Reduced-rank Least Squares Parameter Estimation in the Presence of Byzantine Sensors

Nagananda K G, Rick. S. Blum
Department of ECE,
Lehigh University
Bethlehem, PA 18015, USA.
Email: {kgn209, rblum}@lehigh.edu

Alec Koppel
Information Sciences Division,
U.S. Army Research Laboratory,
Adelphi, MD 20783, USA.
Email: alec.e.koppel.civ@mail.mil

***Abstract*—In this paper, we study the impact of the presence of byzantine sensors on the reduced-rank linear least squares (LS) estimator. A sensor network with N sensors makes observations of the physical phenomenon and transmits them to a fusion center which computes the LS estimate of the parameter of interest. It is well-known that rank reduction exploits the bias-variance tradeoff in the full-rank estimator by putting higher priority on highly informative content of the data. The low-rank LS estimator is constructed using this highly informative content, while the remaining data can be discarded without affecting the overall performance of the estimator. We consider the scenario where a fraction $0 \leq \alpha \leq 1$ of the N sensors are subject to data falsification attack from byzantine sensors, wherein an intruder injects a higher noise power (compared to the unattacked sensors) to the measurements of the attacked sensors.**

Our main contribution is an analytical characterization of the impact of data falsification attack of the above type on the performance of reduced-rank LS estimator. In particular, we show how optimally prioritizing the highly informative content of the data gets affected in the presence of attacks. A surprising result is that, under sensor attacks, when the elements of the data matrix are all positive the error performance of the low-rank estimator experiences a phenomenon wherein the estimate of the mean-squared error comprises negative components. A complex nonlinear programming-based recipe is known to exist that resolves this undesirable effect; however, the phenomenon is oftentimes considered very objectionable in the statistical literature. On the other hand, to our advantage this effect can serve to detect cyber attacks on sensor systems. Numerical results are presented to complement the theoretical findings of the paper.

I. Introduction

There is great interest in sensor signal processing with emphasis on energy efficient statistical procedures for estimation, detection and classification. In order to meet the objective of energy efficiency, the sensors deployed to make observations of the physical phenomenon of interest are typically low-cost with limited computational and communication capabilities, making them vulnerable to cyber attacks. Thus, when sensors are deployed in an adversarial environment where an attacker has malicious intent of disabling or altering the decision-making mechanism, they are subject to various kinds of attacks such as spoofing attacks, man-in-the-middle attacks or a combination of the two. While spoofing attacks falsify the signals obtained from the physical phenomenon before the sensing process, man-in-the-middle attacks are aimed at altering the sensor measurements before reaching the decision-maker (for instance, a fusion center); see [1] for a survey on data falsification attacks on IoT sensor networks. The attacked sensors are commonly referred to as byzantine sensors and the data from these sensors are called byzantine data [2]. In this paper, we address the problem of parameter estimation using the reduced-rank linear least squares (LS) method when the sensor system comprises byzantine data.

Impact of byzantine sensors on standard inference procedures like estimation and detection are widely reported in the literature. For example, effects of data falsification attacks on signal detection and some mitigation techniques can be found in [3]–[6]. Sensor attacks on estimation systems have appeared in [7]–[14], where different types of attacks and after-attack estimation performance in large scale sensor networks are analyzed. Countermeasures, which are fully independent of the network topology, to sensor attacks on distributed parameter estimation systems are suggested in [15], [16]. Performance analysis and detection of byzantine data in cyber physical systems (such as the Smart Grid) were studied in [17], [18]. However, to the best of our knowledge, byzantine attacks on reduced-rank processing has not been investigated, and is the subject topic of this paper.

Reduced-rank approach is a prominent technique in statistical signal processing for dimensionality reduction and data compression, wherein highly informative content of the data is utilized, while the remaining data can be discarded without affecting the performance of statistical inference [19]–[21]. A study of cyber attacks on reduced-rank systems is very important to understand its implications on practical data compression algorithms.

Identification of the highly informative content in data can be achieved by exploiting the fundamental distortion-variance tradeoff in the statistical procedure employed [22]. For example, it has been shown that for the linear LS parameter estimation problem, rank reduction introduces bias into the otherwise unbiased LS estimator while decreasing its variance [22, Chapter 9]. The sum of squared-bias plus variance is smaller than the variance of the unbiased estimator, thereby improving the overall mean-squared error (mse) performance. The bias-variance tradeoff is exploited by arranging the dot products of each singular vector of the data matrix and the

observation vector in the decreasing order of their norms, thus prioritizing highly informative data. The reduced-rank LS estimator is then constructed using only the most informative (*i.e.*, large vector-norm) eigenvectors and discarding the rest. Sensor attacks could be aimed at disrupting this ordering mechanism which will directly impact the selection of highly informative content of the data.

To model the attack on the sensor system, we consider the scenario where a fraction $0 \leq \alpha \leq 1$ of the N sensors are under cyber attack. The sensors make observations of the physical phenomenon and transmit them to a fusion center, which computes the LS estimate of the unknown parameter. In our attack model, the attacker intentionally injects a higher noise power to the measurements of the byzantine sensors, while those of the unattacked sensors are corrupted by the usual measurement noise. In practice, this is the case when an intruder deliberately injects a higher noise power to selected sensors to disrupt the operation of the network.

When the network of sensors is under attack, the reduced-rank LS estimator gets affected in the following manner. Firstly, the optimal value of the number of eigenvectors of the data matrix required to construct the reduced-rank estimator is different from the unattacked case. Though this consequence is expected, it is not clear to what extent an attack alters the number of eigenvectors; thus a proper characterization is necessary and is provided in this paper. Secondly, the aforementioned ordering (or "selection") principle gets affected. In other words, the measure of "informativeness" of the eigenvectors of the data matrix gets altered. Therefore, the highly informative eigenvectors may not get selected in the list of eigenvectors used to construct the reduced-rank LS estimator, resulting in suboptimal utilization of the bias-variance tradeoff. This effect is also not unexpected. Again, we provide a formal description of this phenomenon.

Our most important observation is that, the error performance of the low-rank estimator get affected in a very unexpected manner. As shown in the sequel, even in the absence of attacks, it is not possible to directly evaluate the mse of the reduced-rank estimator. Instead, an estimate of mse is obtained which takes into account the bias introduced into the estimator. When the network is attacked in the fashion described above, and when the elements of the data matrix are all positive, the mse estimate comprises negative components, which is quite misleading to assess the performance of the estimator. In the statistical literature, there exists a procedure proposed by Thomson, Jr. (see [23]) based on nonlinear programming theory to resolve this undesirable effect of negative components in the estimate of a function. However, in the context of cyber attacks, we can take advantage of this effect to detect the presence of attacks on the sensor system. We demonstrate our theoretical findings via computer experiments.

The rest of the paper is organized as follows. In Section II, we derive the structure of the reduced-rank LS estimator in the presence of byzantine sensors. Computer-based experimental results and ensuing discussion are in Section III. Concluding remarks and future directions are provided in Section IV.

II. Characterization of the Reduced-rank LS estimator with byzantine sensors

For sake of clarity and completeness, we first present an overview of the reduced-rank LS estimator in Section II-A. In Section II-B, the attack model is introduced and the reduced-rank LS estimator with byzantine sensors is developed.

A. Overview of reduced-rank LS estimator

Consider a sensor network with N sensors deployed for parameter estimation in a region of interest. Let $\boldsymbol{H}$ denote a $N \times p$ data matrix and $\boldsymbol{z}$ denote the $N \times 1$ vector of observations made by the N sensors in the network. The sensor measurements are transmitted to the fusion center which aims to recover a $p \times 1$ parameter $\boldsymbol{\theta}$ from $\boldsymbol{z}$ by solving the set of linear equations $\boldsymbol{z} = \boldsymbol{H\theta} + \boldsymbol{w}$, where $\boldsymbol{w}$ denotes the $N \times 1$ noise vector assumed to follow a Gaussian distribution with mean 0 and variance σ^2. For simplicity, assume the channel between the sensors and the fusion center to be noiseless. Letting $\boldsymbol{x} \triangleq \boldsymbol{H\theta}$, the LS estimate of $\boldsymbol{x}$ is given by $\hat{\boldsymbol{x}} = \boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{z}$ assuming the matrix inverse exists, where $(\cdot)^{\mathrm{T}}$ denotes the transpose operator. The mse of the estimator $\hat{\boldsymbol{x}}$ is $p\sigma^2$.

The reduced-rank estimator of $\boldsymbol{x}$ denoted by $\hat{\boldsymbol{x}}_r$ is constructed using only $r < p$ most informative singular vectors of $\boldsymbol{H}$, while discarding the rest [22, Chapter 9]. Let $\boldsymbol{u}_i$ for $i = 1, \ldots, p$ denote the p singular vectors of $\boldsymbol{H}$ obtained by eigenvalue decomposition of $\boldsymbol{H}$. If these singular vectors are ordered such that $\|\boldsymbol{u}_{(1)}^{\mathrm{T}}\boldsymbol{z}\|^2 \geq \cdots \geq \|\boldsymbol{u}_{(r)}^{\mathrm{T}}\boldsymbol{z}\|^2 \geq \cdots \geq \|\boldsymbol{u}_{(p)}^{\mathrm{T}}\boldsymbol{z}\|^2$, where $\boldsymbol{u}_{(i)}$ denotes the i^{th} most informative eigenvector and $\|(\cdot)\|$ denotes the norm of a vector, then the reduced-rank LS estimator can be constructed using only r most informative singular vectors $[\boldsymbol{u}_{(1)}, \ldots, \boldsymbol{u}_{(r)}]$, while $[\boldsymbol{u}_{(r+1)}, \ldots, \boldsymbol{u}_{(p)}]$ can be discarded. The value r is obtained as a solution to an optimization problem exploiting the bias-variance tradeoff. The reduced-rank LS estimator has a variance lesser than $p\sigma^2$ but with non-zero bias that is tolerable.

B. Presence of byzantine sensors in the network

When the sensor system is under attack, the optimal value of r is altered, *i.e.*, the minimum number of eigenvectors required to construct the reduced-rank estimator is now different from the unattacked case. Furthermore, since the ordering of eigenvectors gets affected, highly informative eigenvectors may not get selected in the list $[\boldsymbol{u}_{(1)}, \ldots, \boldsymbol{u}_{(r)}]$ used to construct the reduced-rank LS estimator. These will in turn impact the complexity of the prior model and increases the sensitivity of the estimator to measurement errors.

In our attack model, we assume a fraction $0 \leq \alpha \leq 1$ of the N sensors to be byzantine sensors. The fusion center is assumed to have prior knowledge of the fraction of sensors under attack, however, the exact identity of the attacked sensors is unknown. In order to degrade system performance, the byzantine sensors could employ varying strategies. In the attack model considered in this paper, a byzantine sensor deliberately injects higher noise power into its measurements, and thereby its observations are much more corrupted than

the measurements made by the honest sensors. If the attacker has full control of α, the estimation scheme can potentially be completely disrupted. We write, for $i = 1, \ldots, N$,

$$\begin{cases} y_i = \boldsymbol{h}_i\boldsymbol{\theta} + w_{1,i}, \ i \text{ honest sensor} \\ x_i = \boldsymbol{h}_i\boldsymbol{\theta} + w_{2,i}, \ i \text{ byzantine sensor,} \end{cases} \tag{1}$$

where $w_{1,i} \sim \mathcal{N}(0, \sigma_1^2)$ and $w_{2,i} \sim \mathcal{N}(0, \sigma_2^2)$ are both independent and identically distributed and are independent of each other. We assume that $\sigma_2^2 > \sigma_1^2$ to signify that byzantine sensors' observations are (deliberately) much noisier than the measurements of the honest sensors. The distributions of x_i and y_i are given by

$$\begin{cases} y_i \sim \mathcal{N}\left(\boldsymbol{h}_i\boldsymbol{\theta}, \sigma_1^2\right) \\ x_i \sim \mathcal{N}\left(\boldsymbol{h}_i\boldsymbol{\theta}, \sigma_2^2\right), \end{cases} \tag{2}$$

The observations collected at the fusion center are denoted by the $N \times 1$ vector $\boldsymbol{z} = [z_1, \ldots, z_N]$, where

$$z_i = \begin{cases} x_i, \ \text{with probability } \alpha \\ y_i, \ \text{with probability } 1-\alpha. \end{cases} \tag{3}$$

Therefore, we can write

$$\begin{aligned} \boldsymbol{z} &= \alpha(\boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{w}_2) + (1-\alpha)(\boldsymbol{H}\boldsymbol{\theta} + \boldsymbol{w}_1) \\ &= \alpha\boldsymbol{H}\boldsymbol{\theta} + \alpha\boldsymbol{w}_2 + (1-\alpha)\boldsymbol{H}\boldsymbol{\theta} + (1-\alpha)\boldsymbol{w}_1 \\ &= \boldsymbol{H}\boldsymbol{\theta} + (1-\alpha)\boldsymbol{w}_1 + \alpha\boldsymbol{w}_2. \qquad (4) \\ \Rightarrow \boldsymbol{z} &\sim \mathcal{N}\left(\boldsymbol{H}\boldsymbol{\theta}, [(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2]\boldsymbol{I}_N\right). \qquad (5) \end{aligned}$$

For notational simplicity, let us write

$$\boldsymbol{z} = \boldsymbol{q} + \boldsymbol{n}, \tag{6}$$

where $\boldsymbol{q} = \boldsymbol{H}\boldsymbol{\theta}$ and $\boldsymbol{n} = (1-\alpha)\boldsymbol{w}_1 + \alpha\boldsymbol{w}_2$.

The least squares estimates of the parameter $\boldsymbol{\theta}$, the signal $\boldsymbol{z}$ and noise $\boldsymbol{n}$ in the presence of attacks are denoted by $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{x}}$ and $\hat{\boldsymbol{n}}$, respectively. These are given by

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= (\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{z}, \qquad (7) \\ \hat{\boldsymbol{q}} &= \boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{z}, \qquad (8) \\ \hat{\boldsymbol{n}} &= [\boldsymbol{I}_N - \boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}]\boldsymbol{y}, \qquad (9) \end{aligned}$$

and are unique if the inverse of $\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H}$ exists. The estimators $\hat{\boldsymbol{\theta}}$, $\hat{\boldsymbol{q}}$ and $\hat{\boldsymbol{n}}$ are linear transformations on the multivariate normal random vector $\boldsymbol{y}$, governed by the following distributions:

$$\begin{aligned} \hat{\boldsymbol{\theta}} &\sim \mathcal{N}\left(\boldsymbol{\theta}, [(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2](\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\right), \\ \hat{\boldsymbol{q}} &\sim \mathcal{N}\left(\boldsymbol{H}\boldsymbol{\theta}, [(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2]\boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\right), \\ \hat{\boldsymbol{n}} &\sim \mathcal{N}\left(\boldsymbol{0}_N, [(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2][\boldsymbol{I}_N - \boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}]\right). \end{aligned}$$

From (10) it can be seen that $\hat{\boldsymbol{q}}$ is an unbiased estimator of $\boldsymbol{q}$. The squared error $\hat{\boldsymbol{q}}^{\mathrm{T}}\hat{\boldsymbol{q}} \sim [(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2]\chi_p^2$, where χ_p^2 denotes the central Chi squared distribution with p degrees of freedom. The central Chi squared distribution is the result of the sum of squares of p independent zero mean Gaussian random variables. The mse of the estimator $\hat{\boldsymbol{q}}$ is $\mathbb{E}[(\hat{\boldsymbol{q}}-\boldsymbol{q})^2] = p[(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2]$.

In reduced-rank processing, we will seek to achieve a smaller mse than $p\sigma^2$ albeit at the price of nonzero mean. This corresponds to a bias-variance tradeoff. Towards this end, we first express (7) - (9) in terms of the SVD of the matrix $\boldsymbol{H}$ which is given by $\boldsymbol{H} = \boldsymbol{U}_H\boldsymbol{\Gamma}_H\boldsymbol{V}_H^{\mathrm{T}}$, where $\boldsymbol{U}_H = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p] \in \mathbb{R}^{N\times p}$ and $\boldsymbol{V}_H = [\boldsymbol{v}_1, \ldots, \boldsymbol{v}_p] \in \mathbb{R}^{p\times p}$ are orthogonal matrices, and $\boldsymbol{\Gamma}_H = \mathrm{diag}\,[\gamma_1, \ldots, \gamma_p] \in \mathbb{R}^{p\times p}$ is a diagonal matrix comprising the singular values $\gamma_1 \geq \cdots \geq \gamma_p$. Thus, the estimators are governed by the following distributions:

$$\begin{aligned} \hat{\boldsymbol{\theta}} &\sim \mathcal{N}\left(\boldsymbol{\theta}, [(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2]\boldsymbol{V}_H\boldsymbol{\Gamma}_H^{-2}\boldsymbol{V}_H^{\mathrm{T}}\right), \qquad (10) \\ \hat{\boldsymbol{q}} &\sim \mathcal{N}\left(\boldsymbol{H}\boldsymbol{\theta}, [(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2]\boldsymbol{U}_H\boldsymbol{U}_H^{\mathrm{T}}\right), \qquad (11) \\ \hat{\boldsymbol{n}} &\sim \mathcal{N}\left(\boldsymbol{0}_N, [(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2]\left[\boldsymbol{I}_N - \boldsymbol{U}_H\boldsymbol{U}_H^{\mathrm{T}}\right]\right). \qquad (12) \end{aligned}$$

The full-rank estimator $\hat{\boldsymbol{q}}$ will be replaced by a low-rank estimator:

$$\hat{\boldsymbol{q}}_r \triangleq \boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{z}, \tag{13}$$

where the $N \times r$ matrix $\boldsymbol{U}_r = [\boldsymbol{u}_{(1)}, \ldots, \boldsymbol{u}_{(r)}]$ is obtained by discarding $(p-r)$ orthogonal vectors that comprise $\boldsymbol{U}_H$ and $\boldsymbol{u}_{(j)}$ denotes the j^{th} "ordered" orthogonal vector which is not necessarily the j^{th} vector. The notion of ordering will become clearer as we proceed.

The estimation error $(\boldsymbol{q} - \hat{\boldsymbol{q}}_r) \sim \mathcal{N}(\boldsymbol{q} - \boldsymbol{q}_r, [(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2]\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}})$, where $(\boldsymbol{q} - \boldsymbol{q}_r) = \boldsymbol{b}$ denotes the bias of the LS estimator. The mse of the reduced-rank estimator $\hat{\boldsymbol{q}}_r$ is denoted by $\mathrm{mse}(r)$, and is given by

$$\begin{aligned} \mathrm{mse}(r) &= \mathbb{E}\left\{[\boldsymbol{q} - \hat{\boldsymbol{q}}_r]^{\mathrm{T}}[\boldsymbol{q} - \hat{\boldsymbol{q}}_r]\right\} \\ &= \mathbb{E}\left\{[\boldsymbol{q}^{\mathrm{T}}\boldsymbol{q} - \boldsymbol{q}^{\mathrm{T}}\hat{\boldsymbol{q}}_r - \hat{\boldsymbol{q}}_r^{\mathrm{T}}\boldsymbol{q} + \hat{\boldsymbol{q}}_r^{\mathrm{T}}\hat{\boldsymbol{q}}_r]\right\} \\ &= \mathbb{E}\left\{\boldsymbol{q}^{\mathrm{T}}\boldsymbol{q}\right\} - \mathbb{E}\left\{\boldsymbol{q}^{\mathrm{T}}\hat{\boldsymbol{q}}_r\right\} - \mathbb{E}\left\{\hat{\boldsymbol{q}}_r^{\mathrm{T}}\boldsymbol{q}\right\} + \mathbb{E}\left\{\hat{\boldsymbol{q}}_r^{\mathrm{T}}\hat{\boldsymbol{q}}_r\right\} \\ &= \boldsymbol{q}^{\mathrm{T}}\boldsymbol{q} - \boldsymbol{q}^{\mathrm{T}}\boldsymbol{q}_r - \boldsymbol{q}_r^{\mathrm{T}}\boldsymbol{q} + \mathbb{E}\left\{\hat{\boldsymbol{q}}_r^{\mathrm{T}}\hat{\boldsymbol{q}}_r\right\} \\ &= \boldsymbol{q}^{\mathrm{T}}\boldsymbol{q} - \boldsymbol{q}^{\mathrm{T}}\boldsymbol{q}_r - \boldsymbol{q}_r^{\mathrm{T}}\boldsymbol{q} + \mathbb{E}\left\{\boldsymbol{z}^{\mathrm{T}}\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{z}\right\} \\ &= \boldsymbol{q}^{\mathrm{T}}\boldsymbol{q} - \boldsymbol{q}^{\mathrm{T}}\boldsymbol{q}_r - \boldsymbol{q}_r^{\mathrm{T}}\boldsymbol{q} \\ &\quad + \mathbb{E}\left\{(\boldsymbol{q}+\boldsymbol{n})^{\mathrm{T}}\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}(\boldsymbol{q}+\boldsymbol{n})\right\} \\ &= \boldsymbol{q}^{\mathrm{T}}\boldsymbol{U}_H\boldsymbol{U}_H^{\mathrm{T}}\boldsymbol{q} - \boldsymbol{q}^{\mathrm{T}}\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{q} - \boldsymbol{q}^{\mathrm{T}}\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{q} \\ &\quad + \mathbb{E}\left\{\boldsymbol{q}^{\mathrm{T}}\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{q} + \boldsymbol{q}^{\mathrm{T}}\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{n}\right. \\ &\quad \left. + \boldsymbol{n}^{\mathrm{T}}\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{q} + \boldsymbol{n}^{\mathrm{T}}\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{n}\right\} \\ &= \boldsymbol{q}^{\mathrm{T}}\boldsymbol{U}_H\boldsymbol{U}_H^{\mathrm{T}}\boldsymbol{q} - \boldsymbol{q}^{\mathrm{T}}\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{q} - \boldsymbol{q}^{\mathrm{T}}\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{q} \\ &\quad + \boldsymbol{q}^{\mathrm{T}}\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{q} + \mathbb{E}\left\{\boldsymbol{n}^{\mathrm{T}}\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{n}\right\} \\ &= \boldsymbol{q}^{\mathrm{T}}\boldsymbol{U}_H\boldsymbol{U}_H^{\mathrm{T}}\boldsymbol{q} - \boldsymbol{q}^{\mathrm{T}}\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{q} + \mathbb{E}\left\{\boldsymbol{n}^{\mathrm{T}}\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{n}\right\} \\ &= \boldsymbol{q}^{\mathrm{T}}\boldsymbol{U}_H\boldsymbol{U}_H^{\mathrm{T}}\boldsymbol{q} - \boldsymbol{q}^{\mathrm{T}}\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{q} + \mathrm{tr}\left[\mathbb{E}\left\{\boldsymbol{n}\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{n}^{\mathrm{T}}\right\}\right] \\ &\overset{(i)}{=} \sum_{j=r+1}^{p} \|\boldsymbol{u}_{(j)}^{\mathrm{T}}\boldsymbol{q}\|^2 + r[(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2], \qquad (14) \end{aligned}$$

where (i) follows from the fact that the trace of an idempotent matrix is equal to the rank of the matrix. Note that, we still need to find an optimal r that produces a lesser mse for the estimator $\hat{\boldsymbol{q}}_r$ compared to that of $\hat{\boldsymbol{q}}$. The following theorem asserts that such an r indeed exists and can be uncovered by adding an appropriate bias to the estimator $\hat{\boldsymbol{q}}_r$.

Theorem 1: Consider the linear model in (6) and the reduced-rank LS estimator given by (13). There exists an

$r = r^*$ given by

$$r^* = \arg\min_r \left[\left[\sum_{j=r+1}^{p} \|\boldsymbol{u}_{(j)}^{\mathrm{T}} \boldsymbol{z}\|^2 + (2r-p)[(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2] \right] \right] \quad (15)$$

that minimizes the mse between $\boldsymbol{q}$ and $\hat{\boldsymbol{q}}_r$. Since r is integer-valued, r^* is obtained by rounding off to the nearest integer, and the subspace can be found using a combinatorial search.

Theorem 1 provides a full characterization of the impact of byzantine sensors on the rank reduction process.

Proof: The central idea in proving Theorem 1 hinges on the fact that rank reduction is beneficial when the mse of the estimator $\hat{\boldsymbol{q}}_r$ is smaller than that of the full-rank estimator $\hat{\boldsymbol{q}}$. Secondly, deriving an estimator for $\mathrm{mse}(r)$ (due to the presence of the unknown quantity $\boldsymbol{q}$) brings into light the ordering principle that is perhaps the most important outcome of rank reduction, and it is this ordering principle which the intruder tries to disrupt.

The estimator $\hat{\boldsymbol{q}}_r \sim \mathcal{N}(\boldsymbol{q}_r, [(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2]\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}})$, where $\boldsymbol{q}_r = \boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\boldsymbol{q}$ is the projection of $\boldsymbol{q}$ onto the span of $\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}$. The rank reduction procedure will reduce the variance of the estimator of $\boldsymbol{q}$ whenever

$$p[(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2] > \sum_{j=r+1}^{p} \|\boldsymbol{u}_{(j)}^{\mathrm{T}}\boldsymbol{q}\|^2 + r[(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2], \quad (16)$$

which suggests that the optimum choice of the rank r is

$$r^* = \arg\min_r \mathrm{mse}(r) = \arg\min_r \left[\sum_{j=r+1}^{p} \|\boldsymbol{u}_{(j)}^{\mathrm{T}}\boldsymbol{q}\|^2 + r[(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2] \right]. \quad (17)$$

However, since the signal vector $\boldsymbol{q} \triangleq \boldsymbol{H}\boldsymbol{\theta}$ is unknown, we replace $\mathrm{mse}(r)$ with its estimate to solve the setup in (17). Towards this end, we first estimate the bias $\boldsymbol{b}$ using the following statistic:

$$\begin{aligned} \hat{\boldsymbol{b}} &= \left(\boldsymbol{U}_H\boldsymbol{U}_H^{\mathrm{T}} - \boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}\right)\boldsymbol{z} \\ &\sim \mathcal{N}\left(\boldsymbol{b}, [(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2](\boldsymbol{U}_H\boldsymbol{U}_H^{\mathrm{T}} - \boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}})\right). \end{aligned} \quad (18)$$

The mse of the estimator $\hat{\boldsymbol{b}}$ is given by

$$\begin{aligned} \mathbb{E}\{[\hat{\boldsymbol{b}}-\boldsymbol{b}]^{\mathrm{T}}[\hat{\boldsymbol{b}}-\boldsymbol{b}]\} &= \mathbb{E}\{\hat{\boldsymbol{b}}^{\mathrm{T}}\hat{\boldsymbol{b}} - \hat{\boldsymbol{b}}^{\mathrm{T}}\boldsymbol{b} - \boldsymbol{b}^{\mathrm{T}}\hat{\boldsymbol{b}} + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}\} \\ &\overset{(ii)}{=} \mathbb{E}\{\hat{\boldsymbol{b}}^{\mathrm{T}}\hat{\boldsymbol{b}}\} - \boldsymbol{b}^{\mathrm{T}}\boldsymbol{b} \\ &\overset{(iii)}{=} \mathrm{tr}\left[[(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2](\boldsymbol{U}_H\boldsymbol{U}_H^{\mathrm{T}} - \boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}})\right] \\ &= [(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2](p-r), \end{aligned} \quad (19)$$

where (ii) and (iii) follow from (18). From (19), we see that

$$\mathbb{E}\{\hat{\boldsymbol{b}}^{\mathrm{T}}\hat{\boldsymbol{b}}\} = [(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2](p-r) + \boldsymbol{b}^{\mathrm{T}}\boldsymbol{b} \quad (20)$$

which implies that the estimator $\hat{\boldsymbol{b}}^{\mathrm{T}}\hat{\boldsymbol{b}}$ must be corrected by $-[(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2](p-r)$ to be an unbiased estimator of $\boldsymbol{b}^{\mathrm{T}}\boldsymbol{b}$. This leads to the following estimator for $\mathrm{mse}(r)$:

$$\begin{aligned} \hat{\mathrm{mse}}(r) &= \hat{\boldsymbol{b}}^{\mathrm{T}}\hat{\boldsymbol{b}} - [(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2](p-r) \\ &\quad + r[(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2] \\ &= \hat{\boldsymbol{b}}^{\mathrm{T}}\hat{\boldsymbol{b}} + (2r-p)[(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2]. \end{aligned} \quad (21)$$

The optimum choice of r is, therefore, given by

$$\begin{aligned} r^* &= \arg\min_r \hat{\mathrm{mse}}(r) \\ &\overset{(iv)}{=} \arg\min_r \left[\left[\sum_{j=r+1}^{p} \|\boldsymbol{u}_{(j)}^{\mathrm{T}}\boldsymbol{z}\|^2 + (2r-p)[(1-\alpha)^2\sigma_1^2 + \alpha^2\sigma_2^2] \right] \right], \end{aligned} \quad (22)$$

where (iv) follows from (18). This proves Theorem 1. ■

Solving (22) numerically provides the optimal r^*. Once r^* is obtained, the reduced-rank estimator is obtained by discarding $(p-r^*)$ columns in the matrix $\boldsymbol{U}_H = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_p] \in \mathbb{R}^{N\times p}$, and is given by $\hat{\boldsymbol{q}}_{r^*} \triangleq \boldsymbol{U}_{r^*}\boldsymbol{U}_{r^*}^{\mathrm{T}}\boldsymbol{z}$, whose sum of squared-bias plus variance is smaller than the variance of the unbiased estimator $\hat{\boldsymbol{q}} = \boldsymbol{H}(\boldsymbol{H}^{\mathrm{T}}\boldsymbol{H})^{-1}\boldsymbol{H}^{\mathrm{T}}\boldsymbol{z}$ given by (8). From (22), we see that the eigenvectors of $\boldsymbol{U}_H$ should be ordered such that $\|\boldsymbol{u}_{(1)}^{\mathrm{T}}\boldsymbol{z}\|^2 \geq \cdots \geq \|\boldsymbol{u}_{(r)}^{\mathrm{T}}\boldsymbol{z}\|^2 \geq \cdots \geq \|\boldsymbol{u}_{(p)}^{\mathrm{T}}\boldsymbol{z}\|^2$, and the dominant r eigenvectors should be used to construct the rank$-r$ projector $\boldsymbol{U}_r\boldsymbol{U}_r^{\mathrm{T}}$.

It is clear that the attack parameters α and σ_2 control prioritizing the data and the number of eigenvectors used to construct the reduced-rank estimator. Equation (22) fully characterizes the impact of byzantine sensors on rank reduction which is the main objective of this paper. To complement the theoretical findings, in the next section we will show using computer experiments the impact of α on the performance of the reduced-rank estimator $\hat{\boldsymbol{q}}_{r^*}$. Of particular interest is the variation of r^* with α, *i.e.*, the numerical evaluation of (22) and the variation of $\mathrm{mse}(r)$ with α given by (21). We will also show the variations of r^* and $\mathrm{mse}(r)$ with the strength of the attack governed by σ_2. When $\alpha = 0$, we get the unattacked reduced-rank LS estimator presented in [22, Chapter 9]. In practice, the fusion center might not have perfect prior knowledge of the fraction α of byzantines in the network. It is, however, possible to learn α over a fixed duration of time if the attacked sensors transmit continuously to the fusion center; see, for example, [6, Appendix A] for a treatment of this topic.

III. Numerical results and related discussion

For computer experiments, we consider a linear system with data matrix $\boldsymbol{H}$ comprising $N = 500$ rows and $p = 200$ columns. The noise variance at the unattacked sensors is $\sigma_1^2 = 1$. In the first experiment, we vary the fraction α of attacked sensors between 0 and 1, and calculate the optimal reduced-rank r^* numerically by solving the optimization setup

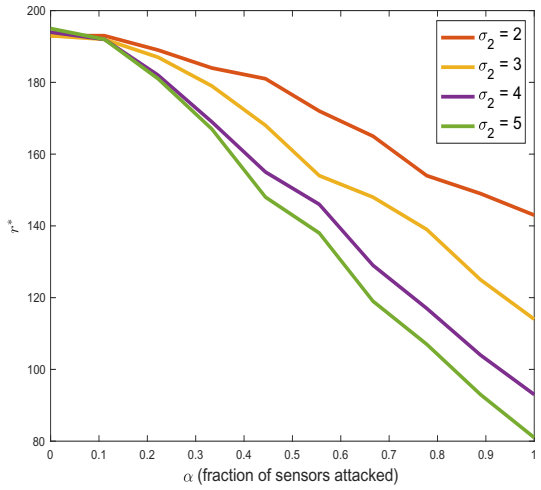

Fig. 1: Variation of r^* with α for different values of σ_2.

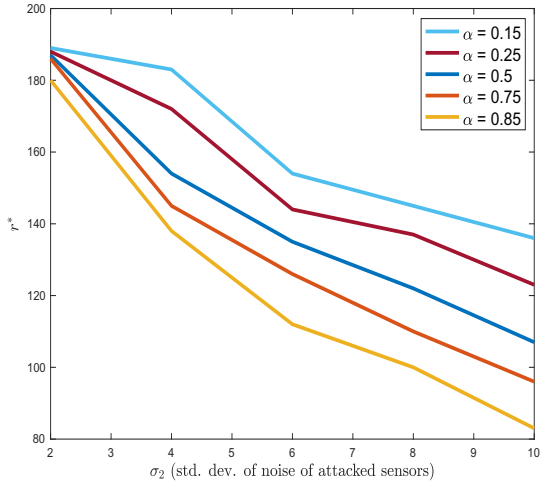

Fig. 2: Variation of r^* with σ_2 for different values of α.

in (22). The variation of r^* with α is shown in Fig. 1 for different values of the standard deviation of the noise at the attacked sensors. It can be seen how the byzantine attack affects rank selection - as more sensors are being attacked, fewer eigenvectors, which are highly informative, get selected for constructing the low-rank estimator.

In the next experiment, we demonstrate the behavior of the variation of r^* with the standard deviation σ_2 of the noise at the attacked sensors. The results are shown in Fig. 2, for different fractions of sensors under attack. The variation of r^* with σ_2 is similar to that noticed in Fig. 1. The variation of the error performance of the low-rank estimator for different fractions of sensors attacked can be seen in Fig. 3. As we showed in the previous section, it is not possible to evaluate $\text{mse}(r)$ due to its dependence on the signal vector $\boldsymbol{q}$, which is unknown. We, therefore, replace $\text{mse}(r)$ with its estimate $\hat{\text{mse}}(r)$, which depends on the estimate of the bias $\hat{\boldsymbol{b}}$ given by

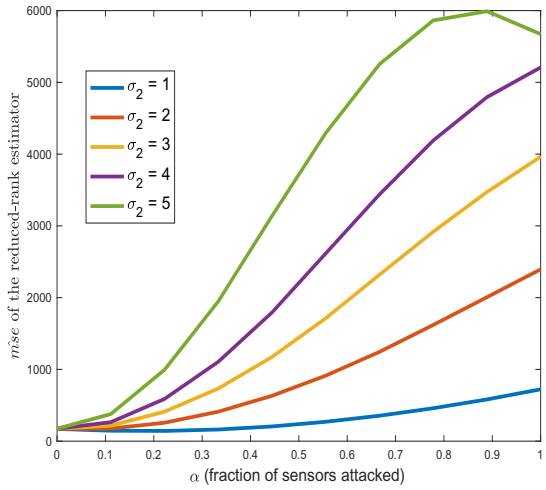

Fig. 3: The problem of negative estimates of $\hat{\text{mse}}$ components when the elements of $\boldsymbol{H}$ are all positive.

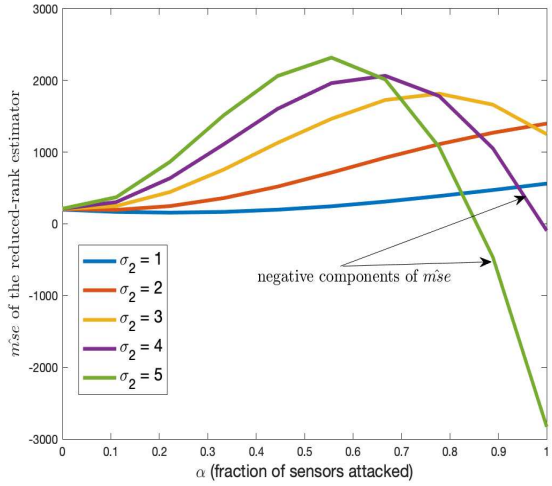

Fig. 4: The problem of negative estimates of $\hat{\text{mse}}$ components when the elements of $\boldsymbol{H}$ are all positive.

(18). Expectedly, the error metric $\hat{\text{mse}}(r)$ grows significantly for increasing values of α. This is the case when there are no restrictions on the sign of the coefficients of the linear system given by (6), *i.e.*, the elements of $\boldsymbol{H}$.

The surprising result is the effect of the byzantine attacks on the estimate $\hat{\text{mse}}(r)$ of the mse of the low-rank estimator when the elements of $\boldsymbol{H}$ are all positive. For this case, when we plot $\hat{\text{mse}}(r)$ as a function of the fraction α of attacked sensors, for different values of the attack parameter σ_2, we observe negative components of $\hat{\text{mse}}(r)$ as shown in Fig. 4. In the statistical literature, this unusual behavior has been reported for the estimates of variance, see [23] and references therein for example. The negative components of estimates is considered to be highly objectionable, and there have been efforts to alleviate this effect. In [23], results from

nonlinear programming theory have been employed to resolve the problem of negative components in the estimates of the variance parameter. From the fusion center's standpoint, this effect can be viewed as an advantage in detecting cyber attacks on the sensor system.

IV. Remarks

In this paper, we consider the problem of cyber attacks on the reduced-rank linear least squares estimator. A fraction α of the total N sensors in the network is under a cyber attack. The intruder intentionally injects a higher noise power to the attacked sensors. We demonstrated that, as a consequence of this type of attack, optimal rank selection is affected. Further, the most informative eigenvectors of the data matrix may not get chosen to construct the low-rank estimator. These two effects have a direct impact on the error performance of the low-rank estimator. As shown by the experimental results, the major consequence of sensor attacks appears in the form of negative components in the estimates of the mse of the low-rank estimator which misleads the assessment of the performance of the estimator.

Experimental results clearly indicate that if the intruder has full control of the fraction α of byzantine sensors, he can potentially disrupt the reduced-rank processing to the extent of making it useless. On the other hand, if the fusion center is able to detect the identity of byzantine sensors, the resulting impact could be minimized. For example, if the fusion center can accurately estimate the noise variance of the received samples, then it can possibly recognize the attacked sensors because of their higher noise power and adapt the rank selection procedure in a manner to alleviate the effect of attacks. However, this comes at the price of higher number of sensor measurements required to estimate the noise variance σ_2^2. Another approach would be to consider a sequence of K measurements at each sensor, which supplies NK measurements to the fusion center. As reported in [10], for $K \to \infty$ and $N \to \infty$, it is possible to ascertain the number of sensors under attack. Extending the results of [10] to reduced-rank processing opens several avenues of research.

For future research, we will also consider the scenario where the sensor network faces different types of attacks. This could, for example, correspond to the situation where several intruders attack different subsets of sensors; the intruders inject different noise powers to each subset of sensors. The effect of such multiple attacks on rank selection, ordering of eigenvectors and error performance of the low-rank estimator will provide useful insights.

Acknowledgement

This material is based upon work partially supported by the U. S. Army Research Laboratory and the U. S. Army Research Office under grant number W911NF-17-1-0331 and by the National Science Foundation under grants ECCS-1744129 and CNS1702555.

References

[1] J. Zhang, R. S. Blum, and H. V. Poor, "Approaches to secure inference in the internet of things: Performance bounds, algorithms, and effective attacks on IoT sensor networks," *IEEE Signal Process. Mag.*, vol. 35, no. 5, pp. 50–63, Sep. 2018.

[2] A. Vempaty, L. Tong, and P. K. Varshney, "Distributed inference with byzantine data: State-of-the-art review on data falsification attacks," *IEEE Signal Process. Mag.*, vol. 30, no. 5, pp. 65–75, Sep. 2013.

[3] S. Marano, V. Matta, and L. Tong, "Distributed detection in the presence of byzantine attacks," *IEEE Trans. Signal Process.*, vol. 57, no. 1, pp. 16–29, Jan. 2009.

[4] B. Kailkhura, Y. S. Han, S. Brahma, and P. K. Varshney, "Asymptotic analysis of distributed Bayesian detection with byzantine data," *IEEE Signal Process. Lett.*, vol. 22, no. 5, pp. 608–612, May 2015.

[5] B. Kailkhura, S. Brahma, B. Dulek, Y. S. Han, and P. K. Varshney, "Distributed detection in tree networks: Byzantines and mitigation techniques," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 7, pp. 1499–1512, Jul. 2015.

[6] B. Kailkhura, Y. S. Han, S. Brahma, and P. K. Varshney, "Distributed Bayesian detection in the presence of byzantine data," *IEEE Trans. Signal Process.*, vol. 63, no. 19, pp. 5250–5263, Oct. 2015.

[7] A. Vempaty, O. Ozdemir, K. Agrawal, H. Chen, and P. K. Varshney, "Localization in wireless sensor networks: Byzantines and mitigation techniques," *IEEE Trans. Signal Process.*, vol. 61, no. 6, pp. 1495–1508, Mar. 2013.

[8] J. Zhang and R. S. Blum, "Distributed estimation in the presence of attacks for large scale sensor networks," in *Conf. Inf. Sci. Syst. (CISS)*, Mar. 2014, pp. 1–6.

[9] B. Alnajjab and R. S. Blum, "After-attack performance of parameter estimation systems," in *Conf. Inf. Sci. Syst. (CISS)*, Mar. 2014, pp. 1–6.

[10] J. Zhang, R. S. Blum, X. Lu, and D. Conus, "Asymptotically optimum distributed estimation in the presence of attacks," *IEEE Trans. Signal Process.*, vol. 63, no. 5, pp. 1086–1101, Mar. 2015.

[11] J. Zhang, R. S. Blum, L. M. Kaplan, and X. Lu, "Functional forms of optimum spoofing attacks for vector parameter estimation in quantized sensor networks," *IEEE Trans. Signal Process.*, vol. 65, no. 3, pp. 705–720, Feb. 2017.

[12] J. Zhang, X. Wang, R. S. Blum, and L. M. Kaplan, "Attack detection in sensor network target localization systems with quantized data," *IEEE Trans. Signal Process.*, vol. 66, no. 8, pp. 2070–2085, Apr. 2018.

[13] M. Al-Salman and R. Niu, "Source location with quantized sensor data corrupted by false information," in *Int. Conf. Inf. Fusion (FUSION)*, Jul. 2018, pp. 391–397.

[14] Y. Liu and C. Li, "Secure distributed estimation over wireless sensor networks under attacks," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 54, no. 4, pp. 1815–1831, Aug. 2018.

[15] Y. Chen, S. Kar, and J. M. F. Moura, "Resilient distributed estimation: Sensor attacks," *IEEE Trans. Auto. Control*, vol. 64, no. 9, pp. 3772–3779, Sep. 2019.

[16] ——, "Resilient distributed parameter estimation with heterogeneous data," *IEEE Trans. Signal Process.*, vol. 67, no. 19, pp. 4918–4933, Oct. 2019.

[17] P. Pradhan, K. Nagananda, P. Venkitasubramaniam, S. Kishore, and R. S. Blum, "GPS spoofing attack characterization and detection in smart grids," in *IEEE Conf. Commun. Net. Security*, Oct. 2016, pp. 391–395.

[18] J. Yao, P. Venkitasubramaniam, S. Kishore, L. V. Snyder, and R. S. Blum, "Network topology risk assessment of stealthy cyber attacks on advanced metering infrastructure networks," in *Conf. Inf. Sci. Syst. (CISS)*, Mar. 2017, pp. 1–6.

[19] L. Scharf, "The SVD and reduced rank signal processing," *Signal Process.*, vol. 25, no. 2, pp. 113–133, Nov. 1991.

[20] P. Stoica and M. Viberg, "Maximum likelihood parameter and rank estimation in reduced-rank multivariate linear regressions," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 3069–3078, Dec. 1996.

[21] J. S. Goldstein and I. S. Reed, "Reduced-rank adaptive filtering," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 492–496, Feb. 1997.

[22] L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison-Wesley, 1991.

[23] W. A. Thompson, Jr., "The problem of negative estimates of variance components," *The Ann. Math. Stat.*, vol. 33, no. 1, pp. 273–289, Mar. 1962.