\text{Tr}\left(\mathbf{H}\right)\text{Tr}\left(\mathbf{R}_{\mathbf{x}_0}\right) = q\ \text{Tr}\left(\mathbf{H}\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}\right)$ where $q \in (0.5, 1]$.

Appendix B
Proof of theorem 4

We are interested in studying the behavior of $\lim_{\rho_o \to \epsilon} G\left(\rho_o\right)$, assuming that ϵ is sufficiently small positive number (i.e., $\epsilon \to 0^+$), and $\epsilon \ll \sigma_i^2$, $\forall i \in [1, n]$. Starting from COPRA function in (38), and by defining $\mathbf{b} = \mathbf{U}^T\mathbf{y}$, we write

$$G\left(\rho_o\right) = \sum_{i=1}^{n} \frac{\sigma_i^2 b_i^2}{\left(\sigma_i^2 + \rho_o\right)^2} \sum_{j=1}^{n_1} \frac{\left(\beta\sigma_j^2 + \rho_o\right)}{\left(\sigma_j^2 + \rho_o\right)^2} - \sum_{i=1}^{n} \frac{b_i^2}{\left(\sigma_i^2 + \rho_o\right)^2} \sum_{j=1}^{n_1} \frac{\sigma_j^2\left(\beta\sigma_j^2 + \rho_o\right)}{\left(\sigma_j^2 + \rho_o\right)^2} + \frac{n_2}{\rho_o} \sum_{i=1}^{n} \frac{\sigma_i^2 b_i^2}{\left(\sigma_i^2 + \rho_o\right)^2}. \tag{B.1}$$

Given how we choose ϵ, Eq. (B.1) can be approximated as

$$G\left(\epsilon\right) \approx \beta \sum_{i=1}^{n} \sigma_i^{-2} b_i^2 \sum_{j=1}^{n_1} \sigma_j^{-2} - \beta n_1 \sum_{i=1}^{n} \sigma_i^{-4} b_i^2 + \frac{n_2}{\epsilon} \sum_{i=1}^{n} \sigma_i^{-2} b_i^2. \tag{B.2}$$

Solving $G\left(\epsilon\right) = 0$ from (B.2) leads to the following root:

$$\epsilon = \frac{n_2 \sum_{i=1}^{n} \sigma_i^{-2} b_i^2}{\beta n_1 \sum_{i=1}^{n} \sigma_i^{-4} b_i^2 - \beta \sum_{i=1}^{n} \sigma_i^{-2} b_i^2 \sum_{j=1}^{n_1} \sigma_j^{-2}}. \tag{B.3}$$

Now, we determine if the root defined by (B.3) is positive. For (B.3) to be positive, the following relation should hold:

$$n_1 \sum_{i=1}^{n} \sigma_i^{-4} b_i^2 \geq \sum_{i=1}^{n} \sigma_i^{-2} b_i^2 \sum_{j=1}^{n_1} \sigma_j^{-2}. \tag{B.4}$$

Starting from the right-hand side of (B.4), and with $\sigma_1 \geq \cdots \geq \sigma_n$, we bound this term as

$$\sum_{i=1}^{n} \sigma_i^{-2} b_i^2 \sum_{j=1}^{n_1} \sigma_j^{-2} \leq \sigma_{n_1}^{-2} \sum_{i=1}^{n} \sigma_i^{-2} b_i^2 \sum_{j=1}^{n_1} 1 = n_1 \sigma_{n_1}^{-2} \sum_{i=1}^{n} \sigma_i^{-2} b_i^2. \tag{B.5}$$

ACCEPTED MANUSCRIPT

However, given how we choose n_1 and n_2, we also have

$$\sum_{i=n_1+1}^{n} \sigma_i^{-2} \geq \sum_{i=1}^{n_1} \sigma_i^{-2}, \tag{B.6}$$

which helps us to bound the left-hand side of (B.4) as

$$n_1 \sum_{i=1}^{n} \sigma_i^{-4} b_i^2 \geq n_1 \sigma_{n_1}^{-2} \sum_{i=1}^{n} \sigma_i^{-2} b_i^2. \tag{B.7}$$

From (B.5) and (B.7), we find that the lower bound for the left-hand side of (B.4) is equal to the upper bound of its right-hand side. Then, we can conclude from these two relations that

$$n_1 \sum_{i=1}^{n} \sigma_i^{-4} b_i^2 \geq \sum_{i=1}^{n} \sigma_i^{-2} b_i^2 \sum_{j=1}^{n_1} \sigma_j^{-2}. \tag{B.8}$$

Thus, ϵ is a positive root for the COPRA function in (38).

Next, we would like to know if ϵ can be considered as a value for our regularization parameter. A direct way to show that can be deduced from the fact that having $\epsilon \ll \sigma_i^2 \quad \forall i \in [1, n]$ will not provide any regularization to the problem. Hence, the RLS estimator in (7) converges to the OLS in (3).

As a remark, we assume that the approximation in (B.2) is uniform, such that it does not affect the position of the roots. Therefore, we can claim that this root is not coming from the negative region of the axis. In fact, we can easily prove that (B.1) does not have a negative root that is close to zero. Thus, this root does not come from the negative region as a result of this function approximation (i.e., perturbed root).

References

[1] Thomas Kailath, Ali H Sayed, and Babak Hassibi, *Linear estimation*, vol. 1, Prentice Hall Upper Saddle River, NJ, 2000.

[2] H Vincent Poor, *An introduction to signal detection and estimation*, Springer Science & Business Media, 2013.

[3] Charles W Groetsch and CW Groetsch, *Inverse problems in the mathematical sciences*, vol. 52, Springer, 1993.

[4] Steven M Kay, *Fundamentals of Statistical Signal Processing: Practical Algorithm Development*, vol. 3, Pearson Education, 2013.

[5] Martin A Fischler and Oscar Firschein, *Readings in Computer Vision: Issues, Problem, Principles, and Paradigms*, Morgan Kaufmann, 2014.

[6] Misha E Kilmer and Dianne P O'Leary, "Choosing regularization parameters in iterative methods for ill-posed problems," *SIAM Journal on matrix analysis and applications*, vol. 22, no. 4, pp. 1204–1221, 2001.

[7] Tomasz Piotrowski and Isao Yamada, "Mv-pure estimator: Minimum-variance pseudo-unbiased reduced-rank estimator for linearly constrained ill-conditioned inverse problems," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3408–3423, 2008.

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

[8] Weifeng Liu, Puskal P Pokharel, and Jose C Principe, “The kernel least-mean-square algorithm,” *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 543–554, 2008.

[9] Mario Bertero, Tomaso A Poggio, and Vincent Torre, “Ill-posed problems in early vision,” *Proceedings of the IEEE*, vol. 76, no. 8, pp. 869–889, 1988.

[10] Tomaso Poggio, Vincent Torre, and Christof Koch, “Computational vision and regularization theory,” *Nature*, vol. 317, no. 6035, pp. 314–319, 1985.

[11] Frank Natterer, *The mathematics of computerized tomography*, vol. 32, Siam, 1986.

[12] Ian Craig and John Brown, “Inverse problems in astronomy,” 1986.

[13] Aggelos K Katsaggelos, Jan Biemond, Ronald W Schafer, and Russell M Mersereau, “A regularized iterative image restoration algorithm,” *IEEE Transactions on Signal Processing*, vol. 39, no. 4, pp. 914–929, 1991.

[14] Per Christian Hansen, James G Nagy, and Dianne P O’leary, *Deblurring images: matrices, spectra, and filtering*, vol. 3, Siam, 2006.

[15] Vincent Torre and Tomaso A Poggio, “On edge detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, , no. 2, pp. 147–163, 1986.

[16] Richard C Aster, Brian Borchers, and Clifford H Thurber, “Parameter estimation and inverse problems elsevier,” *New York*, 2005.

[17] Per Christian Hansen and Dianne Prost O’Leary, “The use of the l-curve in the regularization of discrete ill-posed problems,” *SIAM Journal on Scientific Computing*, vol. 14, no. 6, pp. 1487–1503, 1993.

[18] Weidong Chen, “A new extrapolation algorithm for band-limited signals using the regularization method,” *IEEE Transactions on Signal Processing*, vol. 41, no. 3, pp. 1048–1060, 1993.

[19] James M Varah, “Pitfalls in the numerical solution of linear ill-posed problems,” *SIAM Journal on Scientific and Statistical Computing*, vol. 4, no. 2, pp. 164–176, 1983.

[20] Martin Hanke and Per Christian Hansen, “Regularization methods for large-scale problems,” *Survey on Mathematics for Industry*, vol. 3, no. 4, 1993.

[21] Yonina C Eldar and Alan V Oppenheim, “Covariance shaping least-squares estimation,” *IEEE Transactions on Signal Processing*, vol. 51, no. 3, pp. 686–697, 2003.

[22] Yonina C Eldar, “Improvement of least-squares under arbitrary weighted mse,” in *IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2007*. IEEE, 2007, vol. 3, pp. III–837.

[23] AN Tikhonov and V Ya Arsenin, *Methods for solving ill-posed problems*, John Wiley and Sons, Inc, 1977.

[24] Grace Wahba, *Spline models for observational data*, vol. 59, Siam, 1990.

[25] Per Christian Hansen, “Analysis of discrete ill-posed problems by means of the l-curve,” *SIAM review*, vol. 34, no. 4, pp. 561–580, 1992.

[26] Per Christian Hansen, Toke Koldborg Jensen, and Giuseppe Rodriguez, “An adaptive pruning algorithm for the discrete l-curve criterion,” *Journal of computational and applied mathematics*, vol. 198, no. 2, pp. 483–492, 2007.

[27] Vladimir Alekseevich Morozov, *Methods for solving incorrectly posed problems*, Springer Science & Business Media, 2012.

[28] Frank Bauer and Mark A Lukas, “Comparingparameter choice methods for regularization of ill-posed problems,” *Mathematics and Computers in Simulation*, vol. 81, no. 9, pp. 1795–1841, 2011.

[29] Mohamed Suliman, Tarig Ballal, Abla Kammoun, and Tareq Y. Al-Naffouri, “Penalized linear regression for discrete

ACCEPTED MANUSCRIPT

ACCEPTED MANUSCRIPT

ill-posed problems: A hybrid least-squares and mean-squared error approach," in *Proceedings European Signal Processing Conference (EUSIPCO)*, 2016, pp. 403–407.

[30] Yonina C Eldar, Aharon Ben-Tal, and Arkadi Nemirovski, "Robust mean-squared error estimation in the presence of model uncertainties," *IEEE Transactions on Signal Processing*, vol. 53, no. 1, pp. 168–181, 2005.

[31] Dilip N Ghosh Roy and LS Couchman, *Inverse problems and inverse scattering of plane waves*, Academic Press, 2001.

[32] Laurent El Ghaoui and Hervé Lebret, "Robust solutions to least-squares problems with uncertain data," *SIAM Journal on Matrix Analysis and Applications*, vol. 18, no. 4, pp. 1035–1064, 1997.

[33] S Chandrasekaran, GH Golub, M Gu, and AH Sayed, "Parameter estimation in the presence of bounded data uncertainties," *SIAM Journal on Matrix Analysis and Applications*, vol. 19, no. 1, pp. 235–252, 1998.

[34] Lech Maligranda, "A simple proof of the hölder and the minkowski inequality," *The American mathematical monthly*, vol. 102, no. 3, pp. 256–259, 1995.

[35] Sheng-de Wang, Te-Son Kuo, and Chen-Fa Hsu, "Trace bounds on the solution of the algebraic matrix riccati and lyapunov equation," *IEEE Transactions on Automatic Control*, vol. 31, no. 7, pp. 654–656, 1986.

[36] Dave Zachariah and Petre Stoica, "Online hyperparameter-free sparse estimation method," *IEEE Transactions on Signal Processing*, vol. 63, no. 13, pp. 3348–3359, 2015.

[37] William J Hemmerle, "An explicit solution for generalized ridge regression," *Technometrics*, vol. 17, pp. 309–314, 1975.

[38] Mohamed A Suliman, Ayed M Alrashdi, Tarig Ballal, and Tareq Y Al-Naffouri, "Snr estimation in linear systems with gaussian matrices," *IEEE Signal Processing Letters*, vol. 24, no. 12, pp. 1867–1871, 2017.

[39] Willliam Feller, *An introduction to probability theory and its applications*, vol. 2, John Wiley & Sons, 2008.

[40] David Vernon Widder, *Laplace Transform (PMS-6)*, Princeton University Press, 2015.

[41] David W Kammler, "Chebyshev approximation of completely monotonic functions by sums of exponentials," *SIAM Journal on Numerical Analysis*, vol. 13, no. 5, pp. 761–774, 1976.

[42] David W Kammler, "Least squares approximation of completely monotonic functions by sums of exponentials," *SIAM Journal on Numerical Analysis*, vol. 16, no. 5, pp. 801–818, 1979.

[43] Yuri K Shestopaloff, *Sums of exponential functions and their new fundamental properties, with applications to natural phenomena*, Akvy Press, 2008.

[44] Christopher J Zarowski, *An introduction to numerical analysis for electrical and computer engineers*, John Wiley & Sons, 2004.

[45] Mohamed Suliman, Tarig Ballal, and Tareq Y Al-Naffouri, "Robust regularized least-squares beamforming approach to signal estimation," in *Signal and Information Processing (GlobalSIP), 2016 IEEE Global Conference on*. IEEE, 2016, pp. 75–79.

[46] Mohamed Suliman, Tarig Ballal, Abla Kammoun, and Tareq Y Al-Naffouri, "Constrained perturbation regularization approach for signal estimation using random matrix theory," *IEEE signal processing letters*, vol. 23, no. 12, pp. 1727–1731, 2016.

[47] Tarig Ballal, Mohamed A Suliman, and Tareq Y Al-Naffouri, "Bounded perturbation regularization for linear least squares estimation," *IEEE Access*, vol. 5, pp. 27551–27562, 2017.

[48] Per Christian Hansen, "Regularization tools: A matlab package for analysis and solution of discrete ill-posed problems," *Numerical algorithms*, vol. 6, no. 1, pp. 1–35, 1994.

[49] Arthur E Hoerl and Robert W Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, no. 1, pp. 55–67, 1970.

ACCEPTED MANUSCRIPT