ROBUST ESTIMATION IN LINEAR ILL-POSED PROBLEMS WITH ADAPTIVE REGULARIZATION SCHEME

Mohamed A. Suliman, Houssem Sifaou, Tarig Ballal, Mohamed-Slim Alouini, and Tareq Y. Al-Naffouri

CEMSE Division, KAUST, Thuwal, Makkah Province, KSA.

ABSTRACT

In this paper, we propose a new regularized robust estimation approach based on the robust τ-estimator applied to linear ill-posed problems in the presence of noise outliers. Additionally, we introduce a new approach to obtain the optimal regularization parameter for the proposed robust estimator by using tools from random matrix theory. Simulation results demonstrate that the proposed approach with its automated regularization parameter selection outperforms a set of benchmark methods.

***Index Terms*—** Linear inverse problem, regularization, robust estimation, tau estimator.

1. INTRODUCTION

Robust linear estimation has attracted a lot of interest in many fields of engineering such as wireless communication [1], control theory, and computer vision [2]. In this paper, we tackle the problem of recovering a vector of coefficients $\mathbf{x} \in \mathbb{R}^K$ from an observation vector $\mathbf{y} \in \mathbb{R}^M$ related to $\mathbf{x}$ through

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

where $\mathbf{H} \in \mathbb{R}^{M\times K}$ is a known measurement matrix that has independent and identically distributed (i.i.d.) Gaussian entries and $\mathbf{z} \in \mathbb{R}^M$ is an unknown noise vector. The entries of $\mathbf{x}$ are i.i.d. from a certain distribution of zero mean and variance $\sigma^2_{\mathbf{x}}$. We focus on the case where the problem based on (1) is ill-posed and the noise $\mathbf{z}$ is subject to the occurrence of outliers.

Owing to the lack of prior knowledge on $\mathbf{x}$, the least-square (LS) estimator, which is based on minimizing the norm of the residual error $||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2$, is normally used. The LS estimator is the best linear unbiased estimator when the noise $\mathbf{z}$ has i.i.d. Gaussian entries. However, the LS estimator is known to be very sensitive to the occurrence of outliers in $\mathbf{z}$. Another difficulty associated with the LS estimator is when the problem is ill-posed. The LS solution for this category of problems may not exist, is not unique, and/or does not depend continuously on the initial data (i.e., unstable) [3].

A common way to overcome the outliers effect on the LS estimator is to replace the norm function by a slowly increasing function that absorbs the large residuals influence. This class of estimators is referred to as robust estimators [4]; among them are the M-estimators [5], the S-estimators [6], the MM-estimators, and the τ-estimators [7]. The performance of robust estimators is measured by their breakdown point (BP) which refers to the proportion of outliers that the estimator can handle before the outlier effect overwhelms the model [4].

The S-estimators and the MM-estimators are developed to provide higher breakdown points. The MM-estimators is known to be more efficient than the S-estimators. The τ-estimators enjoy the robustness and the efficiency of the MM-estimators in addition to having lower maximum bias curves.

Regularization methods are used to overcome the difficulties associated with ill-posed problems [3, 8, 9]. The most common form of regularization is Tikhonov which is given in its simplified form by

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}} \{ ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2^2 + \lambda\, ||\mathbf{x}||_2^2 \}, \tag{2}$$

where λ is a regularization parameter. In [10, 11, 12], the authors derived a new regularization approach for linear ill-posed problems, and in [13] for linear problems with Gaussian random matrices that are based on enhancing the singular-value structure of $\mathbf{H}$. The problem tends to be formulated as

$$\hat{\mathbf{x}} := \arg\min_{\mathbf{x}} \{ ||\mathbf{y} - \mathbf{H}\mathbf{x}||_2 + \gamma\,(\mathbf{H}, \mathbf{y})\ ||\mathbf{x}||_2 \}, \tag{3}$$

where γ is a function of both $\mathbf{y}$ and $\mathbf{H}$. These estimators are shown to outperform many benchmark regularization methods in terms of the mean-squared error (MSE). However, they are derived based on the assumption that $\mathbf{z}$ has i.i.d. entries without outliers.

Combining robust estimators with regularization leads to robust regularized estimators such as the regularized M-estimators, S-estimators [14], and MM-estimators [15]. Recently, in [16], a new regularized τ-estimator is derived and shown to be the most efficient regularized estimator with higher BP than the other estimators.

One of the major issues associated with robust regularized estimators is the selection of the regularization parameter. Several robust estimators obtain this parameter by using exhaustive search algorithm [16, 17, 18], which is computationally intractable. Other algorithms use the generalized cross-validation (GCV) [19], which does not necessarily produce an optimal (or close to optimal) regularizer since it assumes Gaussian zero mean noise.

In this paper, we develop a new robust regularized estimator by modifying the cost function in [10, 11, 12], (i.e., (3)) to handle the presence of noise outliers. Moreover, we introduce a new approach to obtain the optimal regularization parameter in a way that minimizes the MSE of the estimator by using tools from random matrix theory (RMT). This regularized estimator is shown to offer remarkable performance enhancement in addition to significant computational complexity reduction.

2. THE PROPOSED REGULARIZED τ-ESTIMATOR

Given a set of residuals $r_i = y_i - \hat{\mathbf{h}}_i^T \mathbf{x}; i = 1, \dots, M$, where $\hat{\mathbf{h}}_i^T \in \mathbb{R}^{1\times K}$ is the i-th row of $\mathbf{H}$, the M-scale estimate $s_{\mathrm{M}}\,(\mathbf{r}\,(\mathbf{x}))$ is obtained by [5]

$$\frac{1}{M}\sum_{i=1}^{M} \rho_1 \left(\frac{r_i\,(\mathbf{x})}{s_{\mathrm{M}}\,(\mathbf{r}\,(\mathbf{x}))} \right) = b_1, \tag{4}$$

This work was funded by a CRG3 grant ORS# 2221 from the Office of Competitive Research at King Abdullah University of Science and Technology (KAUST).

where $\rho_1(\cdot)$ is even, bounded, non-decreasing in $[0,\infty)$, and differentiable function. The constant $b_1 = \mathbb{E}_{\mathbf{\Phi}}(\rho_1(\mathbf{r}(\mathbf{x})))$, where $\mathbf{\Phi}$ is the standard normal distribution. Based on (4), the τ-scale of $\mathbf{r}(\mathbf{x})$ is defined as [7]

$$\tau^2(\mathbf{r}(\mathbf{x})) = s_{\mathrm{M}}^2(\mathbf{r}(\mathbf{x}))\frac{1}{M}\sum_{i=1}^{M}\rho_2\left(\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}\right), \tag{5}$$

where the function $\rho_2(\cdot)$ determines the estimator efficiency while $\rho_1(\cdot)$ determines its BP [7].

2.1. Problem Formulation and Solution

As mentioned in Section 1, the proposed regularized τ-estimator is the robust version of the estimators in [10, 11, 12]. Based on (3), we formulate the proposed regularized τ-estimator as

$$\hat{\mathbf{x}} := \underset{\mathbf{x}\in\mathbb{R}^K}{\arg\min}\, \tau_{\text{reg}}(\mathbf{r}(\mathbf{x})) = \underset{\mathbf{x}\in\mathbb{R}^K}{\arg\min}\left\{\tau(\mathbf{r}(\mathbf{x})) + \lambda\,||\mathbf{x}||_2\right\}. \tag{6}$$

Comparing (6) to the cost function in [16], we can see that the two functions are mathematically inequivalent since in (6) the two terms of the cost function are not squared. The cost function in (6) is non-convex and may have multiple local minima. In the following, we obtain an implicit expression for its local minimum, then we discuss how to obtain a solution.

Theorem 1. *The penalized τ-estimation in (6) is given by*

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

where

$$\gamma(\mathbf{y},\mathbf{H},\hat{\mathbf{x}}) = \frac{2MQ(\hat{\mathbf{x}})\lambda}{s_M(\mathbf{r}(\hat{\mathbf{x}}))\,\kappa(\hat{\mathbf{x}})\,||\hat{\mathbf{x}}||_2}, \tag{8}$$

$$\kappa(\hat{\mathbf{x}}) = \frac{1}{\mathbf{r}(\hat{\mathbf{x}})^T\mathbf{F}_1(\hat{\mathbf{x}})\mathbf{r}(\hat{\mathbf{x}})}\left[2MQ^2(\hat{\mathbf{x}}) - \mathbf{r}(\hat{\mathbf{x}})^T\mathbf{F}_2(\hat{\mathbf{x}})\mathbf{r}(\hat{\mathbf{x}})\right]. \tag{9}$$

The scalar $Q(\hat{\mathbf{x}}) = \left[\frac{1}{M}\sum_{i=1}^{M}\rho_2\left(\frac{r_i(\hat{\mathbf{x}})}{s_M(\mathbf{r}(\hat{\mathbf{x}}))}\right)\right]^{\frac{1}{2}}$ *while*

$$\mathbf{F}_t(\hat{\mathbf{x}}) = \mathit{diag}\left(F_t^i(\hat{\mathbf{x}})\right) \in \mathbb{R}^{M\times M}, t = 1, 2. \tag{10}$$

$$F_t^i(\hat{\mathbf{x}}) = \rho_t^{(1)}\left(\frac{r_i(\hat{\mathbf{x}})}{s_M(\mathbf{r}(\hat{\mathbf{x}}))}\right)\frac{1}{r_i(\hat{\mathbf{x}})\,s_M(\mathbf{r}(\hat{\mathbf{x}}))};\ i = 1,\ldots,M. \tag{11}$$

The diagonal matrix $\mathbf{\Sigma}$ is given by

$$\mathbf{\Sigma} = \mathbf{F}_1 + \frac{1}{\kappa(\hat{\mathbf{x}})}\mathbf{F}_2. \tag{12}$$

Proof. The first derivative of the cost function in (6) is

$$\begin{aligned}\nabla_{\mathbf{x}}\tau_{\text{reg}}(\mathbf{r}(\mathbf{x})) &= \nabla_{\mathbf{x}}s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))\left[\frac{1}{M}\sum_{i=1}^{M}\rho_2\left(\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}\right)\right]^{\frac{1}{2}}\\ &+ \frac{\sum_{i=1}^{M}\rho_2^{(1)}\left(\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}\right)\left(\frac{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))\hat{\mathbf{h}}_i - r_i(\mathbf{x})\nabla_{\mathbf{x}}s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))^2}\right)}{2M\left[\frac{1}{M}\sum_{i=1}^{M}\rho_2\left(\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}\right)\right]^{\frac{1}{2}}}s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))\\ &+ \frac{\lambda}{||\mathbf{x}||_2}\mathbf{x}.\end{aligned} \tag{13}$$

By taking the derivative of (4) w.r.t. $\mathbf{x}$, we obtain

$$\nabla_{\mathbf{x}}s_{\mathrm{M}}(\mathbf{r}(\mathbf{x})) = \frac{-\sum_{i=1}^{M}\rho_1^{(1)}\left(\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}\right)\hat{\mathbf{h}}_i}{\sum_{i=1}^{M}\rho_1^{(1)}\left(\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}\right)\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}}. \tag{14}$$

Substituting (14) in (13) and using $Q(\mathbf{x})$ yields

$$\begin{aligned}\nabla_{\mathbf{x}}\tau_{\text{reg}}(\mathbf{r}(\mathbf{x})) &= \frac{-2MQ^2(\mathbf{x})\sum_{i=1}^{M}\rho_1^{(1)}\left(\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}\right)\hat{\mathbf{h}}_i}{\sum_{i=1}^{M}\rho_1^{(1)}\left(\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}\right)\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}}\\ &+ \frac{\sum_{i=1}^{M}\rho_2^{(1)}\left(\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}\right)\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}\sum_{i=1}^{M}\rho_1^{(1)}\left(\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}\right)\hat{\mathbf{h}}_i}{\sum_{i=1}^{M}\rho_1^{(1)}\left(\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}\right)\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}}\\ &- \sum_{i=1}^{M}\rho_2^{(1)}\left(\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}\right)\hat{\mathbf{h}}_i + \frac{2MQ(\mathbf{x})\lambda}{||\mathbf{x}||_2}\mathbf{x}.\end{aligned} \tag{15}$$

Now, based on the definition in (10), we can write

$$\sum_{i=1}^{M}\rho_t^{(1)}\left(\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}\right)\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))} = \mathbf{r}(\mathbf{x})^T\mathbf{F}_t(\mathbf{x})\mathbf{r}(\mathbf{x}), \tag{16}$$

$$\sum_{i=1}^{M}\rho_t^{(1)}\left(\frac{r_i(\mathbf{x})}{s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))}\right)\hat{\mathbf{h}}_i = s_{\mathrm{M}}(\mathbf{r}(\mathbf{x}))\mathbf{H}^T\mathbf{F}_t(\mathbf{x})\mathbf{r}(\mathbf{x}). \tag{17}$$

Substituting (16) and (17) in (15), then using $\kappa(\mathbf{x})$ as in (9) with $\mathbf{r}(\mathbf{x}) = \mathbf{y} - \mathbf{H}\mathbf{x}$, and finally solving $\nabla_{\mathbf{x}}\tau_{\text{reg}}(\mathbf{r}(\hat{\mathbf{x}})) = 0$, results in

$$\begin{aligned}&\left(\mathbf{H}^T\left(\mathbf{F}_1(\hat{\mathbf{x}}) + \frac{1}{\kappa(\hat{\mathbf{x}})}\mathbf{F}_2(\hat{\mathbf{x}})\right)\mathbf{H} + \gamma(\hat{\mathbf{x}})\mathbf{I}_{\mathrm{K}}\right)\hat{\mathbf{x}}\\ &\quad -\mathbf{H}^T\left(\mathbf{F}_1(\hat{\mathbf{x}}) + \frac{1}{\kappa(\hat{\mathbf{x}})}\mathbf{F}_2(\hat{\mathbf{x}})\right)\mathbf{y} = 0,\end{aligned} \tag{18}$$

which leads directly to the expression in (7). □

Since (7) represents a local minimum of (6), different starting points $\mathbf{x}_0$ converge to various local minima.

3. OBTAINING THE REGULARIZATION PARAMETER

In this section, we use tools from RMT to obtain λ in (6). Our goal here is to avoid the exhaustive search that robust estimators normally use to find λ and to obtain a regularizer that minimizes the MSE of the estimator in (6). We start by writing the solution in (7) as

$$\hat{\mathbf{x}} = \left(\mathbf{H}^T\mathbf{R}\mathbf{H} + \lambda\,\mathbf{I}_{\mathrm{K}}\right)^{-1}\mathbf{H}^T\mathbf{R}\mathbf{y}, \tag{19}$$

where

$$\mathbf{R} = \frac{s_{\mathrm{M}}(\mathbf{r}(\hat{\mathbf{x}}))\,\kappa(\hat{\mathbf{x}})\,||\hat{\mathbf{x}}||_2}{2MQ(\hat{\mathbf{x}})}\mathbf{\Sigma}. \tag{20}$$

The formulation in (19) suggests that we can express the local minimum of (6) as the minimizer of the convex optimization

$$\hat{\mathbf{x}} := \underset{\mathbf{x}\in\mathbb{R}^K}{\arg\min}\left\{||\mathbf{R}^{\frac{1}{2}}(\mathbf{y} - \mathbf{H}\mathbf{x})||_2^2 + \lambda\,||\mathbf{x}||_2^2\right\}. \tag{21}$$

The regularizer λ in (6) and (21) must be chosen judiciously to achieve high estimator accuracy. We start our derivation by stating our basic assumptions on (1) and (21).

Assumption 1. Let $\mathbf{H} \in \mathbb{R}^{M \times K}$ have i.i.d. entries with $[\mathbf{H}]_{i,j} \sim \mathcal{N}(0,1)$, and let $\mathbf{R}$ be a deterministic uniformly bounded real diagonal matrix of size $M \times M$ which does not have any fluctuations.

Assumption 2. Consider the linear asymptotic regime where M and K grow to infinity with $M/K \to \rho \in (0, \infty)$.

Assumption 3. We assume that the entries of $\mathbf{x}$ are i.i.d. from a certain distribution density (not necessarily known) of zero mean and unknown variance $\sigma_{\mathbf{x}}^2$ and that the noise vector $\mathbf{z}$ is modeled as

$$\mathbf{z} = \mathbf{z}_{\text{g}} + \mathbf{z}_{\text{s}}, \tag{22}$$

where the entries of $\mathbf{z}_{\text{g}}$ are i.i.d. from a certain distribution of zero mean and unknown variance $\sigma_{\mathbf{z}_{\text{g}}}^2$. The vector $\mathbf{z}_{\text{s}}$ has $p \ll M$ non-zero i.i.d. elements that represent the sparse noise and can be drawn from any distribution of zero mean and unknown variance $\sigma_{\mathbf{z}_{\text{s}}}^2$. Finally, we assume that $\mathbf{z}_{\text{g}}$ and $\mathbf{z}_{\text{s}}$ are independent and that the variance of $\mathbf{z}$ is given by $\sigma_{\mathbf{z}}^2$.

Based on Assumption 3, the optimal regularizer λ_{optimal} that minimizes the MSE of (6) is given by [20, 21]

$$\lambda_{\text{optimal}} = \frac{1}{\text{SNR}} = \frac{\sigma_{\mathbf{z}}^2}{\sigma_{\mathbf{x}}^2}. \tag{23}$$

Since both $\sigma_{\mathbf{x}}^2$ and $\sigma_{\mathbf{z}}^2$ are unknowns, our main goal will be to obtain good estimates of $\sigma_{\mathbf{x}}^2$ and $\sigma_{\mathbf{z}}^2$.

Theorem 2. *Under the settings of Assumptions 1, 2, and 3, and by considering the following function*

$$\mathbb{E}\left[\Psi(\mathbf{H})\right] = \mathbb{E}\Big[\frac{1}{K}||\mathbf{R}^{\frac{1}{2}}\left(\mathbf{y} - \mathbf{H}\hat{\mathbf{x}}\right)||_2^2 + \frac{\lambda}{K}\,||\hat{\mathbf{x}}||_2^2\Big], \tag{24}$$

there exists a deterministic function $\alpha(t)$ defined as

$$\alpha(t) = \frac{Tr\left(\mathbf{R}\mathbf{T}(t)\right)}{K\left(1 + t\delta(t)\right)}\sigma_{\mathbf{x}}^2 + \left[\frac{Tr\left(\mathbf{R}\right)}{K} - \frac{t\,Tr\left(\mathbf{R}^2\mathbf{T}(t)\right)}{K\left(1+t\delta(t)\right)}\right]\sigma_{\mathbf{z}}^2 + \mathcal{O}\left(K^{-1}\right), \tag{25}$$

such that

$$\mathop{\mathbb{E}}_{\mathbf{x},\mathbf{z}}\left[\Psi(\mathbf{H})\right] - \alpha(t) \xrightarrow{a.s.} 0, \tag{26}$$

where $t = \frac{K}{\lambda}$, "$\xrightarrow{a.s.}$" denotes the almost sure convergence, and the notation $q = \mathcal{O}\left(K^{-1}\right)$ indicates that $|\frac{q}{K^{-1}}|$ is bounded as $K \to \infty$. Let $\mathbf{D}$ be the diagonal matrix that contains the eigenvalues of $\mathbf{R}$, i.e., $\mathbf{R} = \mathbf{U}\mathbf{D}\mathbf{U}^T$, then

$$\mathbf{T}(t) = \mathbf{U}\left(\mathbf{I}_M + \frac{t}{\left(1+t\delta(t)\right)}\mathbf{D}\right)^{-1}\mathbf{U}^T. \tag{27}$$

Finally, the variable $\delta(t)$ is defined as the unique positive solution of the following fixed-point equation

$$\delta(t) = \frac{1}{K}Tr\left(\mathbf{D}\left(\mathbf{I}_M + \frac{t}{1+t\,\delta(t)}\mathbf{D}\right)^{-1}\right). \tag{28}$$

Proof. Let us start by evaluating the cost function in (21) at its optimal solution in (19). By doing so, we obtain

$$||\mathbf{R}^{\frac{1}{2}}\left(\mathbf{y}-\mathbf{H}\hat{\mathbf{x}}\right)||_2^2 + \lambda\,||\hat{\mathbf{x}}||_2^2 = \mathbf{y}^T\mathbf{R}\mathbf{y} - 2\mathbf{y}^T\mathbf{R}\mathbf{H}\mathbf{G}\mathbf{H}^T\mathbf{R}\mathbf{y} + \mathbf{y}^T\mathbf{R}\mathbf{H}\mathbf{G}\mathbf{H}^T\mathbf{R}\mathbf{H}\mathbf{G}\mathbf{H}^T\mathbf{R}\mathbf{y} + \lambda\mathbf{y}^T\mathbf{R}\mathbf{H}\mathbf{G}^2\mathbf{H}^T\mathbf{R}\mathbf{y}, \tag{29}$$

where $\mathbf{G} \triangleq \left(\mathbf{H}^T\mathbf{R}\mathbf{H} + \lambda\mathbf{I}_{\text{K}}\right)^{-1}$. Now, by taking the expected value of (25) over $\mathbf{H}, \mathbf{x}$, and $\mathbf{z}$, and based on Assumption 3, we can express the first term in (24) using (1) and (29) as

$$\begin{aligned}&\mathbb{E}_{\mathbf{H},\mathbf{x},\mathbf{z}}\Big[\frac{1}{K}||\mathbf{R}^{\frac{1}{2}}\left(\mathbf{y}-\mathbf{H}\hat{\mathbf{x}}\right)||_2^2\Big] = \\ &\frac{\sigma_{\mathbf{x}}^2\lambda^2}{K}\,\mathbb{E}_{\mathbf{H}}\Big[\text{Tr}\left(\mathbf{H}^T\mathbf{R}\mathbf{H}\left(\mathbf{H}^T\mathbf{R}\mathbf{H}+\lambda\mathbf{I}_{\text{K}}\right)^{-2}\right)\Big] \\ &+ \frac{\sigma_{\mathbf{z}}^2}{K}\mathbb{E}_{\mathbf{H}}\Big[\text{Tr}\left(\mathbf{R}\right) - \text{Tr}\left(\mathbf{H}^T\mathbf{R}^2\mathbf{H}\left(\mathbf{H}^T\mathbf{R}\mathbf{H}+\lambda\mathbf{I}_{\text{K}}\right)^{-1}\right) \\ &- \lambda\text{Tr}\left(\mathbf{H}^T\mathbf{R}^2\mathbf{H}\left(\mathbf{H}^T\mathbf{R}\mathbf{H}+\lambda\mathbf{I}_{\text{K}}\right)^{-2}\right)\Big].\end{aligned} \tag{30}$$

Given that $\left(\mathbf{H}^T\mathbf{H}+\mathbf{I}_{\text{K}}\right)^{-1}\mathbf{H} = \mathbf{H}\left(\mathbf{H}\mathbf{H}^T+\mathbf{I}_{\text{M}}\right)^{-1}$ and by defining $\widehat{\mathbf{H}} \triangleq \mathbf{R}^{\frac{1}{2}}\mathbf{H}$, we can write (30) as

$$\begin{aligned}&\mathbb{E}_{\mathbf{H},\mathbf{x},\mathbf{z}}\Big[\frac{1}{K}||\mathbf{R}^{\frac{1}{2}}\left(\mathbf{y}-\mathbf{H}\hat{\mathbf{x}}\right)||_2^2\Big] = \frac{\sigma_{\mathbf{z}}^2}{K}\,\mathbb{E}_{\mathbf{H}}\Bigg[\text{Tr}\left(\mathbf{R}\right) \\ &- t\,\text{Tr}\left(\mathbf{R}\left(\frac{t}{K}\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T+\mathbf{I}_{\text{M}}\right)^{-1}\frac{\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T}{K}\right) \\ &- t\,\text{Tr}\left(\mathbf{R}\left(\frac{t}{K}\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T+\mathbf{I}_{\text{M}}\right)^{-2}\frac{\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T}{K}\right)\Bigg] \\ &+ \frac{\sigma_{\mathbf{x}}^2}{K}\,\mathbb{E}_{\mathbf{H}}\Bigg[\text{Tr}\left(\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T\left(\frac{t}{K}\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T+\mathbf{I}_{\text{M}}\right)^{-2}\right)\Bigg].\end{aligned} \tag{31}$$

Following the same procedure, we can obtain the corresponding expression for the second term in (24). Then, we can prove that

$$\begin{aligned}&\mathbb{E}_{\mathbf{H},\mathbf{x},\mathbf{z}}\Big[\frac{1}{K}||\mathbf{R}^{\frac{1}{2}}\left(\mathbf{y}-\mathbf{H}\hat{\mathbf{x}}\right)||_2^2 + \frac{\lambda}{K}\,||\hat{\mathbf{x}}||_2^2\Big] = \frac{\sigma_{\mathbf{z}}^2}{K}\,\mathbb{E}_{\mathbf{H}}\Bigg[\text{Tr}\left(\mathbf{R}\right) \\ &- t\,\text{Tr}\left(\mathbf{R}\left(\frac{t}{K}\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T+\mathbf{I}_{\text{M}}\right)^{-1}\frac{\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T}{K}\right)\Bigg] \\ &+ \frac{\sigma_{\mathbf{x}}^2}{K}\mathbb{E}_{\mathbf{H}}\Bigg[\text{Tr}\left(\left(\frac{t}{K}\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T+\mathbf{I}_{\text{M}}\right)^{-1}\frac{\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T}{K}\right)\Bigg].\end{aligned} \tag{32}$$

Now, based on the result obtained in [22] (Equation (23)), and after some algebraic manipulations, we can prove that

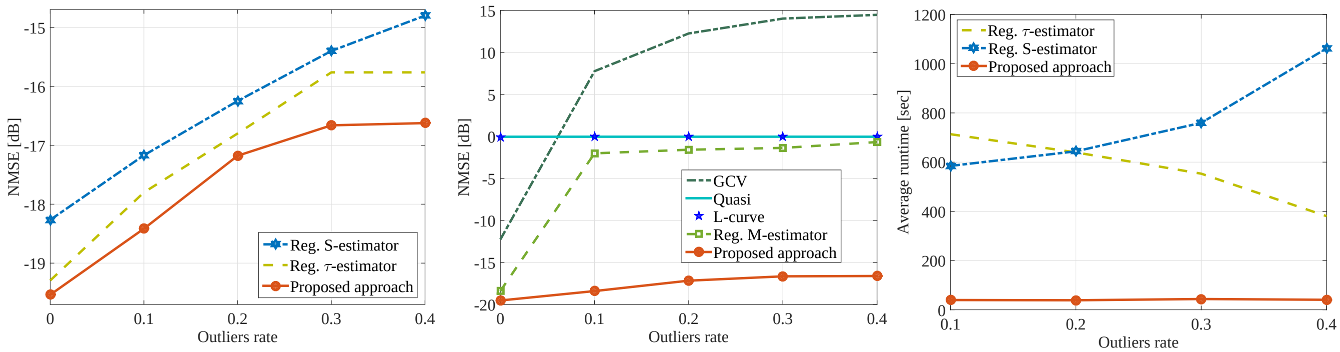
$$\mathbb{E}_{\mathbf{H}}\left[\frac{1}{K}\text{Tr}\left(\mathbf{R}\left(\frac{t}{K}\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T+\mathbf{I}_{\text{M}}\right)^{-1}\frac{\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T}{K}\right)\right] = \frac{1}{K}\frac{\text{Tr}\left(\mathbf{R}^2\mathbf{T}(t)\right)}{\left(1+t\delta(t)\right)} + \mathcal{O}\left(K^{-1}\right). \tag{33}$$

$$\mathbb{E}_{\mathbf{H}}\left[\frac{1}{K}\text{Tr}\left(\left(\frac{t}{K}\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T+\mathbf{I}_{\text{M}}\right)^{-1}\frac{\widehat{\mathbf{H}}\widehat{\mathbf{H}}^T}{K}\right)\right] = \frac{\text{Tr}\left(\mathbf{R}\mathbf{T}(t)\right)}{K\left(1+t\,\delta(t)\right)} + \mathcal{O}\left(K^{-1}\right). \tag{34}$$

By substituting (33) and (34) in (32), we can obtain (25). From (32), (33), (34), and (25), we can conclude that

$$\mathop{\mathbb{E}}_{\mathbf{H},\mathbf{x},\mathbf{z}}\left[\Psi(\mathbf{H})\right] - \alpha(t) \to 0. \tag{35}$$

However, based on the result obtained in [22], the variance of the terms inside the two expectations in (33) and (34) is $\mathcal{O}\left(K^{-2}\right)$.



(a) Perfromance comparision. (b) Perfromance comparision. (c) Average runtime of the superior methods.

Fig. 1. Performance comparison and the average runtime of the proposed robust estimator compared to other robust and non-robust methods.

Based on this fact and equation (35), and by using Borel-Cantelli lemma [23], we can easily prove that

$$\mathop{\mathbb{E}}_{\mathbf{H},\mathbf{x},\mathbf{z}}\left[\Psi\left(\mathbf{H}\right)\right] - \mathop{\mathbb{E}}_{\mathbf{x},\mathbf{z}}\left[\Psi\left(\mathbf{H}\right)\right] \xrightarrow{\text{a.s.}} 0. \tag{36}$$

Finally, based on (35) and (36), we can obtain (26). $\square$

3.1. Using Theorem 2 to obtain λ

The result in Theorem 2 indicates that the average value of the cost function in (21) converges to (25) at $\mathbf{x} = \hat{\mathbf{x}}$. At a certain value of (21), we can write

$$\begin{bmatrix} \frac{\operatorname{Tr}(\mathbf{R}\mathbf{T}(t))}{K(1+t\delta(t))} & \frac{\operatorname{Tr}(\mathbf{R})}{K} - \frac{t\operatorname{Tr}\left(\mathbf{R}^2\mathbf{T}(t)\right)}{K(1+t\delta(t))} \end{bmatrix} \begin{bmatrix} \sigma_{\mathbf{x}}^2 \\ \sigma_{\mathbf{z}}^2 \end{bmatrix} + \epsilon = \left[\frac{1}{K}\|\mathbf{R}^{\frac{1}{2}}\left(\mathbf{y}-\mathbf{H}\hat{\mathbf{x}}\right)\|_2^2 + \frac{\lambda}{K}\|\hat{\mathbf{x}}\|_2^2\right], \tag{37}$$

where ϵ is an approximation error. Evaluating (37) at multiple $\lambda_i, i = 1, \ldots, n$, yields

$$\begin{bmatrix} \Gamma_1(\lambda_1) & \Gamma_2(\lambda_1) \\ \vdots & \vdots \\ \Gamma_1(\lambda_n) & \Gamma_2(\lambda_n) \end{bmatrix} \begin{bmatrix} \sigma_{\mathbf{x}}^2 \\ \sigma_{\mathbf{z}}^2 \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_n \end{bmatrix} = \begin{bmatrix} \psi(\lambda_1) \\ \vdots \\ \psi(\lambda_n) \end{bmatrix} \Rightarrow \boldsymbol{\Gamma}\boldsymbol{\sigma} + \boldsymbol{\epsilon} = \boldsymbol{\psi}. \tag{38}$$

Now, by solving the following constrained linear LS

$$\min_{\boldsymbol{\sigma}} \frac{1}{2}\|\boldsymbol{\psi} - \boldsymbol{\Gamma}\boldsymbol{\sigma}\|_2^2 \quad \text{subject to } \boldsymbol{\sigma} \geq \mathbf{0}, \tag{39}$$

we can obtain estimates $\hat{\sigma}_{\mathbf{x}}^2$ and $\hat{\sigma}_{\mathbf{z}}^2$, given that $n \geq 2$.

Remark 1. The error vector $\boldsymbol{\epsilon}$ is due to the fact that we are equating the cost function in (21) evaluated at its optimal solution $\hat{\mathbf{x}}$ with its average value. This will be accurate for high dimensions but becomes less so as we decrease the dimensions of the problem. Our main goal here is to facilitates the process of obtaining a pair $\left(\hat{\sigma}_{\mathbf{x}}^2, \hat{\sigma}_{\mathbf{z}}^2\right)$ that closely approximate the signal and the noise statistics.

4. NUMERICAL RESULTS

In this section, we demonstrate the performance of the proposed approach using numerical simulations. A matrix $\mathbf{H} \in \mathbb{R}^{300\times100}$ with i.i.d. entries $\mathcal{N}(0,1)$ and a condition number equal to 10^3 is generated. The elements of $\mathbf{x}$ are chosen to be i.i.d. with $[\mathbf{x}]_i \sim \mathcal{N}(0,1)$. The noise vector $\mathbf{z}$ is generated to satisfy Assumption 3 with $\mathbf{z}_{\text{g}} \sim \mathcal{N}(0,1)$ and $\mathbf{z}_{\text{s}}$ being generated from a Bernoulli distribution with success probability p. The non-zero entries of $\mathbf{z}_{\text{s}}$ are set to be Gaussian i.i.d. with a variance that is 10 times that of $\mathbf{Hx}$. Different values of the parameter p, which controls the sparsity of the noise $\mathbf{z}$, are used in the experiments. The values of λ required in (38) are chosen to be $\{1, 2, \ldots, 10\} \times 10^{-2}$. Performance is evaluated using the normalized MSE (NMSE) (i.e., MSE normalized by $\|\mathbf{x}\|_2^2$).

In Fig 1(a), we compare the performance of the proposed approach, with λ_{optimal} obtained using the method in Section 3, to that of the regularized τ-estimator [16] and the regularized S-estimator when their λ_{optimal} is obtained using exhaustive search. The noise outlier rate p is set to vary from 0 to 40%. The loss functions of the proposed approach and the τ-estimator are chosen to be the optimal weight functions in [17] with $c_1 = 1.214$, $b_1 = 0.5$, and $c_2 = 3.270$, while for the S-estimator we use the Turkey's biweight loss function in [14] with $d = 1.547$. The results are obtained over 10^3 Monte-Carlo trials. From Fig 1(a), we can see that the proposed approach outperforms other robust methods over all the p range.

In Fig 1(b), we compare the performance of the proposed approach with the regularized M-estimator and three RLS algorithms: quasi-optimal, GCV, and L-curve [19]. These results are presented separately to provide better visualization given that these methods provide high NMSE. We use the loss function in [18] for the regularized M-estimator. From Fig 1(b), we observe that the proposed method outperforms all the other methods by offering the lowest NMSE while the RLS algorithms have the worst performance.

Fig 1(c) compares the runtime of our approach with the best two benchmarks methods. It is clear that the proposed estimator with its regularization parameter selection method provides significant complexity reduction. It should be noted that when the search algorithm is applied to obtain λ_{optimal} for the regularized τ-estimator and S-estimator, the search is confined to a prespecified feasible range $[\lambda_{\min}, \lambda_{\max}]$. Since this information may not be available in reality, the actual complexity of the search algorithm is worse than what Fig 1(c) shows.

5. CONCLUSIONS

A new robust regularized estimator based on the τ-estimator is proposed. Tools from RMT are used to obtain the optimal regularizer that minimizes the MSE of the estimator. Simulations show that the proposed method provides the lowest MSE among robust and non-robust benchmark methods.

6. REFERENCES

[1] Ulrich Hammes, Eric Wolsztynski, and Abdelhak M Zoubir, “Robust tracking and geolocation for wireless networks in nlos environments,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 5, pp. 889–901, 2009.

[2] Charles V Stewart, “Robust parameter estimation in computer vision,” *SIAM review*, vol. 41, no. 3, pp. 513–537, 1999.

[3] Tomaso Poggio, Vincent Torre, and Christof Koch, “Computational vision and regularization theory,” *Nature*, vol. 317, pp. 314–319, 1985.

[4] Abdelhak M Zoubir, Visa Koivunen, Yacine Chakhchoukh, and Michael Muma, “Robust estimation in signal processing: A tutorial-style treatment of fundamental concepts,” *IEEE Signal Processing Magazine*, vol. 29, no. 4, pp. 61–80, 2012.

[5] Peter J Huber, *Robust statistics*, Springer, 2011.

[6] Peter Rousseeuw and Victor Yohai, “Robust regression by means of s-estimators,” in *Robust and nonlinear time series analysis*, pp. 256–272. Springer, 1984.

[7] Victor J Yohai and Ruben H Zamar, “High breakdown-point estimates of regression by means of the minimization of an efficient scale,” *Journal of the American statistical association*, vol. 83, pp. 406–413, 1988.

[8] Mohamed Suliman, *Regularization Techniques for Linear Least-Squares Problems*, Ph.D. thesis, 2016.

[9] 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.

[10] 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 *24th European Signal Processing Conference (EUSIPCO)*, 2016, pp. 403–407.

[11] Mohamed Suliman, Tarig Ballal, and Tareq Y Al-Naffouri, “Perturbation-based regularization for signal estimation in linear discrete ill-posed problems,” *arXiv preprint arXiv:1611.09742*, 2016.

[12] 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.

[13] 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.

[14] Kukatharmini Tharmaratnam, Gerda Claeskens, Christophe Croux, and Matias Salibián-Barrera, “S-estimation for penalized regression splines,” *Journal of Computational and Graphical Statistics*, 2012.

[15] Ricardo A Maronna, “Robust ridge regression for high-dimensional data,” *Technometrics*, vol. 53, no. 1, pp. 44–53, 2011.

[16] Marta Martinez-Camara, Michael Muma, Abdelhak M Zoubir, and Martin Vetterli, “A new robust and efficient estimator for ill-conditioned linear inverse problems with outliers,” in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2015, pp. 3422–3426.

[17] Matias Salibian-Barrera, Gert Willems, and Ruben Zamar, “The fast-τ estimator for regression,” *Journal of Computational and Graphical Statistics*, 2012.

[18] Thomas CM Lee and Hee-Seok Oh, “Robust penalized regression spline fitting with application to additive mixed modeling,” *Computational Statistics*, vol. 22, no. 1, pp. 159–171, 2007.

[19] 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.

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

[21] 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.

[22] W. Hachem, O. Khorunzhiy, P. Loubaton, J. Najim, and L. Pastur, “A new approach for mutual information analysis of large dimensional multi-antenna channels,” *IEEE Transactions on Information Theory*, vol. 54, no. 9, pp. 3987–4004, Sept 2008.

[23] RARD Maronna, R Douglas Martin, and Victor Yohai, *Robust statistics*, John Wiley & Sons, Chichester. ISBN, 2006.