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

Mohamed Suliman, *Student Member, IEEE*, Tarig Ballal, *Member, IEEE*, and Tareq Y. Al-Naffouri, *Member, IEEE*

***Abstract*—Estimating the values of unknown parameters from corrupted measured data faces a lot of challenges in ill-posed problems. In such problems, many fundamental estimation methods fail to provide a meaningful stabilized solution. In this work, we propose a new regularization approach and a new regularization parameter selection approach for linear least-squares discrete ill-posed problems. The proposed approach is based on enhancing the singular-value structure of the ill-posed model matrix to acquire a better solution. Unlike many other regularization algorithms 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 typical 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 demonstrate that the proposed approach outperforms a set of benchmark regularization methods in most cases. In addition, the approach also enjoys the lowest runtime and offers the highest level of robustness amongst all the tested benchmark regularization methods.**

***Index Terms*—Linear estimation, ill-posed problems, linear least squares, regularization.**

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 m noisy, linear observations given by $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$. Here, $\mathbf{A} \in \mathbb{R}^{m\times n}$ is a known linear measurement matrix, and, $\mathbf{z} \in \mathbb{R}^{m\times 1}$ is the noise vector; the latter is assumed to be additive white Gaussian noise (AWGN) vector with unknown variance $\sigma_{\mathbf{z}}^2$ that is independent of $\mathbf{x}_0$. Such problem has been extensively studied over the years due to its obvious practical importance as well as its theoretical interest [1]–[3]. It arises in many fields of science and engineering, e.g., communication, signal processing, computer vision, control theory, and economics.

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) [4] that finds an estimate $\hat{\mathbf{x}}_{\text{OLS}}$ of $\mathbf{x}_0$ by minimizing the Euclidean norm of the residual error

$$\hat{\mathbf{x}}_{\text{OLS}} = \arg\min_{\mathbf{x}} ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2. \tag{1}$$

The behavior of the OLS has been extensively studied in the literature and it is now very well understood. In particular, if $\mathbf{A}$ is a full column rank, (1) has a unique solution given by

$$\hat{\mathbf{x}}_{\text{OLS}} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{y} = \mathbf{V}\mathbf{\Sigma}^{-1}\mathbf{U}^T\mathbf{y}, \tag{2}$$

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 the right orthogonal singular vectors, while the singular values σ_i are assumed to satisfy $\sigma_1 \geq \sigma_1 \geq \cdots \geq \sigma_n$. A major difficulty associated with the OLS approach is in discrete ill-posed problems. A problem is considered to be well-posed if its solution always exists, unique, and depends continuously on the initial data. Ill-posed problems fail to satisfy at least one of these conditions [5]. In such problems, the matrix $\mathbf{A}$ is ill-conditioned and the computed LS solution in (2) is potentially very sensitive to perturbations in the data such as the noise $\mathbf{z}$ [6].

Discrete ill-posed problems are of great practical interest in the field of signal processing and computer vision [7]–[10]. They arise in a variety of applications such as computerized tomography [11], astronomy [12], image restoration and deblurring [13], [14], edge detection [15], seismography [16], stereo matching [17], and the computation of lightness and surface reconstruction [18]. Interestingly, in all these applications and even more, data are gathered by convolution of a noisy signal with a detector [19], [20]. A linear representation of such process is normally given by

$$\int_{b_1}^{b_2} a\left(s,t\right)\mathbf{x}_0\left(t\right)\mathrm{d}t = \mathbf{y}_0\left(s\right) + \mathbf{z}\left(s\right) = \mathbf{y}\left(s\right), \tag{3}$$

where $\mathbf{y}_0\left(s\right)$ is the true signal, while the kernal function $a\left(s,t\right)$ represents the response. It is shown in [21] how a problem with a formulation similar to (3) fails to satisfy the well-posed conditions introduced above. The discretized version of (3) can be represented by $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$.

To solve ill-posed problems, regularization methods are commonly used. These methods are based on introducing an additional prior information in the problem. All regularization methods are used to generate a reasonable solution for the ill-posed problem by replacing the problem with a well-posed one whose solution is acceptable. This must be done after careful analysis to the ill-posed problem in terms of its physical plausibility and its mathematical properties.

Several regularization approaches have been proposed throughout the years. Among them are the truncated SVD [22], the maximum entropy principle [23], the hybrid methods [24], the covariance shaping LS estimator [25], and the weighted LS [26]. The most common and widely used approach is the regularized M-estimator that obtains an estimate $\hat{\mathbf{x}}$ of $\mathbf{x}_0$ from

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

$\mathbf{y}$ by solving the convex problem

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}} \mathcal{L}\left(\mathbf{y} - \mathbf{A}\mathbf{x}\right) + \gamma f\left(\mathbf{x}\right), \tag{4}$$

where the loss function $\mathcal{L} : \mathbb{R}^m \to \mathbb{R}$ measures the fit of $\mathbf{A}\mathbf{x}$ 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 leads to different estimators. The most popular among them is the Tikhonov regularization which is given in its simplified form by [27]

$$\hat{\mathbf{x}}_{\text{RLS}} := \arg\min_{\mathbf{x}} \ ||\mathbf{y} - \mathbf{A}\mathbf{x}||_2^2 + \gamma \ ||\mathbf{x}||_2^2. \tag{5}$$

The solution to (5) is given by the regularized least-square (RLS) estimator

$$\hat{\mathbf{x}}_{\text{RLS}} = \left(\mathbf{A}^T\mathbf{A} + \gamma\mathbf{I}_n\right)^{-1}\mathbf{A}^T\mathbf{y}, \tag{6}$$

where $\mathbf{I}_n$ is $(n \times n)$ identity matrix. In general, γ is unknown and has to be chosen judiciously.

On the other hand, several regularization parameter selection methods have been proposed to find the regularization parameter in regularization methods. These include the *generalized cross validation* (GCV) [28], the *L-curve* [29], [30], and the *quasi-optimal* method [31]. The GCV obtains the regularizer by minimizing the GCV function which suffers from the shortcoming that it may have a very flat minimum that makes it very challenging to be located numerically. The L-curve, on the other hand, is a graphical tool to obtain the regularization parameter which has a very high computational complexity. Finally, the quasi-optimal criterion chooses the regularization parameter without taking into account the noise level. In general, the performance of these methods varies significantly depending on the nature of the problem[1].

A. Paper Contributions

1) *New regularization approach*: We proposed a new approach for linear discrete ill-posed problems that is based on adding an artificial perturbation matrix with a bounded norm to $\mathbf{A}$. The objective of this artificial perturbation is to improve the challenging singular-value structure of $\mathbf{A}$. This perturbation affects the fidelity of the model $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$, and as a result, the equality relation becomes invalid. We show that using such modification provides a solution with better numerical stability.
2) *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 $\mathbf{x}_0$ and its estimate $\hat{\mathbf{x}}$, $\mathbb{E}\ ||\hat{\mathbf{x}} - \mathbf{x}_0||_2^2$. [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 about it. Moreover, we assume that the noise variance $\sigma_{\mathbf{z}}^2$ is unknown. Finally, the approach can be applied to a large number of linear discrete ill-posed problems.

[1]The work presented in this paper is an extended version of [32].

[2]Less work have been done in the literature to provide estimators that are based on the MSE as for example in [33] where the authors derived an estimator for the linear model problem that is based on minimizing the *worst-case* MSE (as opposed to the actual MSE) while imposing a constraint on the unknown vector $\mathbf{x}_0$.

B. Paper Organization

This paper is organized as follows. Section II presents the formulation of the problem and derive its solution. In Section III, we derive the artificial perturbation bound that minimizes the MSE. Further, we derive the proposed approach characteristic equation for obtaining the regularization parameter. Section IV studies the properties of the approach characteristic equation while Section V presents the performance of the proposed approach using simulations. Finally, we conclude our work in Section VI.

C. Notations

Matrices are given in boldface upper case letters (e.g., $\mathbf{X}$), column vectors are represented by boldface lower case letters (e.g., $\mathbf{x}$), and $(.)^T$ stands for the transpose operator. Further, $\mathbb{E}\left(.\right)$, $\mathbf{I}_n$, and $\mathbf{0}$ denote the expectation operator, the $(n \times n)$ identity matrix, and the zero matrix, respectively. Notation $||.||_2$ refers to the spectral norm for matrices and Euclidean norm for vectors. The operator $\text{diag}\left(.\right)$ returns a vector that contains the diagonal elements of a matrix, and a diagonal matrix if it operates on a vector where the diagonal entries of the matrix are the elements of the vector.

II. Proposed Regularization Approach

A. Background

We consider the linear discrete ill-posed problems in the form $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$ and we focus mainly on the case where $m \geq n$ without imposing any assumptions on $\mathbf{x}_0$. The matrix $\mathbf{A}$ in such problems is ill-conditioned that has a very fast singular values decay [34]. A comparison between the singular values decay of the full rank matrices, the rank deficient matrices, and the ill-posed problems matrices is given in Fig. 1.

From Fig. 1, we observe that the singular values of the full column rank matrix are decaying constantly while in the rank definite matrix there is a jump (gap) between the nonzero and the zero singular values. Finally, the singular values of the ill-posed problem matrix are decaying very fast, without a gap, to a significantly small positive number.

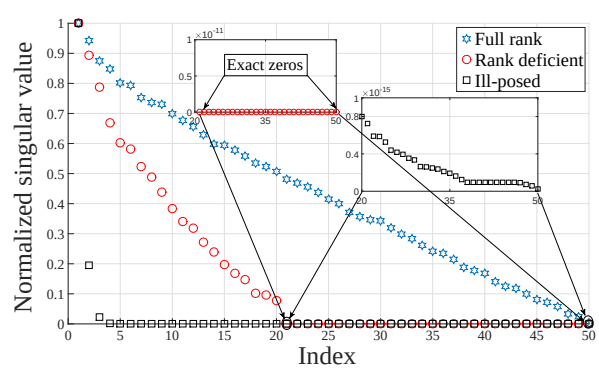


Fig. 1: Different singular values decay for different matrices, $\mathbf{A} \in \mathbb{R}^{50\times 50}$.

B. Problem Formulation

Let us start by considering the LS solution in (2). In many ill-posed problems, and due to the singular-value structure of $\mathbf{A}$ and the interaction that it has with the noise, equation (2) is not capable of producing a sensible estimate of $\mathbf{x}_0$. Herein, we propose adding an artificial perturbation $\Delta\mathbf{A} \in \mathbb{R}^{m\times n}$ to $\mathbf{A}$. We assume that this perturbation, which replaces $\mathbf{A}$ by $(\mathbf{A}+\Delta\mathbf{A})$, improves the challenging singular-value structure, and hence is capable of producing a better estimate of $\mathbf{x}_0$. In other words, we assume that using $(\mathbf{A}+\Delta\mathbf{A})$ in estimating $\mathbf{x}_0$ from $\mathbf{y}$ provides a better estimation result than using $\mathbf{A}$. Finally, to provide a balance between improving the singular-value structure and maintaining the fidelity of the basic linear model, we add the constraint $||\Delta\mathbf{A}||_2 \leq \delta$, $\delta \in \mathbb{R}^+$. Therefore, the basic linear model is modified to

$$\mathbf{y} \approx (\mathbf{A}+\Delta\mathbf{A})\,\mathbf{x}_0 + \mathbf{z};\ \ ||\Delta\mathbf{A}||_2 \leq \delta. \tag{7}$$

The question now is what is the best $\Delta\mathbf{A}$ and the bound on this perturbation. It is clear that these values are important since they affect the model's fidelity and dictate the quality of the estimator. This question is addressed further ahead in this section. For now, let us start by assuming that δ is known[3].

Before proceeding, it worth mentioning that the model in (7) has been considered for signal estimation in the presence of data errors but with strict equality (e.g., [33], [35], [36]). These methods assume that $\mathbf{A}$ is not known perfectly due to some error contamination, and that a prior knowledge on the real error bound (which corresponds to δ in our case) is available. However, in our case the matrix $\mathbf{A}$ is known perfectly, whereas δ is unknown.

To obtain an estimate of $\mathbf{x}_0$, we consider minimizing the worst-case residual function of the new perturbed model in (7) which is given by

$$\min_{\mathbf{x}} \max_{||\Delta\mathbf{A}||_2\leq\delta} Q(\mathbf{x},\Delta\mathbf{A}) := \ ||\mathbf{y}-(\mathbf{A}+\Delta\mathbf{A})\,\mathbf{x}||_2. \tag{8}$$

Theorem 1. *The unique minimzer $\hat{\mathbf{x}}$ of the min-max constrained problem in (8) for fixed $\delta > 0$ is given by*

$$\hat{\mathbf{x}} = \left(\mathbf{A}^T\mathbf{A} + \rho(\delta,\hat{\mathbf{x}})\,\mathbf{I}_n\right)^{-1}\mathbf{A}^T\mathbf{y}, \tag{9}$$

where $\rho(\delta,\hat{\mathbf{x}})$ is a regularization parameter that is related to the perturbation bound δ through

$$\rho(\delta,\hat{\mathbf{x}}) = \delta\ \frac{||\mathbf{y}-\mathbf{A}\hat{\mathbf{x}}||_2}{||\hat{\mathbf{x}}||_2}. \tag{10}$$

Proof: By using Minkowski inequality [37], we find an upper bound for the cost function $Q(\mathbf{x},\Delta\mathbf{A})$ in (8) as

$$\begin{aligned}||\mathbf{y}-(\mathbf{A}+\Delta\mathbf{A})\,\mathbf{x}||_2 &\leq ||\mathbf{y}-\mathbf{A}\mathbf{x}||_2 + ||\Delta\mathbf{A}\ \mathbf{x}||_2\\ &\leq ||\mathbf{y}-\mathbf{A}\mathbf{x}||_2 + ||\Delta\mathbf{A}||_2||\mathbf{x}||_2\\ &\leq ||\mathbf{y}-\mathbf{A}\mathbf{x}||_2 + \delta\ ||\mathbf{x}||_2.\end{aligned} \tag{11}$$

However, upon setting $\Delta\mathbf{A}$ to be the following rank one matrix

$$\Delta\mathbf{A} = \frac{(\mathbf{A}\mathbf{x}-\mathbf{y})}{||\mathbf{y}-\mathbf{A}\mathbf{x}||_2}\frac{\mathbf{x}^T}{||\mathbf{x}||_2}\delta, \tag{12}$$

we can show that the bound in (11) is achievable by

$$\begin{aligned}||\mathbf{y}-(\mathbf{A}+\Delta\mathbf{A})\,\mathbf{x}||_2 &= ||\,(\mathbf{y}-\mathbf{A}\mathbf{x}) + \frac{(\mathbf{y}-\mathbf{A}\mathbf{x})}{||\mathbf{y}-\mathbf{A}\mathbf{x}||_2}\frac{\mathbf{x}^T}{||\mathbf{x}||_2}\mathbf{x}\delta||_2\\ &= ||\,(\mathbf{y}-\mathbf{A}\mathbf{x}) + \frac{(\mathbf{y}-\mathbf{A}\mathbf{x})}{||\mathbf{y}-\mathbf{A}\mathbf{x}||_2}||\mathbf{x}||_2\delta||_2.\end{aligned} \tag{13}$$

Since the two added vectors $(\mathbf{y}-\mathbf{A}\mathbf{x})$ and $\frac{(\mathbf{y}-\mathbf{A}\mathbf{x})}{||\mathbf{y}-\mathbf{A}\mathbf{x}||_2}||\mathbf{x}||_2\delta$ in (13) are positively linearly dependent (i.e., pointing in the same direction), we conclude that

$$||\,(\mathbf{y}-\mathbf{A}\mathbf{x}) + \frac{(\mathbf{y}-\mathbf{A}\mathbf{x})}{||\mathbf{y}-\mathbf{A}\mathbf{x}||_2}||\mathbf{x}||_2\ \delta||_2 = \underbrace{||\mathbf{y}-\mathbf{A}\mathbf{x}||_2 + \delta||\mathbf{x}||_2}_{W(\mathbf{x})} \tag{14}$$

As a result, (8) can be expressed equivalently by

$$\min_{\mathbf{x}} \max_{||\Delta\mathbf{A}||_2\leq\delta} Q(\mathbf{x},\Delta\mathbf{A}) \equiv \min_{\mathbf{x}} W(\mathbf{x}). \tag{15}$$

Therefore, the solution of (8) depends only on δ and is agnostic to the structure of $\Delta\mathbf{A}$[4]. It is easy to check that the solution space for $W(\mathbf{x})$ 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 $W(\mathbf{x})$ is zero, or $W(\mathbf{x})$ is not differentiable. More precisely, $W(\mathbf{x})$ is not differentiable only at $\mathbf{x}$ = 0 and when $\mathbf{y}-\mathbf{A}\mathbf{x} = 0$. However, the former case is a trivial case that is not being considered in this paper, while the latter case is not possible by definition. The gradient of $W(\mathbf{x})$ can be obtained as

$$\begin{aligned}\nabla_{\mathbf{x}}W(\mathbf{x}) &= \frac{1}{||\mathbf{y}-\mathbf{A}\mathbf{x}||_2}\mathbf{A}^T(\mathbf{A}\mathbf{x}-\mathbf{y}) + \frac{\delta\ \mathbf{x}}{||\mathbf{x}||_2}\\ &= \frac{1}{||\mathbf{y}-\mathbf{A}\mathbf{x}||_2}\left(\mathbf{A}^T\mathbf{A}\mathbf{x} + \frac{\delta\ ||\mathbf{y}-\mathbf{A}\mathbf{x}||_2\ \mathbf{x}}{||\mathbf{x}||_2} - \mathbf{A}^T\mathbf{y}\right).\end{aligned} \tag{16}$$

Solving for $\nabla_{\mathbf{x}}W(\hat{\mathbf{x}}) = 0$ upon defining $\rho(\delta,\hat{\mathbf{x}})$ as in (10), we obtain (9). ■

Remark 1. The regularization parameter ρ in (10) is a function of the unknown estimate $\hat{\mathbf{x}}$, as well as the upper perturbation bound δ (we have dropped the dependence of ρ on δ and $\hat{\mathbf{x}}$ to simplify notation). In addition, it is clear from (14) that δ controls the weight given to the minimization of the side constraint relative to the minimization of the residual norm. We have assumed that δ is known to obtain the min-max optimization solution. However, this assumption is not valid in reality. Thus, is it impossible to obtain ρ directly from (10) given that both δ and $\hat{\mathbf{x}}$ are unknowns.

Now, it is obvious with (9) and (10) in hand, we can eliminate the dependency of ρ on $\hat{\mathbf{x}}$. By substituting (9) in (10) we obtain after some algebraic manipulations

$$\begin{aligned}&\delta^2\Big[\mathbf{y}^T\mathbf{y} - 2\mathbf{y}^T\mathbf{A}\left(\mathbf{A}^T\mathbf{A}+\rho\mathbf{I}_n\right)^{-1}\mathbf{A}^T\mathbf{y}\\ &+ ||\mathbf{A}\left(\mathbf{A}^T\mathbf{A}+\rho\mathbf{I}_n\right)^{-1}\mathbf{A}^T\mathbf{y}||^2\Big]\\ &= \rho^2\mathbf{y}^T\mathbf{A}\left(\mathbf{A}^T\mathbf{A}+\rho\mathbf{I}_n\right)^{-2}\mathbf{A}^T\mathbf{y}.\end{aligned} \tag{17}$$

[3]We will use this assumption to obtain the proposed estimator solution as a function of δ, then, we will address the problem of obtaining the value of δ.

[4]Interestingly, setting the norm of the penalty term in $W(\mathbf{x})$ to be of l1 norm (i.e., $||\mathbf{x}||_1$) leads to the square-root LASSO [38] which is used in sparse signal estimation.

In this following subsection, we will utilize and simplify (17) using some manipulations to obtain δ that corresponds to an optimal choice of ρ in ill-posed problems.

C. Finding the Optimal Perturbation Bound

Let us denote the optimal choices of ρ and δ by $\rho_{\rm o}$ and $\delta_{\rm o}$, respectively. To simplify (17), we substitute the SVD of $\mathbf{A}$, then we solve for δ^2, and finally we take the trace $\text{Tr}\left(.\right)$ of the two sides considering the evaluation point to be $(\delta_{\rm o}, \rho_{\rm o})$ to get

$$\underbrace{\delta_{\rm o}^2 \ \text{Tr}\left(\left(\mathbf{\Sigma}^2+\rho_{\rm o}\mathbf{I}_n\right)^{-2}\mathbf{U}^T\left(\mathbf{y}\mathbf{y}^T\right)\mathbf{U}\right)}_{D(\rho_{\rm o})} = \underbrace{\text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho_{\rm o}\mathbf{I}_n\right)^{-2}\mathbf{U}^T\left(\mathbf{y}\mathbf{y}^T\right)\mathbf{U}\right)}_{N(\rho_{\rm o})}. \tag{18}$$

In order to obtain a useful expression, let us think of $\delta_{\rm o}$ 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 by its expected value $\mathbb{E}(\mathbf{y}\mathbf{y}^T)$. In other words, we are looking for an optimal choice of δ, say $\delta_{\rm o}$, that is optimal for all realizations of $\mathbf{y}$. At this point, we assume that such value exists. Then this parameter $\delta_{\rm o}$ is clearly deterministic. If we sum (18) for all realizations of $\mathbf{y}$, and a fixed $\delta_{\rm o}$, we end replacing $\mathbf{y}\mathbf{y}^T$ with $\mathbb{E}(\mathbf{y}\mathbf{y}^T)$ which can be expressed using our basic linear model as

$$\begin{aligned}\mathbb{E}\left(\mathbf{y}\mathbf{y}^T\right) &= \mathbf{A}\mathbf{R}_{\mathbf{x}_0}\mathbf{A}^T+\sigma_{\mathbf{z}}^2\mathbf{I}_m\\ &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}\mathbf{\Sigma}\mathbf{U}^T+\sigma_{\mathbf{z}}^2\mathbf{I}_m,\end{aligned} \tag{19}$$

where $\mathbf{R}_{\mathbf{x}_0} \triangleq \mathbb{E}\left(\mathbf{x}_0\mathbf{x}_0^T\right)$ is the covariance matrix of $\mathbf{x}_0$. For a deterministic $\mathbf{x}_0$, $\mathbf{R}_{\mathbf{x}_0} = \mathbf{x}_0\mathbf{x}_0^T$ is used for notational simplicity. Substituting (19) in both terms of (18) results

$$\begin{aligned}N\left(\rho_{\rm o}\right) = &\ \text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho_{\rm o}\mathbf{I}_n\right)^{-2}\mathbf{\Sigma}^2\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}\right)\\ &+\sigma_{\mathbf{z}}^2\text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho_{\rm o}\mathbf{I}_n\right)^{-2}\right),\end{aligned} \tag{20}$$

and

$$\begin{aligned}D\left(\rho_{\rm o}\right) = \delta_{\rm o}^2\Big[&\text{Tr}\left(\left(\mathbf{\Sigma}^2+\rho_{\rm o}\mathbf{I}_n\right)^{-2}\mathbf{\Sigma}^2\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}\right)\\ &+\sigma_{\mathbf{z}}^2\text{Tr}\left(\left(\mathbf{\Sigma}^2+\rho_{\rm o}\mathbf{I}_n\right)^{-2}\right)\Big].\end{aligned} \tag{21}$$

Considering the singular-value structure for the ill-posed problems, we can divide the singular values into two groups of *significant*, or relatively large, and *trivial*, or nearly zero singular value[5]. As an example, we can see from Fig. 1 that the singular values of the ill-posed problem matrix are decaying very fast, making it possible to identify the two groups. Based on this, the matrix $\mathbf{\Sigma}$ can be divided into two diagonal sub-matrices, $\mathbf{\Sigma}_{n1}$, which contains the first (significant) n_1 diagonal entries, and $\mathbf{\Sigma}_{n2}$, which contains the last (trivial) $n_2 = n - n_1$ diagonal entries[6]. As a result, $\mathbf{\Sigma}$ can be written as

$$\mathbf{\Sigma} = \begin{bmatrix}\mathbf{\Sigma}_{n1} & \mathbf{0}\\ \mathbf{0} & \mathbf{\Sigma}_{n2}\end{bmatrix}. \tag{22}$$

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

[6]The splitting threshold is obtained as the mean of the eigenvalues multiplied by a certain constant c, where c $\in (0,1)$.

Similarly, we can 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 write $N\left(\rho_{\rm o}\right)$ in (20) in terms of the partitioned $\mathbf{\Sigma}$ and $\mathbf{V}$ as

$$\begin{aligned}N\left(\rho_{\rm o}\right) = &\ \text{Tr}\left(\mathbf{\Sigma}_{n1}^2\left(\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\mathbf{\Sigma}_{n1}^2\mathbf{V}_{n1}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}_{n1}\right)\\ &+\text{Tr}\left(\mathbf{\Sigma}_{n2}^2\left(\mathbf{\Sigma}_{n2}^2+\rho_{\rm o}\mathbf{I}_{n2}\right)^{-2}\mathbf{\Sigma}_{n2}^2\mathbf{V}_{n2}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}_{n2}\right)\\ &+\sigma_{\mathbf{z}}^2\text{Tr}\left(\mathbf{\Sigma}_{n1}^2\left(\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\right)\\ &+\sigma_{\mathbf{z}}^2\text{Tr}\left(\mathbf{\Sigma}_{n2}^2\left(\mathbf{\Sigma}_{n2}^2+\rho_{\rm o}\mathbf{I}_{n2}\right)^{-2}\right).\end{aligned} \tag{23}$$

Given how n_1 and n_2 are selected, we have $\|\mathbf{\Sigma}_{n2}\| \approx 0$, and so we can approximate $N\left(\rho_{\rm o}\right)$ as

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

Similarly, $D\left(\rho_{\rm o}\right)$ in (21) can be approximated equivalently as

$$\begin{aligned}D\left(\rho_{\rm o}\right) \approx &\ \sigma_{\mathbf{z}}^2\ \text{Tr}\left(\left(\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\right)+\frac{n_2\sigma_{\mathbf{z}}^2}{\rho_{\rm o}^2}\\ &+\text{Tr}\left(\left(\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\mathbf{\Sigma}_{n1}^2\mathbf{V}_{n1}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}_{n1}\right).\end{aligned} \tag{25}$$

Substituting (24) and (25) in (18) and manipulating, we obtain

$$\begin{aligned}\delta_{\rm o}^2 \approx &\Big[\sigma_{\mathbf{z}}^2\text{Tr}\left(\mathbf{\Sigma}_{n1}^2\left(\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\right)\\ &+\text{Tr}\left(\mathbf{\Sigma}_{n1}^2\left(\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\mathbf{\Sigma}_{n1}^2\mathbf{V}_{n1}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}_{n1}\right)\Big]\Big/\\ &\Big[\sigma_{\mathbf{z}}^2\text{Tr}\left(\left(\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\right)+\frac{n_2\sigma_{\mathbf{z}}^2}{\rho_{\rm o}^2}\\ &+\text{Tr}\left(\left(\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\mathbf{\Sigma}_{n1}^2\mathbf{V}_{n1}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}_{n1}\right)\Big].\end{aligned} \tag{26}$$

The bound $\delta_{\rm o}$ in (26) is a function of the unknowns $\rho_{\rm o}$, $\sigma_{\mathbf{z}}^2$, and $\mathbf{R}_{\mathbf{x}_0}$. In fact, estimating $\sigma_{\mathbf{z}}^2$ and $\mathbf{R}_{\mathbf{x}_0}$ without any prior knowledge is a very tedious process. The problem becomes 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 will use the MSE as a criterion to eliminate the dependence of $\delta_{\rm o}$ on these unknowns and a result to set the value of the perturbation bound.

III. Minimizing the MSE for the Solution of the Proposed Perturbation Approach

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

$$\text{MSE} = \mathbb{E}\big[||\hat{\mathbf{x}}-\mathbf{x}_0||^2\big] = \text{Tr}\left(\mathbb{E}\left((\hat{\mathbf{x}}-\mathbf{x}_0)(\hat{\mathbf{x}}-\mathbf{x}_0)^T\right)\right). \tag{27}$$

Since the solution of the proposed approach problem in (8) is given by (9), we can substitute for $\hat{\mathbf{x}}$ from (9) in (27) and then use the SVD of $\mathbf{A}$ to obtain

$$\begin{aligned}\text{MSE}\left(\rho\right) = &\ \sigma_{\mathbf{z}}^2\text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho\mathbf{I}_n\right)^{-2}\right)\\ &+\rho^2\text{Tr}\left(\left(\mathbf{\Sigma}^2+\rho\mathbf{I}_n\right)^{-2}\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}\right).\end{aligned} \tag{28}$$

Theorem 2. *For $\sigma_{\mathbf{z}}^2 > 0$, the approximate value for the optimal regularizer ρ_o of (28) that approximately minimizes the MSE is given by*

$$\rho_o \approx \frac{\sigma_{\mathbf{z}}^2}{Tr\left(\mathbf{R}_{\mathbf{x}_0}\right)/n}. \tag{29}$$

Proof: We can easily prove that the function in (28) is convex in ρ, and hence its global minimizer (i.e., $\rho_{\rm o}$) can be obtained by differentiating (28) with respect to ρ and setting the result to zero, i.e.,

$$\nabla_\rho \text{ MSE}(\rho) = -2\sigma_{\mathbf{z}}^2 \text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho\mathbf{I}_n\right)^{-3}\right) + 2\rho \underbrace{\text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho\mathbf{I}_n\right)^{-3}\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}\right)}_{S} = 0. \quad (30)$$

Equation (30) dictates the relationship between the optimal regularization parameter and the problem parameters. By solving (30), we can obtain the optimal regularizer $\rho_{\rm o}$. However, in the general case, and with lack of knowledge about $\mathbf{R}_{\mathbf{x}_0}$, we cannot obtain a closed-form expression for $\rho_{\rm o}$. As a result, we will seek to obtain a suboptimal regularizer in the MSE sense that minimizes (28) approximately. In what follows, we show how through some bounds and approximations, we can obtain this suboptimal regularizer.

By using the trace inequalities in [[39], eq.(5)], we can bound the second term in (30) by

$$\begin{aligned}&\lambda_{\min}(\mathbf{R}_{\mathbf{x}_0})\,\text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho\mathbf{I}_n\right)^{-3}\right)\\ &\le S = \text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho\mathbf{I}_n\right)^{-3}\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}\right)\\ &\le \lambda_{\max}(\mathbf{R}_{\mathbf{x}_0})\,\text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho\mathbf{I}_n\right)^{-3}\right), \end{aligned} \quad (31)$$

where λ_i is the i'th eigenvalue of $\mathbf{R}_{\mathbf{x}_0}$. 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 $\rho_{\rm o}$, for all (or almost all) possible $\mathbf{A}$, that minimizes the MSE approximately. To achieve this, we consider an *average* value of S based on the inequalities in (31) as our evaluation point, i.e.,

$$S \approx \text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho\mathbf{I}_n\right)^{-3}\right)\frac{\text{Tr}(\mathbf{R}_{\mathbf{x}_0})}{n}. \quad (32)$$

Substituting (32) in (30) yields[7]

$$\nabla_\rho \text{ MSE}(\rho) \approx -2\sigma_{\mathbf{z}}^2 \text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho\mathbf{I}_n\right)^{-3}\right) + 2\rho\frac{\text{Tr}(\mathbf{R}_{\mathbf{x}_0})}{n}\text{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho\mathbf{I}_n\right)^{-3}\right) = 0. \quad (33)$$

Note that the same approximation can be applied from the beginning to the second term in (28) and the same result in (33) will be obtained after taking the derivative of the new approximated MSE function. In Appendix A, we provide the error analysis for this approximation and show that it is bounded in very small feasible region.
Equation (33) can now be solved to obtain $\rho_{\rm o}$ as in (29).[8] ■

Remark 2. The solution in (29) shows that there always exists a positive $\rho_{\rm o}$, for $\sigma_{\mathbf{z}}^2 \neq 0$, which approximately minimizes the MSE in (28). The conclusion that the regularization parameter is generally dependent on the noise variance has been shown before in different contexts (see for example [40], [41]). For the special case where the entries of $\mathbf{x}_0$ are independent and identically distributed (i.i.d.) with zero mean, we have $\mathbf{R}_{\mathbf{x}_0} = \sigma_{\mathbf{x}_0}^2\mathbf{I}_n$. Since the linear minimum mean-squared error (LMMSE) estimator of $\mathbf{x}_0$ in $\mathbf{y} = \mathbf{A}\mathbf{x}_0 + \mathbf{z}$ is defined as [4]

$$\hat{\mathbf{x}}_{\text{LMMSE}} = \left(\mathbf{A}^T\mathbf{A} + \sigma_{\mathbf{z}}^2\mathbf{R}_{\mathbf{x}_0}^{-1}\mathbf{I}_n\right)^{-1}\mathbf{A}^T\mathbf{y}, \quad (34)$$

substituting $\mathbf{R}_{\mathbf{x}_0} = \sigma_{\mathbf{x}_0}^2\mathbf{I}$ makes the LMMSE regularizer in (34) equivalent to $\rho_{\rm o}$ in (29) since $\rho_{\rm o} = \frac{\sigma_{\mathbf{z}}^2}{\text{Tr}(\mathbf{R}_{\mathbf{x}_0})/n} = \frac{\sigma_{\mathbf{z}}^2}{\sigma_{\mathbf{x}_0}^2}$. This shows that (29) is exact when the input is white, while for a general input $\mathbf{x}_0$, the optimum matrix regularizer is given in (34). In other words, the result in (29) provides an approximate optimum scalar regularizer for a general colored input. Note that since $\sigma_{\mathbf{z}}^2$ and $\mathbf{R}_{\mathbf{x}_0}$ are unknowns, $\rho_{\rm o}$ cannot be obtained directly from (29).

We are now ready in the following subsection to use the result in (29) along with the perturbation bound expression in (26) and some reasonable manipulations and approximations to eliminate the dependency of $\delta_{\rm o}$ in (26) on the unknowns $\sigma_{\mathbf{z}}^2$, and $\mathbf{R}_{\mathbf{x}_0}$. Then, we will select a pair of $\delta_{\rm o}$ and $\rho_{\rm o}$ from the space of all possible values of δ and ρ that minimizes the MSE of the proposed estimator solution.

A. Setting the Optimal Perturbation Bound that Minimizes the MSE

We start by applying the same reasoning leading to (32) for both the numerator and the denominator of (26) and manipulate to obtain

$$\begin{aligned}&\text{Tr}\left(\mathbf{\Sigma}_{n1}^2\left(\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\left(\mathbf{\Sigma}_{n1}^2+\frac{n_1\sigma_{\mathbf{z}}^2}{\text{Tr}(\mathbf{R}_{\mathbf{x}_0})}\mathbf{I}_{n1}\right)\right)\\ &\approx \delta_{\rm o}^2\left[\text{Tr}\left(\left(\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\left(\mathbf{\Sigma}_{n1}^2+\frac{n_1\sigma_{\mathbf{z}}^2}{\text{Tr}(\mathbf{R}_{\mathbf{x}_0})}\mathbf{I}_{n1}\right)\right)\right.\\ &\left.+\frac{n_2 n_1\sigma_{\mathbf{z}}^2}{\rho_{\rm o}^2\text{Tr}(\mathbf{R}_{\mathbf{x}_0})}\right]. \end{aligned} \quad (35)$$

In Section V, we verify this approximation using simulations.

Now, we will use the relationship of $\sigma_{\mathbf{z}}^2$ and $\text{Tr}(\mathbf{R}_{\mathbf{x}_0})$ in (29) to the suboptimal regularizer $\frac{n_1}{n}\rho_{\rm o} \approx \frac{n_1\sigma_{\mathbf{z}}^2}{\text{Tr}(\mathbf{R}_{\mathbf{x}_0})}$ to impose a constrain on (35) that makes the selected perturbation bound minimizes the MSE and as a result to make (35) an implicit equation in $\delta_{\rm o}$ and $\rho_{\rm o}$ only. By doing this, we obtain after some algebraic manipulations

$$\delta_{\rm o}^2 \approx \frac{\text{Tr}\left(\mathbf{\Sigma}_{n1}^2\left(\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\left(\frac{n}{n_1}\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)\right)}{\text{Tr}\left(\left(\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\left(\frac{n}{n_1}\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)\right)+\frac{n_2}{\rho_{\rm o}}}. \quad (36)$$

The expression in (36) reveals that any $\delta_{\rm o}$ satisfying (36) minimizes the MSE approximately. Now, we have two equations (17) (evaluated at δ_0 and ρ_0) and (36) in two unknowns δ_0 and ρ_0. Solving these equations and then applying the SVD of $\mathbf{A}$ to the result equation, result in the characteristic equation for the proposed constrained perturbation regularization approach (COPRA) in (37), where $\beta = \frac{n}{n_1}$.

The COPRA characteristic equation in (37) is a function of the problem matrix $\mathbf{A}$, the received signal $\mathbf{y}$, and regularization

[7] Another way to look at (32) is that we can replace $\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}$ inside the trace in (30) by a diagonal matrix $\mathbf{F} = \text{diag}\left(\text{diag}\left(\mathbf{V}^T\mathbf{R}_{\mathbf{x}_0}\mathbf{V}\right)\right)$ without affecting the result. Then, we replace this positive definite matrix $\mathbf{F}$ by an identity matrix multiplied by scalar which is given by the average value of the diagonal entries of $\mathbf{F}$.

[8] In fact, one can prove that when $\mathbf{x}_0$ is i.i.d., (33) and (30) are exactly equivalent to each other (see Appendix A).

$$G(\rho_{\rm o}) = \underbrace{\mathrm{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho_{\rm o}\mathbf{I}_n\right)^{-2}\mathbf{U}^T\mathbf{y}\mathbf{y}^T\mathbf{U}\right)\mathrm{Tr}\left(\left(\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\left(\beta\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)\right)+\frac{n_2}{\rho_{\rm o}}\mathrm{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho_{\rm o}\mathbf{I}_n\right)^{-2}\mathbf{U}^T\mathbf{y}\mathbf{y}^T\mathbf{U}\right)}_{G_1(\rho_{\rm o})}$$
$$-\underbrace{\mathrm{Tr}\left(\left(\mathbf{\Sigma}^2+\rho_{\rm o}\mathbf{I}_n\right)^{-2}\mathbf{U}^T\mathbf{y}\mathbf{y}^T\mathbf{U}\right)\mathrm{Tr}\left(\mathbf{\Sigma}_{n1}^2\left(\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)^{-2}\left(\beta\mathbf{\Sigma}_{n1}^2+\rho_{\rm o}\mathbf{I}_{n1}\right)\right)}_{G_2(\rho_{\rm o})}=0. \tag{37}$$

parameter $\rho_{\rm o}$ which is the only unknown in (37. Solving for $G(\rho_{\rm o}) = 0$ should lead to the regularization parameter $\rho_{\rm o}$ that approximately minimizes the MSE of the estimator. Our main interest then is to find a positive root $\rho_{\rm o}^* > 0$ for (37). In the following section, we study the main properties of this equation and examine the existence and uniqueness of its positive root. Before that, it worth mentioning the following remark

Remark 3. A special case of the proposed COPRA approach is when all the singular values are significant and so no truncation is required (full column rank matrix, see Fig. 1). This is the case where $n_1 = n$, $n_2 = 0$, and $\mathbf{\Sigma}_{n1} = \mathbf{\Sigma}$. Substituting these values in (37) we obtain

$$\begin{aligned}&\bar{G}(\rho_{\rm o}) = \\ &\mathrm{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho_{\rm o}\mathbf{I}_n\right)^{-2}\mathbf{U}^T\mathbf{y}\mathbf{y}^T\mathbf{U}\right)\mathrm{Tr}\left(\left(\mathbf{\Sigma}^2+\rho_{\rm o}\mathbf{I}_n\right)^{-1}\right)\\ &-\mathrm{Tr}\left(\left(\mathbf{\Sigma}^2+\rho_{\rm o}\mathbf{I}_n\right)^{-2}\mathbf{U}^T\mathbf{y}\mathbf{y}^T\mathbf{U}\right)\mathrm{Tr}\left(\mathbf{\Sigma}^2\left(\mathbf{\Sigma}^2+\rho_{\rm o}\mathbf{I}_n\right)^{-1}\right)\\ &=0.\end{aligned} \tag{38}$$

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

In this section, we provide a detailed analysis for the COPRA function $G(\rho_{\rm o})$ in (37). We start by examining some main properties of $G(\rho_{\rm o})$ that are straightforward to proof.

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

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

Property 6. *Each of the functions $G_1(\rho_o)$ and $G_2(\rho_o)$ in (37) is completely monotonic in the interval $(0, +\infty)$.*

Proof: According to [42], [43], a function $F(\rho_{\rm o})$ is completely monotonic if it satisfies

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

where $F^{(n)}(\rho_{\rm o})$ is the n'th derivative of $F(\rho_{\rm o})$.
By continuously differentiating $G_1(\rho_{\rm o})$ and $G_2(\rho_{\rm o})$, we can easily show that both functions satisfy the monotonic condition in (39). ■

Theorem 3. *The COPRA function $G(\rho_o)$ in (37) has at most two roots in the interval $(0, +\infty)$.*

Proof: The proof of Theorem 3 will be conducted in two steps. Firstly, it has been proved in [44], [45] that any completely monotonic function can be approximated as a sum of exponential functions. That is, if $F(\rho_{\rm o})$ is a completely monotonic, it can be approximated as

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

where l is the number of the terms in the sum and l_i and k_i are two constants. It is shown that there always exists a best uniform approximation to $F(\rho_{\rm o})$ where the error in this approximation gets smaller as we increase the number of the terms l. However, our main concern here is the relation given by (40) more than finding the best number of the terms or the unknown parameters l_i and k_i. To conclude, both functions $G_1(\rho_{\rm o})$ and $G_2(\rho_{\rm o})$ in (37) can be approximated by a sum of exponential functions.

Secondly, it is shown in [46] that the sum of exponential functions has at most two intersections with the abscissa. Consequently, and since the relation in (37) can be expressed as a sum of exponential functions, the function $G(\rho_{\rm o})$ has at most two roots in the interval $(0, +\infty)$. ■

Theorem 4. *There exists a sufficiently small positive value ϵ, such that $\epsilon \to 0^+$ and $\epsilon \ll \sigma_i^2$, $\forall i \in [1, n]$ where the value of the COPRA function $G(\rho_o)$ in (37) is zero (i.e., ϵ is a positive root for (37)). However, we are not interested in this root in the proposed COPRA.*

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

Theorem 5. *A sufficient condition for the function $G(\rho_o)$ to approach zero at $\rho_o = +\infty$ from a positive direction 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: Let $\mathbf{b} = \mathbf{U}^T\mathbf{y}$ as in (37). Given that $\mathbf{\Sigma}^2$ is a diagonal matrix, $\mathbf{\Sigma}^2 = \mathrm{diag}\left(\sigma_1^2, \sigma_2^2, \cdots, \sigma_n^2\right)$, 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 (37) by a diagonal matrix $\mathbf{b}\mathbf{b}_d^T$ that contains $\mathbf{b}\mathbf{b}^T$ diagonal entries without affecting the result. By defining $\mathbf{b}\mathbf{b}_d^T =$

diag $\left(b_1^2, b_2^2, \cdots, b_n^2\right)$, we can write (37) as

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

Then, we use some algebraic manipulations to obtain

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

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

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

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

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

It is obvious that the limit in (45) is zero. However, the direction where the limit approaches zero depends on the sign of the term between the square brackets. For $G\left(\rho_{\mathrm{o}}\right)$ to approach zero from the positive direction, knowing that the terms that are independent of τ are the dominants, the following condition must hold:

$$n \left(\sum_{j=1}^{n} \sigma_j^2 b_j^2\right) > \left(\sum_{i=1}^{n_1} \sigma_i^2\right) \left(\sum_{j=1}^{n} b_j^2\right). \tag{46}$$

Which is the same as (41). ■

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

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

From Property 3 and Theorem 4, we can conclude that the COPRA function has a positive value before ϵ, then it switches to the negative region after that. The condition in (41) guarantees that $G\left(\rho_{\mathrm{o}}\right)$ approaches zero from a positive direction as ρ_{o} approaches $+\infty$. This means that $G\left(\rho_{\mathrm{o}}\right)$ 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 $\rho_{\mathrm{o,m}}$, then the function starts increasing for $\rho_{\mathrm{o}} > \rho_{\mathrm{o,m}}$ until it approaches the second zero crossing at $\rho_{\mathrm{o},2}$. Since Theorem 3 states clearly that we cannot have more than two roots, we conclude that when (41) holds, we have only one unique positive root over the interval $(\epsilon, +\infty)$. ■

A. Finding the Root of $G\left(\rho_o\right)$

To find the positive root of the COPRA function $G\left(\rho_{\mathrm{o}}\right)$ in (37), Newton's method [47] can be used. The function $G\left(\rho_{\mathrm{o}}\right)$ is differentiable in the interval $(\epsilon, +\infty)$ and the expression of the first derivative $G^{'}\left(\rho_{\mathrm{o}}\right)$ can be easily obtained. Newton's method can then be applied in a straightforward manner to find this root. Starting from an initial value $\rho_{\mathrm{o}}^{n=0} > \epsilon$ that is sufficiency small, the following iterations are performed:

$$\rho_{\mathrm{o}}^{n+1} = \rho_{\mathrm{o}}^{n} - \frac{G\left(\rho_{\mathrm{o}}^{n}\right)}{G'\left(\rho_{\mathrm{o}}^{n}\right)}. \tag{47}$$

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

B. Convergence

When condition (41) is satisfied, the convergence of Newton's method can be easily proved. As a result from Theorem 5, the function $G\left(\rho_{\mathrm{o}}\right)$ has always a negative value in the interval $(\epsilon, \rho_{o,2})$. It is also clear that $G\left(\rho_{\mathrm{o}}\right)$ is an increasing function in the interval $[\rho_{\mathrm{o}}^{n=0}, \rho_{\mathrm{o},2}]$. Thus, starting from $\rho_{\mathrm{o}}^{n=0}$, (47) will produce a consecutively increasing estimate for ρ_{o}. Convergence occurs when $G\left(\rho_{\mathrm{o}}^{n}\right) \to 0$ and $\rho_{\mathrm{o}}^{n+1} \to \rho_{\mathrm{o}}^{n}$. When the condition in (41) is not satisfied, the regularization parameter should be set to ϵ.

C. COPRA Summary

The proposed COPRA discussed in the previous sections is summarized in Algorithm 1.

V. Numerical Results

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

Three different scenarios of simulation experiments are performed. Firstly, the proposed COPRA is applied to a set of

Algorithm 1 COPRA summary

if (41) is not satisfied **then**
  $\rho_{\text{o}} = \epsilon$.
else
  Define $\tilde{\xi}$ as the iterations stopping criterion.
  Set $\rho_{\text{o}}^{n=0}$ to a sufficiently small positive quantity.
  Find $G\left(\rho_{\text{o}}^{n=0}\right)$ using (37), and compute its derivative $G^{'}\left(\rho_{\text{o}}^{n=0}\right)$.
  while $|G\left(\rho_{\text{o}}^{n}\right)| > \tilde{\xi}$ **do**
    Solve (47) to get ρ_{o}^{n+1}.
    $\rho_{\text{o}}^{n} = \rho_{\text{o}}^{n+1}$.
  end while
end if
Find $\hat{\mathbf{x}}$ using (9).

nine real-world discrete ill-posed problems that are commonly used in testing the performance of regularization methods in discrete ill-posed problems. Secondly, COPRA is used to estimate the signal when $\mathbf{A}$ is a random rank deficient matrix generated as

$$\mathbf{A} = \frac{1}{n}\mathbf{B}\mathbf{B}^{T}, \tag{48}$$

where $\mathbf{B}\left(m \times n, m > n\right)$ is a random matrix with i.i.d. zero mean unit variance Gaussian random entries. This is a theoretical test example which is meant to illustrate the robustness of COPRA and to make sure that the obtained results are applicable for broad class of matrices. Finally, an image restoration in image tomography ill-posed problem is considered[9].

A. Real-World Discrete Ill-posed Problems

The Matlab regularization toolbox [48] is used to generate pairs of a matrix $\mathbf{A} \in \mathbb{R}^{50\times 50}$ and a signal $\mathbf{x}_0$. The toolbox provides many real-world discrete ill-posed problems that can be used to test the performance of regularization methods. The problems are derived from discretization of Fredholm integral equation as in (3) and they arise in many signal processing applications[10].

Experiment setup: The performance of COPRA is compared with three benchmark regularization methods, the quasi-optimal, the GCV, the L-curve in addition to the LS. The performance is evaluated in terms of 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 a certain signal-to-noise-ratio (SNR) defined as SNR $\triangleq \|\mathbf{A}\mathbf{x}_0\|_2^2/n\sigma_{\mathbf{z}}^2$ to generate $\mathbf{y}$. The performance is presented as the NMSE (in dB) (NMSE in dB = $10\log_{10}$ (NMSE)) versus SNR (in dB) 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 for all the selected 9 problems. Each sub-figure quote the condition number (CN) of the problem's matrix. The NMSE curve for some methods disappears in certain cases. This indicates extremely poor performance for these methods such that they are out of scale. For example, LS does not show up in all the tested scenarios, while the other benchmark methods disappear in quite a few cases.

Generally speaking, It can be said that an estimator offering NMSE above 0 dB is not robust and is worthless. From Fig. 2, it is clear that COPRA offers the highest level of robustness among all the methods as it is the only approach whose NMSE performance remains below 0 dB in almost all cases. Comparing the NMSE over the range of the SNR values in each problem, we find *on average* that COPRA exhibits the lowest NMSE amongst all the methods in 8 problems (the first 8 sub-figures). Considering all the problems, the closest contender to COPRA is the quasi method. However, this method and the remaining methods show lack of robustness in certain situations as evident by the extremely high NMSE.

In Fig. 3, we provide the NMSE for the approximation of the perturbation bound expression in (26) by (35) for a selected ill-posed problem matrices. The two expressions are evaluated at each SNR using the suboptimal regularizer in (29). The sub-figures show that the NMSE of the approximation is extremely small (below -20 dB in most cases) and that the error increases as the SNR increases. The increase of the approximation 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 $\|\mathbf{\Sigma}_{n2}\| = 0$ in (22). This theoretical test is meant to illustrate the robustness of COPRA.

Experiment setup: The matrix $\mathbf{A}$ is generated as a random matrix that satisfies $\mathbf{A} = \frac{1}{50}\mathbf{B}\mathbf{B}^{T}$, where $\mathbf{B} \in \mathbb{R}^{50\times 45}, B_{ij} \sim \mathcal{N}\left(0,1\right)$. The elements of $\mathbf{x}_0$ are chosen to be Gaussian i.i.d. with zero mean unit variance, and i.i.d. with uniform distribution in the interval $(0,1)$. Results are obtained as an average over 10^5 different realizations of $\mathbf{A}$, $\mathbf{x}_0$, and $\mathbf{z}$.

From Fig. 4(a), we observe that when the elements of $\mathbf{x}_0$ are Gaussian i.i.d. COPRA outperforms all the benchmark regularization methods. In fact, COPRA is the only approach that provides a NMSE below 0 dB overall the SNR range while other algorithms are providing a very high NMSE. The same behavior can be observed when the elements of $\mathbf{x}_0$ are uniformly distributed as Fig. 4(b) shows. Finally, the performance of the LS is above 250 dB for both cases.

C. Image Restoration

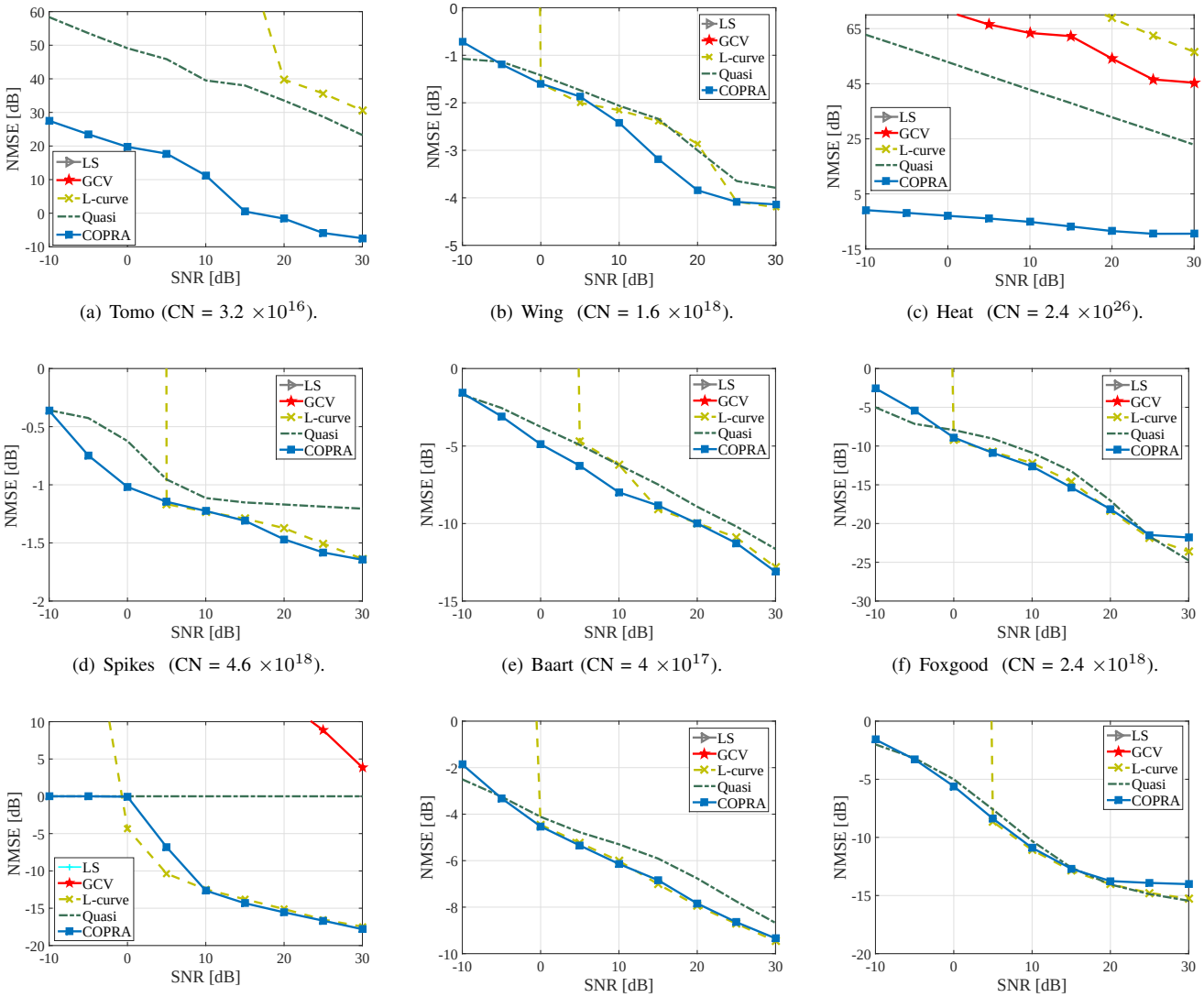
The tomo example in Section V-A and Fig. 2(a) discusses the NMSE of the tomography inverses problem solution. In this subsection, we present visual results for the restored images.

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

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

[9]The MATLAB code of the COPRA is provided at http://faculty.kfupm.edu.sa/ee/naffouri/publications.html.

[10]For more details about the test problems consult [48].



(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] vs SNR [dB] (CN $\equiv$ condition number).

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

In Fig. 5, we present the original image, the received image, and the performance of the methods. Fig. 5 demonstrates that COPRA outperforms all methods through providing a clear image that is very close to the original image. This also appear when we compare the peak signal-to-noise ratio (PSNR) of the algorithms as in Table. I with COPRA having the largest PSNR among them. Moreover, algorithm such GCV provides an unreliable result, while L-curve and quasi fail to restore the internal parts clearly, especially those who have colors close to the image background color.

TABLE I: Algorithms PSNR at SNR=30

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

D. Average Runtime

In Fig. 6, we plot the average runtime for each method against the SNR as calculated in 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 lowest runtime in compare to all benchmark methods.

VI. Conclusion

In this work, we developed a new approach to find the regularization parameter for linear discrete ill-posed problems. Due to the challenging singular-value structure for such problems, many regularization approaches fail to provide a good stabilize solution. In the proposed approach, the singular-value structure of the model matrix is modified by allowing an artificial perturbation into it. To maintain the fidelity of the model, an upper bound constraint on the perturbation is allowed. The proposed approach minimizes the worst-case residual

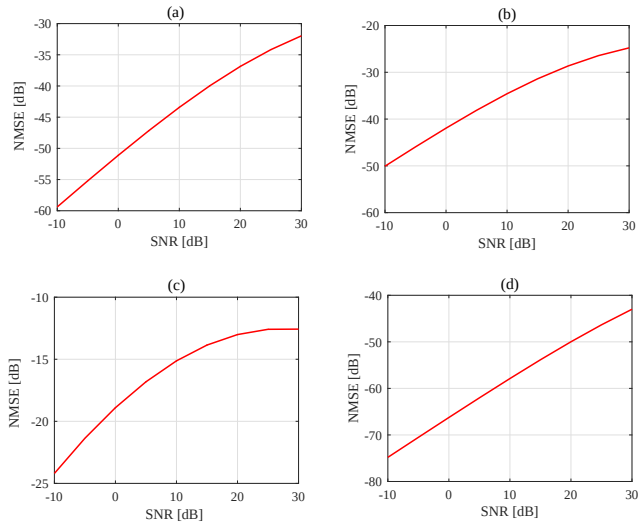
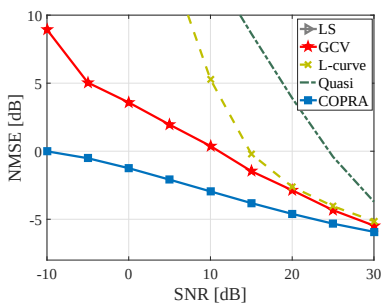
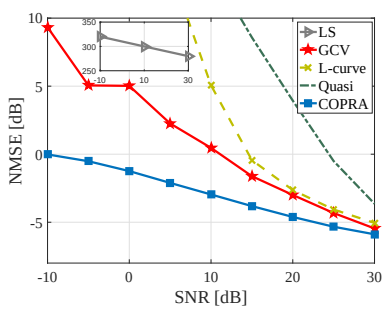


Fig. 3: NMSE [dB] vs SNR [dB] between (26) and (35) for a various $\mathbf{A}$. (a) Wing problem. (b) Heat problem. (c) Foxgood problem. (d) Deriv2 problem.



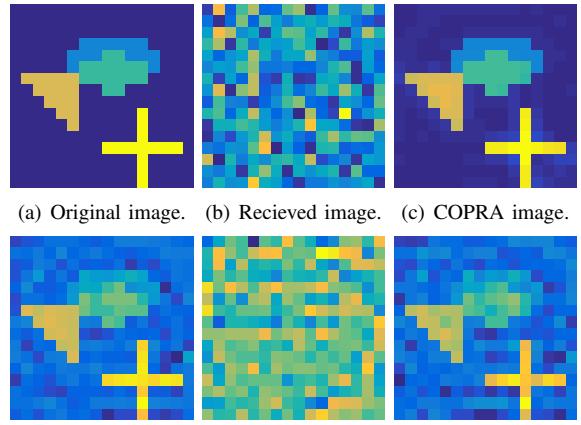
(a) $\mathbf{x}_0 \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ with i.i.d. elements.



(b) The elements of $\mathbf{x}_0$ are i.i.d. with uniform distribution in the interval $(0, 1)$.

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

error of the estimator and selects the perturbation bound in a way that approximately minimizes the MSE. As a result, the approach combines the simplicity of the least-squares criterion with the robustness of the MSE based estimation. The regularization parameter is obtained as a solution of a non-linear equation in one unknown variable. Simulation results demonstrate that the proposed approach outperforms a set of benchmark regularization methods.



(a) Original image. (b) Recieved image. (c) COPRA image.

(d) L-curve image. (e) GCV image. (f) Quasi image.

Fig. 5: Image tomography restoration.

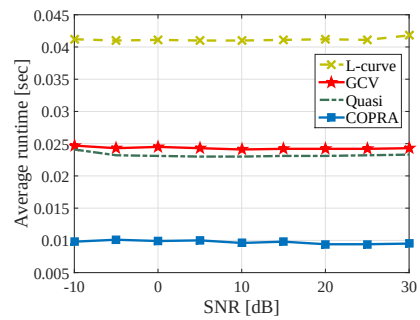


Fig. 6: Average runtime.

Appendix A
Error Analysis

In this appendix, we analyze the error of the approximation that is made to obtain (29). To simpify the analysis, we will consider the case when the approximation is applied directly to the MSE function in (28). Let us start by defining $\mathbf{H} \triangleq \mathbf{\Sigma}^{2k}\left(\mathbf{\Sigma}^2 + \rho \boldsymbol{I}_n\right)^{-p}$, whose diagonal entries can be written as

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

Note that in our case $k = 0$ and $p = 2$ for the diagonal matrix inside the trace function of the second term in (28). However, we will use these two variables to obtain the error expression for the general case, then we will substitute for k and p. By using the inequalities in [[39], eq.(5)], we obtain

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

Similarly, we can write

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

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

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

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 found to be more useful in carrying out the following error analysis. Based on (A.4), we see that $|\varepsilon| \geq 1$ indicates an inaccurate approximation. Although at the end it depends totally on the application, we will 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 will derive two bounds based on (A.2) and (A.3). Then, we will 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 according to

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

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

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

By using (A.1) we can transform (A.7) to

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

Note that the bound μ_x depends only on $\mathbf{R}_{\mathbf{x}_0}$, while μ_a depends both the singular values of $\mathbf{A}$ and the unknown regularizaer 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 a minimum of two independent bounds. We will analyze each bound separately and then we will 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, $\lambda_{\text{min}}(\mathbf{R}_{\mathbf{x}_0}) = \lambda_{\text{avg}}(\mathbf{R}_{\mathbf{x}_0}) = \lambda_{\text{min}}(\mathbf{R}_{\mathbf{x}_0}) = \sigma_{\mathbf{x}_0}^2$, and as a result

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

which means based on (A.10) that the approximation is exact regardless of the contribution of the error from μ_a. When $\mathbf{x}_0$ deviate from being i.i.d., it will be very difficult to obtain a value for μ_x. Therefore, and since no previous knowledge about $\mathbf{x}_0$ is assumed in this paper, it seems that this bound is very loose for a general $\mathbf{x}_0$ and we should rather rely on μ_a to tighten and evaluate the bound of the error as we will discuss. 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 easily prove that the two functions are decreasing in ρ. This means that we can 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 for the low SNR regime: In the extreme low SNR regime, we have $\rho \to \infty$, and therefore, we can obtain the minimum bound on the absolute error. Based on (A.1) we can 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) will boil down after some manipulations to

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

The result in (A.15) indicates that as the SNR decreases, the approximation becomes more accurate. 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 for the high SNR regime: At the extremely high SNR, ρ is sufficiently small and thus doing such analysis allows us to obtain the upper worst case possible value for ε.

Since the main objective of this paper is to provide an estimator that minimizes the MSE, and based on the result proven in [49], there always exists a positive regularizer $\rho > 0$, such that the regularized estimator offers lower MSE than the OLS estimator. This implies also for well-conditioned problems. However, if the condition number is too small, both the regularization parameter and the corresponding improvement in the MSE performance comparing to that of the OLS are too small. Therefore, we conclude that for the extreme high SNR, ρ converges to a minimum value ρ_{min}. In what follow, we find a lower bound expression for ρ_{min} and then examine the absolute error in this value.

Starting from the definition of the 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) then doing some algebraic manipulations, we obtain

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

where $\mathbf{R}_{\mathbf{x}_0} = \mathbf{x}_0\mathbf{x}_0^T$. Now, using (A.17) with the suboptimal regularizer ρ_{o} expression in (29), we can write

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

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

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

Give the nature of the ill-posed problems and there singular values behavior as Fig 1 shows, we will partition $\boldsymbol{\Sigma}$ and $\mathbf{V}$ into to two sub-matrices (same as in Section II-C) and then approximate (A.19) as

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

where $\boldsymbol{\Sigma}_{n1}^2 = \text{diag}\left(\sigma_1^2, \ldots, \sigma_{n1}^2\right)$. 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 will consider the lower bound of ρ_{min} as in (A.21). Moreover, and based on the unitary matrix property and the partitioning of $\mathbf{V}$, we can 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

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

To evaluate (A.24), we will relay on numerical results. Firstly, let us 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 $\rho_{\text{min}}^l = 0.018\sigma_{n1}^2$. Now, substituting the result in (A.24), then evaluating the expression for the 9 ill-posed problems in Section V, we find that $|\varepsilon_a^h| \leq \mu_a^h \approx 1$. As this represents the worst case upper value for the absolute error, we conclude that

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

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

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

The conclusion in (A.26) 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 $\lim_{\rho_{\text{o}}\to\epsilon} G\left(\rho_{\text{o}}\right)$, 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 (37), and by defining $\mathbf{b} = \mathbf{U}^T\mathbf{y}$, we can write

$$\begin{aligned} G\left(\rho_{\text{o}}\right) &= \sum_{i=1}^{n}\frac{\sigma_i^2 b_i^2}{\left(\sigma_i^2+\rho_{\text{o}}\right)^2}\sum_{j=1}^{n_1}\frac{\left(\beta\sigma_j^2+\rho_{\text{o}}\right)}{\left(\sigma_j^2+\rho_{\text{o}}\right)^2} \\ &- \sum_{i=1}^{n}\frac{b_i^2}{\left(\sigma_i^2+\rho_{\text{o}}\right)^2}\sum_{j=1}^{n_1}\frac{\sigma_j^2\left(\beta\sigma_j^2+\rho_{\text{o}}\right)}{\left(\sigma_j^2+\rho_{\text{o}}\right)^2} + \frac{n_2}{\rho_{\text{o}}}\sum_{i=1}^{n}\frac{\sigma_i^2 b_i^2}{\left(\sigma_i^2+\rho_{\text{o}}\right)^2}. \end{aligned} \tag{B.1}$$

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 would like to know if the root define 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 given that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$, we can bound this term as

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

On the other hand, given how we choose n_1 and n_2, we have

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

which can help us to bound the left hand side of (B.4) as

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

Now, from (B.5) and (B.7) we find that a lower bound for the left hand side of (B.4) is equal to the upper bound for 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 (37).

Now, we would like to know if ϵ can be considered as a value for our regularization parameter ρ_o. A direct way to prove that can be noted from the fact that having $\epsilon \ll \sigma_i^2$ $\forall i \in [1, n]$ will not provide any source of regularization to the problem. Hence, the RLS in (6) converges to the LS in (2).

As a remark, we can assume that the approximation in (B.2) is uniform such that it does not affect the position of the roots. Thus, 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 is not coming from the negative region as a result of the 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.

[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] John A Scales and Adam Gersztenkorn, "Robust methods in inverse theory," *Inverse problems*, vol. 4, no. 4, pp. 1071, 1988.

[17] Daniel Scharstein and Richard Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International journal of computer vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[18] Volker Blanz, Albert Mehl, Thomas Vetter, and Hans-Peter Seidel, "A statistical method for robust 3d surface reconstruction from sparse data," in *International Symposium on 3D Data Processing Visualization and Transmission, 2004*. IEEE, 2004, pp. 293–300.

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

[20] 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.

[21] 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.

[22] 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.

[23] C Ray Smith and Walter Grandy, *Maximum-Entropy and bayesian methods in inverse problems*, vol. 14, Springer Science & Business Media, 2013.

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

[25] 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.

[26] 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.

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

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

[29] 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.

[30] 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.

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

[32] Mohamed Suliman, Tarig Ballal, Abla Kammoun, and Tareq Y. Al-Naffouri, "Penalized linear regression for discrete ill-posed problems: A hybrid least-squares and mean-squared error approach," in *Proceedings European Signal Processing Conference (EUSIPCO)*, 2016, pp. 403–407.

[33] 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.

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

[35] 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.

[36] 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.

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

[38] Alexandre Belloni, Victor Chernozhukov, and Lie Wang, "Square-root lasso: pivotal recovery of sparse signals via conic programming," *Biometrika*, vol. 98, no. 4, pp. 791–806, 2011.

[39] 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.

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

[41] William J Hemmerle, "An explicit solution for generalized ridge regression," *Technometrics*, vol. 17, no. 3, pp. 309–314, 1975.

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

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

[44] 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.

[45] 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.

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

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

[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.