# On Estimation under Noisy Order Statistics

Alex Dytso†, Martina Cardone*, Mishfad S. Veedu*, H. Vincent Poor†
† Princeton University, Princeton, NJ 08544, USA, Email: {adytso, poor}@princeton.edu
* University of Minnesota, Minneapolis, MN 55404, USA, Email: {cardo089, veedu002}@umn.edu

***Abstract*—This paper presents an estimation framework to assess the performance of the sorting function over data that is perturbed. In particular, the performance is measured in terms of the Minimum Mean Square Error (MMSE) between the values of the sorting function computed on the data without perturbation and the estimate that uses the sorting function applied to the perturbed data. It is first shown that, under certain conditions satisfied by the practically relevant Gaussian noise perturbation, the optimal estimator can be expressed as a linear combination of estimators on the unsorted data. Then, a suboptimal estimator is proposed, and its performance is evaluated and compared to the optimal estimator. Finally, a lower bound on the desired MMSE is derived when data is i.i.d. and has a Gaussian distribution. This is accomplished by solving a new problem that consists of estimating the norm of an unsorted vector from a noisy observation of it.**

## I. Introduction

Sorting is a widely used function and a benchmark for several modern recommender and distributed computing systems (e.g., Hadoop MapReduce). Today, sorting is often performed over massive amounts of data, which might be sensitive (e.g., clinical/genomic health) and hence it is required to remain confidential/private. For instance, in a recommender system, a user may not wish to fully reveal her interests or previous purchases. In order to ensure data confidentiality, one solution would consist of perturbing the data with some noise. This gives rise to a natural question: How does data perturbation affect the performance of the sorting function?

In this paper, we focus on assessing the performance of the sorting function over data that is perturbed (e.g., because of data privacy purposes). Our goal is to quantify the performance loss of the sorting function versus different levels of noise perturbation. Towards this end, we pose the problem within an estimation framework. In particular, as a metric, we adopt the Minimum Mean Square Error (MMSE) between the values of the sorting function computed on the original data (i.e., with no perturbation) and the estimate that uses the sorting function applied to the noisy version of the data.

We first analyze the optimal estimator, i.e., the conditional expectation of observing the original data as sorted, given the observation of the noisy sorted data. We show that, under certain conditions satisfied by the practically relevant Gaussian noise perturbation, the optimal estimator can be expressed as a linear combination of estimators on the unsorted data. Then, by leveraging the structure of the optimal estimator, we propose a suboptimal estimator and prove that, with Gaussian noise perturbation, it is asymptotically optimal in the low noise regime, and its MMSE performance is always to within a constant gap of the optimal MMSE. Finally, we derive a lower bound on the desired MMSE term, when the original data is i.i.d. and has a Gaussian distribution. This consists of another MMSE term obtained by estimating the norm of an unsorted vector from a noisy observation of it. As such, this new MMSE term, beyond providing a closed-form lower bound on the desired MMSE, might also be of independent interest since it provides the solution to another estimation problem.

**Related Work.** Recently, sorting with noisy data has gained significant traction. For instance, in [1] the authors focused on estimating the *permutation structure* of the original sorted vector given a noisy observation of it. In [2], the authors considered the estimation of the unknown *deterministic* but randomly permuted vector, which is also perturbed by additive noise. Different from these works, we are interested in estimating the values of the components of the original sorted vector, by performing joint estimation and sorting.

The estimation of parameters of a family of distributions from an ordered vector has received considerable attention. For example, for a location-scale family of distributions with density function $\frac{1}{\sigma} f\left(\frac{x-\mu}{\sigma}\right)$, the best linear unbiased estimator (BLUE) of $\sigma$ and $\mu$ has been found in [3]. In particular, the BLUE was shown to be a function of only the covariance matrix and the mean vector of the order statistics. However, note that the computation of the covariance matrix and the mean vector of the order statistics is often a formidable task. Explicit expressions for moments of order statistics are in fact known only for some specific distributions and have been tabulated in [4]. A rich body of literature also exists on universal bounds on moments of order statistics [5].

Having observed a partial sample of the first $r$ order statistics from a sample of size $n$, the best linear unbiased predictor (BLUP) of the remaining $n - r$ terms has been characterized in [6]. Specifically, the predictor in [6] was shown to depend only on the covariance matrix and the mean of the order statistics. Interesting connections between the BLUE and BLUP have been derived in [7].

Due to its application to life-testing experiments, inference of censored (i.e., incomplete data) order statistics has received significant attention. Interestingly, for various types of censoring protocols, closed-form expressions for maximum likelihood estimators of parameters are available. A comprehensive survey of several censoring scenarios can be found in [8].

Order statistics also appear in the study of outliers since these are expected to be a few extreme order statistics. Several effective tests are formed from extreme order statistics that

The work of A. Dytso and H. V. Poor was supported in part by the U.S. National Science Foundation under Grants CCF-0939370 and CCF-1513915. The work of M. Cardone was supported in part by the U.S. National Science Foundation under Grant CCF-1849757.

seek to compute the deviation of the candidate outliers from the rest of the data [9]. Another application of order statistics is on the goodness-of-fit tests. The most classical example of such a test is the Shapiro and Wilk's test for normality [10].

**Paper Organization.** Section II introduces the notation, presents some definitions and describes the system model. Section III studies the MMSE of estimating a sorted vector from a noisy sorted observation of it, and proposes a suboptimal estimator. Section IV considers the practically relevant Gaussian noise case and derives a lower bound on the desired MMSE. Finally, Section V concludes the paper.

## II. Notation, Definitions and System Model

Boldface upper case letters $\mathbf{X}$ denote vector random variables; the boldface lower case letter $\mathbf{x}$ indicates a specific realization of $\mathbf{X}$; $\vec{\mathbf{X}}$ is $\mathbf{X}$ ordered in ascending order; $X_i$ and $X_{(i)}$ specify the $i$-th entry of $\mathbf{X}$ and $\vec{\mathbf{X}}$, respectively; $\|\mathbf{x}\|$ denotes the $\ell_2$-norm of the vector $\mathbf{x}$; the set $\vec{\mathbb{R}}_n = \{\mathbf{x} : \tilde{\mathbf{x}} \in \mathbb{R}_{\mathbf{n}}\}$; $\delta(x)$ is the Dirac delta function; $1_{\{X_i\}}(x_i)$ is the indicator function of $X_i$ taking value $x_i$; $[n]$is the set of integers $\{1, \ldots, n\}$; $\mathbf{P}_n$ is the matrix of dimension $n! \times n$ containing in each row a permutation of the elements of $[n]$, and $\mathbf{P}_n(i, :)$ is the $i$-th row of $\mathbf{P}_n$; $\vec{\mathbf{y}}_{\mathbf{v}}$ is a vector whose components are ordered according to the $n$-length vector $\mathbf{v}$, i.e., $y_{v_1} \leq y_{v_2} \leq \ldots \leq y_{v_n}$; $\mathbf{I}_n$ is the identity matrix of dimension $n$.

We next provide three definitions, which will be used in the proof of our main results.

**Definition 1.** *A sequence of random variables $U_1, U_2, \ldots, U_n$ is said to be exchangeable (or interchangeable) if the distribution of the random vector $(U_1, U_2, \ldots, U_n)$ is the same as that of $(U_{\pi_1}, U_{\pi_2}, \ldots, U_{\pi_n})$ for any permutation $(\pi_1, \pi_2, \ldots, \pi_n)$ of the elements of $[n]$. More formally,*

$$(U_1, U_2, \ldots, U_n) \stackrel{d}{=} (U_{\pi_1}, U_{\pi_2}, \ldots, U_{\pi_n}),$$

*where $\stackrel{d}{=}$ denotes equality in distribution.*

Note that any convex combination or mixture distribution of independent and identically distributed sequences of random variables is exchangeable.

**Definition 2.** *The confluent hypergeometric function is given by*

$$F_{1,1}(a, b; x) = \sum_{k=0}^{\infty} \frac{\Gamma(a+k)\Gamma(b)}{\Gamma(a)\Gamma(b+k)} \frac{x^k}{k!}, \ \min\{a, b, x\} > 0, \quad (1)$$

*where $\Gamma(\cdot)$ is the gamma function.*

**Definition 3.** *The modified Bessel function of the first kind of order $v \geq 0$ is defined as*

$$\mathsf{I}_v(x) = \frac{x^v}{2^v \sqrt{\pi}\Gamma\left(v + \frac{1}{2}\right)} \int_0^{\pi} \mathrm{e}^{x \cos\theta} (\sin\theta)^{2v} \mathrm{d}\theta, \ x \in \mathbb{R}. \quad (2)$$

We consider the framework shown in Fig. 1, where an $n$-dimensional random vector $\mathbf{X}$ is generated according to a probability density function (PDF) $f_{\mathbf{X}}(\cdot)$ and then passed through a noisy channel with a transition probability equal to $f_{\mathbf{Y}|\mathbf{X}}(\cdot|\cdot)$. In particular, this is assumed to be a *parallel* channel, i.e.,

$$f_{\mathbf{Y}|\mathbf{X}}(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{n} f_{Y_i|X_i}(y_i|x_i).$$

The output of the channel – denoted as $\mathbf{Y}$ – is finally sorted in ascending order, i.e.,$\vec{\mathbf{Y}}$ is the sorted version of $\mathbf{Y}$.

In this work, as thoroughly explained in Section III, we are interested in characterizing the MMSE of estimating $\vec{\mathbf{X}}$ – which denotes the sorted version of $\mathbf{X}$ (i.e., the ground truth) – when $\vec{\mathbf{Y}}$ is observed.

## III. On the Optimal MMSE Estimator

The objective of this section is to study the MMSE of estimating $\vec{\mathbf{X}}$ from an observation $\vec{\mathbf{Y}}$. The MMSE is given by

$$\mathrm{mmse}(\vec{\mathbf{X}}|\vec{\mathbf{Y}}) = \mathbb{E}\left[\left\|\vec{\mathbf{X}} - \mathbb{E}\left[\vec{\mathbf{X}}|\vec{\mathbf{Y}}\right]\right\|^2\right]. \quad (3)$$

It is well known that the conditional expectation $\mathbb{E}\left[\vec{\mathbf{X}}|\vec{\mathbf{Y}}\right]$ is the optimal estimator under the square error criterion. The next theorem provides a characterization of $\mathbb{E}\left[\vec{\mathbf{X}}|\vec{\mathbf{Y}}\right]$ in terms of the distribution of $(\mathbf{X}, \mathbf{Y})$ under certain symmetry conditions.

**Theorem 1.** *Let $(\mathbf{X}, \mathbf{Y})$ be continuous random vectors. Assume that $\mathbf{X}$ is exchangeable (see Definition 1) and that*

$$\begin{aligned} &\frac{1}{n!} \sum_{\ell=1}^{n!} \sum_{j=1}^{n!} f_{\mathbf{Y}|\mathbf{X}}(\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)} | \vec{\mathbf{x}}_{\mathbf{P}_n(\ell,:)}) \\ &= \sum_{j=1}^{n!} f_{\mathbf{Y}|\mathbf{X}}(\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)} | \vec{\mathbf{x}}), \ \forall (\vec{\mathbf{x}}, \vec{\mathbf{y}}). \end{aligned} \quad (4)$$

*Then, for any $k \in [n]$*

$$\begin{aligned} &\mathbb{E}[X_{(k)}|\vec{\mathbf{Y}} = \vec{\mathbf{y}}] \\ &= n! \sum_{j=1}^{n!} \frac{f_{\mathbf{Y}}(\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)})}{f_{\vec{\mathbf{Y}}}(\vec{\mathbf{y}})} \mathbb{E}\left[X_k \cdot 1_{\vec{\mathbb{R}}_n}(\mathbf{X}) | \mathbf{Y} = \vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}\right]. \end{aligned} \quad (5)$$

*In addition, if $\mathbf{Y}$ is exchangeable (see Definition 1), then for any $k \in [n]$*

$$\mathbb{E}[X_{(k)}|\vec{\mathbf{Y}} = \vec{\mathbf{y}}] = \sum_{j=1}^{n!} \mathbb{E}\left[X_k \cdot 1_{\vec{\mathbb{R}}_n}(\mathbf{X}) | \mathbf{Y} = \vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}\right]. \quad (6)$$

*Proof:* We have

$$\begin{aligned} &\mathbb{E}\left[X_{(k)}|\vec{\mathbf{Y}} = \vec{\mathbf{y}}\right] \\ &= \int_{\vec{\mathbb{R}}_n} t_k f_{\vec{\mathbf{X}}|\vec{\mathbf{Y}}}(\mathbf{t}|\vec{\mathbf{y}}) \mathrm{d}\mathbf{t} \\ &\stackrel{\text{(a)}}{=} \int_{\vec{\mathbb{R}}_n} t_k \frac{n!}{f_{\vec{\mathbf{Y}}}(\vec{\mathbf{y}})} f_{\mathbf{X}}(\mathbf{t}) \sum_{j=1}^{n!} f_{\mathbf{Y}|\mathbf{X}}(\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}|\mathbf{t}) \mathrm{d}\mathbf{t} \\ &= \frac{n!}{f_{\vec{\mathbf{Y}}}(\vec{\mathbf{y}})} \sum_{j=1}^{n!} \int_{\vec{\mathbb{R}}_n} t_k f_{\mathbf{Y}|\mathbf{X}}(\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}|\mathbf{t}) f_{\mathbf{X}}(\mathbf{t}) \mathrm{d}\mathbf{t} \end{aligned}$$

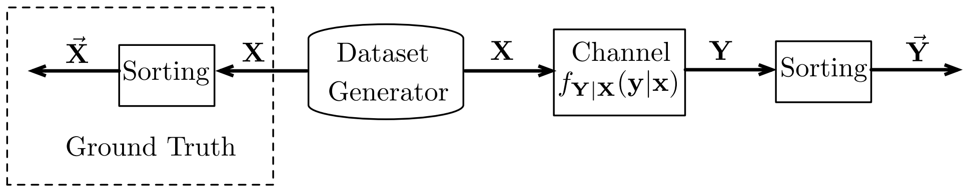


Fig. 1: Graphical representation of the proposed framework.

$$
\begin{aligned}
&= \frac{n!}{f_{\vec{\mathbf{Y}}}(\vec{\mathbf{y}})} \sum_{j=1}^{n!} \int_{\mathbb{R}_n} t_k \cdot 1_{\vec{\mathbb{R}}_n}(\mathbf{t}) f_{\mathbf{Y}|\mathbf{X}}(\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}|\mathbf{t}) f_{\mathbf{X}}(\mathbf{t}) \mathrm{d}\mathbf{t} \\
&= \frac{n!}{f_{\vec{\mathbf{Y}}}(\vec{\mathbf{y}})} \sum_{j=1}^{n!} \frac{f_{\mathbf{Y}}(\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)})}{f_{\mathbf{Y}}(\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)})} \int_{\mathbb{R}_n} t_k 1_{\vec{\mathbb{R}}_n}(\mathbf{t}) f_{\mathbf{X}}(\mathbf{t}) \\
&\qquad \cdot f_{\mathbf{Y}|\mathbf{X}}(\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}|\mathbf{t}) \mathrm{d}\mathbf{t} \\
&\overset{(b)}{=} \frac{n!}{f_{\vec{\mathbf{Y}}}(\vec{\mathbf{y}})} \sum_{j=1}^{n!} f_{\mathbf{Y}}(\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}) \int_{\mathbb{R}_n} t_k 1_{\vec{\mathbb{R}}_n}(\mathbf{t}) f_{\mathbf{X}|\mathbf{Y}}(\mathbf{t}|\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}) \mathrm{d}\mathbf{t} \\
&\overset{(c)}{=} \frac{n!}{f_{\vec{\mathbf{Y}}}(\vec{\mathbf{y}})} \sum_{j=1}^{n!} f_{\mathbf{Y}}(\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}) \mathbb{E}\left[X_k 1_{\vec{\mathbb{R}}_n}(\mathbf{X}) | \mathbf{Y} = \vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}\right],
\end{aligned}
$$

where the equalities follow from: (a) the identity

$$
f_{\vec{\mathbf{X}}|\vec{\mathbf{Y}}}(\vec{\mathbf{x}}|\vec{\mathbf{y}}) = \frac{n!}{f_{\vec{\mathbf{Y}}}(\vec{\mathbf{y}})} f_{\mathbf{X}}(\vec{\mathbf{x}}) \sum_{j=1}^{n!} f_{\mathbf{Y}|\mathbf{X}}(\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}|\vec{\mathbf{x}}),
$$

whose proof can be found in [11, Lemma 6]; (b) using Bayes' rule; and (c) using the definition of the conditional expectation. This concludes the proof of (5). To show that (6) holds observe that, if $\mathbf{Y}$ is exchangeable, then

$$
\frac{f_{\mathbf{Y}}(\vec{\mathbf{y}}_{\mathbf{P}(j,:)})}{f_{\vec{\mathbf{Y}}}(\vec{\mathbf{y}})} = \frac{f_{\mathbf{Y}}(\vec{\mathbf{y}})}{f_{\vec{\mathbf{Y}}}(\vec{\mathbf{y}})} = \frac{1}{n!}, \tag{7}
$$

where the last equality follows by [11, Lemma 5]. This concludes the proof of Theorem 1. ■

**Remark 1.** *Examples of noisy transformations that satisfy the condition in* (4) *include the practically relevant case of a Gaussian noisy channel transition probability* $f_{\mathbf{Y}|\mathbf{X}}(\cdot|\cdot)$.

Theorem 1 allows to express the optimal estimator of elements of $\vec{\mathbf{X}}$ from an observation $\vec{\mathbf{Y}}$ as a linear combination of estimators of the original unsorted $\mathbf{X}$ from the unsorted channel observation $\mathbf{Y}$. This has the potential of significantly simplifying the computation of the conditional expectation as there is no need to find the joint distribution of $(\vec{\mathbf{X}}, \vec{\mathbf{Y}})$.

Observe that the conditional expectation in (5) and (6) can be written as follows

$$
\begin{aligned}
&\mathbb{E}\left[X_k \cdot 1_{\vec{\mathbb{R}}_n}(\mathbf{X}) | \mathbf{Y} = \vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}\right] \\
&= \mathbb{E}\Big[X_k | \mathbf{X} \in \vec{\mathbb{R}}_n, \mathbf{Y} = \vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}\Big] \mathbb{P}\Big[\mathbf{X} \in \vec{\mathbb{R}}_n | \mathbf{Y} = \vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}\Big].
\end{aligned}
$$

We now use this identity to propose and study the following suboptimal estimator of $X_{(k)}$

$$
\begin{aligned}
\hat{f}_k(\vec{\mathbf{y}}) = \sum_{j=1}^{n!} &\mathbb{E}\Big[X_k | \mathbf{Y} = \vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}\Big] \\
&\cdot \mathbb{P}\left[\mathbf{X} \in \vec{\mathbb{R}}_n | \mathbf{Y} = \vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}\right],
\end{aligned} \tag{8a}
$$

and the following suboptimal estimator of $\vec{\mathbf{X}}$

$$
\hat{\mathbf{f}}(\vec{\mathbf{Y}}) = [\hat{f}_1(\vec{\mathbf{Y}}), \ldots, \hat{f}_n(\vec{\mathbf{Y}})]. \tag{8b}
$$

Observe that the only difference between the optimal estimator in (6) and the suboptimal estimator in (8a) is that in the latter the conditioning on $\mathbf{X} \in \vec{\mathbb{R}}_n$ has been dropped. In other words, we are implicitly using the approximation $\mathbb{E}\left[X_k | \mathbf{X} \in \vec{\mathbb{R}}_n, \mathbf{Y}\right] \approx \mathbb{E}[X_k | \mathbf{Y}]$.

The next theorem, whose proof can be found in [11, Appendix B], compares the performances of the optimal estimator in (6) and the proposed suboptimal estimator in (8a).

**Theorem 2.** *Suppose that* $\mathbf{X}$ *and* $\mathbf{Y}$ *are exchangeable and the assumption in* (4) *holds. Let*

$$
\Delta = \left| \mathbb{E}\left[ \left\| \vec{\mathbf{X}} - \hat{\mathbf{f}}(\vec{\mathbf{Y}}) \right\|^2 \right] - \mathrm{mmse}(\vec{\mathbf{X}}|\vec{\mathbf{Y}}) \right|. \tag{9}
$$

*Then,*

$$
\Delta \leq \Delta_{up} = \sum_{j=1}^{n!} \mathbb{E}\left[ \|\mathbf{X}\|^2 g\left(\mathbf{Y}_{\mathbf{P}_n(j,:)}\right) \mid \mathbf{Y} \in \vec{\mathbb{R}}_n \right], \tag{10}
$$

*where*

$$
g(\mathbf{y}) = \mathbb{P}\Big[\mathbf{X} \in \vec{\mathbb{R}}_n | \mathbf{Y} = \mathbf{y}\Big] \cdot \Big(1 - \mathbb{P}\Big[\mathbf{X} \in \vec{\mathbb{R}}_n | \mathbf{Y} = \mathbf{y}\Big]\Big). \tag{11}
$$

In Section IV, under the assumption of an additive Gaussian noise, the estimator in (8a) will be shown to be optimal in a small noise regime. Moreover, in the very noisy regime the mean squared error (MSE) of the estimator in (8a) will be shown to be within a constant gap of the MMSE (i.e., the error attained by the optimal estimator in (6)).

## IV. Analysis with Gaussian Statistics

In this section, we consider the practically relevant case of Gaussian noise, i.e., we assume that $\mathbf{Y}|\mathbf{X} = \mathbf{x} \sim \mathcal{N}(\mathbf{x}, \sigma^2 \mathbf{I}_n)$.

We start by noting that, under the additional assumption that the input $\mathbf{X} \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$, the proposed suboptimal estimator in (8a) takes the form of a weighted linear function given by

$$
\hat{f}_k(\vec{\mathbf{y}}) = \sum_{j=1}^{n!} a_j(\vec{\mathbf{y}}) \left[\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}\right]_k, \tag{12}
$$

where $\left[\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}\right]_k$ is $k$-th component of the vector $\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}$ and where

$$
a_j(\vec{\mathbf{y}}) = \frac{\sigma}{1+\sigma^2} \mathbb{P}\left[\mathbf{X} \in \vec{\mathbb{R}}_n | \mathbf{Y} = \vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}\right]. \tag{13}
$$

Fig. 2 compares the performance, in terms of the MSE, of the estimator in (12) (dashed lines) and the optimal estimator in (6) (solid lines), versus different values of the noise standard deviation $\sigma$. From Fig. 2 we observe that the suboptimal estimator performs closely to the optimal estimator for small values of $\sigma$, whereas for higher values of $\sigma$ Fig. 2 suggests that the MSE of the estimator in (12) is to within a constant gap of the MSE of the optimal estimator. We next formalize these observations in Theorem 3, and show that they hold also when the input $\mathbf{X}$ is not necessarily Gaussian.

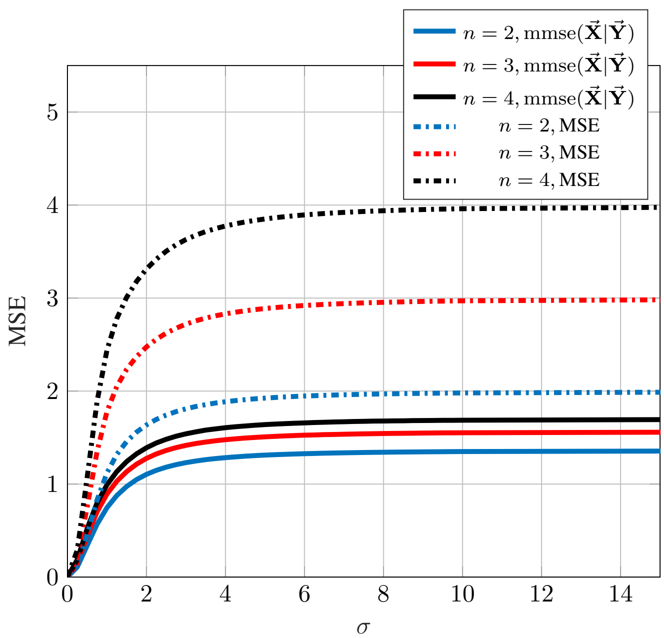


Fig. 2: Comparison of the $\mathrm{mmse}(\vec{\mathbf{X}}|\vec{\mathbf{Y}})$ and the MSE of the estimator in (12) versus $\sigma$.

In particular, Theorem 3 shows that, for any exchangeable input $\mathbf{X}$, the suboptimal estimator in (8) is asymptotically optimal in the low noise regime. Moreover, Theorem 3 provides an upper bound on the penalty in the high noise regime.

**Theorem 3.** *Let* $\mathbf{X}$ *be exchangeable and let* $\mathbf{Y}|\mathbf{X}=\mathbf{x}\sim\mathcal{N}(\mathbf{x},\sigma^2\mathbf{I}_n)$. *Assume that* $\mathbb{E}[X_k^2]<\infty$ *for all* $k\in[n]$. *Then, the approximation error in* (10) *satisfies the following*

$$\lim_{\sigma\to 0}\Delta_{up}=0, \tag{14}$$

$$\lim_{\sigma\to\infty}\Delta_{up}=\mathbb{E}\left[\|\mathbf{X}\|^2\right]\left(1-\frac{1}{n!}\right). \tag{15}$$

*Proof:* We start by noting that, by a simple application of the dominated convergence theorem, we have that

$$\lim_{\sigma\to 0}\mathbb{P}\left[\mathbf{X}\in\vec{\mathbb{R}}_n|\mathbf{Y}=\mathbf{y}\right]=1_{\vec{\mathbb{R}}_n}(\mathbf{y}), \tag{16}$$

$$\lim_{\sigma\to\infty}\mathbb{P}\left[\mathbf{X}\in\vec{\mathbb{R}}_n|\mathbf{Y}=\mathbf{y}\right]=\mathbb{P}\left[\mathbf{X}\in\vec{\mathbb{R}}_n\right]=\frac{1}{n!}. \tag{17}$$

Now observe that

$$\begin{aligned}
&\lim_{\sigma\to 0}\mathbb{E}\left[X_k^2\cdot g\left(\mathbf{Y}_{\mathbf{P}_n(j,:)}\right)\mid\mathbf{Y}\in\vec{\mathbb{R}}_n\right]\\
&=\frac{\lim_{\sigma\to 0}\mathbb{E}\left[X_k^2\cdot g\left(\mathbf{Y}_{\mathbf{P}_n(j,:)}\right)\cdot 1_{\vec{\mathbb{R}}_n}(\mathbf{Y})\right]}{n!}\\
&\overset{(a)}{=}\frac{\mathbb{E}\left[\lim_{\sigma\to 0}X_k^2\cdot g\left(\mathbf{Y}_{\mathbf{P}_n(j,:)}\right)\cdot 1_{\vec{\mathbb{R}}_n}(\mathbf{Y})\right]}{n!}\\
&\overset{(b)}{=}\frac{\mathbb{E}\left[X_k^2\cdot 1_{\vec{\mathbb{R}}_n}\left(\mathbf{X}_{\mathbf{P}_n(j,:)}\right)\left(1-1_{\vec{\mathbb{R}}_n}\left(\mathbf{X}_{\mathbf{P}_n(j,:)}\right)\right)\cdot 1_{\vec{\mathbb{R}}_n}(\mathbf{X})\right]}{n!}\\
&\overset{(c)}{=}0,
\end{aligned} \tag{18}$$

where the labeled equalities follow from: (a) using the dominated convergence theorem with the bound

$$\begin{aligned}
&X_k^2\cdot g\left(\mathbf{Y}_{\mathbf{P}_n(j,:)}\right)\\
&=X_k^2\cdot p_{\vec{\mathbf{X}}|\mathbf{Y}}\left(\mathbf{Y}_{\mathbf{P}_n(j,:)}\right)\left(1-p_{\vec{\mathbf{X}}|\mathbf{Y}}(\mathbf{Y}_{\mathbf{P}_n(j,:)})\right)\cdot 1_{\vec{\mathbb{R}}_n}(\mathbf{Y})\\
&\le X_k^2,
\end{aligned} \tag{19}$$

with

$$p_{\vec{\mathbf{X}}|\mathbf{Y}}(\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}=\mathbb{P}\left[\mathbf{X}\in\vec{\mathbb{R}}_n|\mathbf{Y}=\vec{\mathbf{y}}_{\mathbf{P}_n(j,:)}\right],$$

and the assumption that $\mathbb{E}[X_k^2]<\infty$; (b) using the limit in (16); and (c) using the fact that $1_{\vec{\mathbb{R}}_n}\left(\mathbf{X}_{\mathbf{P}_n(j,:)}\right)\left(1-1_{\vec{\mathbb{R}}_n}\left(\mathbf{X}_{\mathbf{P}_n(j,:)}\right)\right)=0$. Combining (18) with the definition of $\Delta_{\text{up}}$ in (10) we arrive at

$$\lim_{\sigma\to 0}\Delta=0.$$

We now focus on the case of $\sigma\to\infty$, and we obtain

$$\begin{aligned}
&\lim_{\sigma\to\infty}\mathbb{E}\left[X_k^2\cdot g\left(\mathbf{Y}_{\mathbf{P}_n(j,:)}\right)\mid\mathbf{Y}\in\vec{\mathbb{R}}_n\right]\\
&=\lim_{\sigma\to\infty}\mathbb{E}\left[X_k^2\cdot g\left(\vec{\mathbf{Y}}_{\mathbf{P}_n(j,:)}\right)\right]\\
&\overset{(a)}{=}\mathbb{E}\left[\lim_{\sigma\to\infty}X_k^2\cdot g\left(\vec{\mathbf{Y}}_{\mathbf{P}_n(j,:)}\right)\right]\\
&\overset{(b)}{=}\mathbb{E}\left[X_k^2\cdot\frac{1}{n!}\left(1-\frac{1}{n!}\right)\right]\\
&=\mathbb{E}\left[X_k^2\right]\cdot\frac{1}{n!}\left(1-\frac{1}{n!}\right),
\end{aligned} \tag{20}$$

where the labeled equalities follow from: (a) using the dominated convergence theorem with the bound in (19) and the assumption that $\mathbb{E}[X_k^2]<\infty$; and (b) using the limit in (17). Combining (18) with the definition of $\Delta_{\text{up}}$ in (10), we obtain

$$\lim_{\sigma\to\infty}\Delta_{\text{up}}=\sum_{k=1}^{n}\mathbb{E}\left[X_k^2\right]\cdot\left(1-\frac{1}{n!}\right).$$

This concludes the proof of Theorem 3. ■

As highlighted above, Theorem 3 shows that the estimator proposed in (8) is asymptotically optimal in the low noise regime, and its MMSE performance is always to within a constant gap of the optimal MMSE. The plot of the upper bound on the penalty in (10) for the Gaussian input $\mathbf{X}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_n)$ is shown in Fig. 3, versus different values of $\sigma$.

Obtaining a closed-form expression for $\mathrm{mmse}(\vec{\mathbf{X}}|\vec{\mathbf{Y}})$ is in general not possible, and hence computable lower bounds become necessary. The next theorem presents a lower bound on the desired MMSE term, when the input $\mathbf{X}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_n)$.

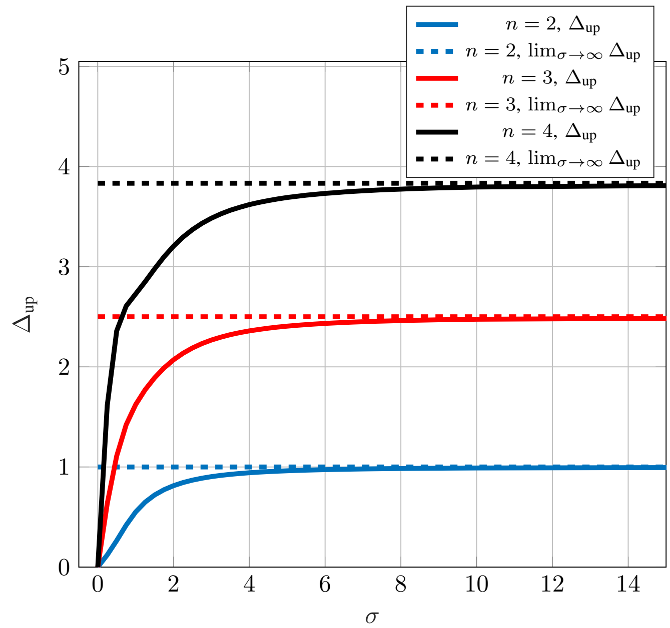


Fig. 3: Plot of $\Delta_{\text{up}}$ evaluated with $\mathbf{X}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_n)$ versus the noise level $\sigma$.

**Theorem 4.** *Let $\mathbf{X}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_n)$. Then,*

$$\begin{aligned}\mathrm{mmse}\left(\vec{\mathbf{X}}\mid\vec{\mathbf{Y}}\right)&\geq\mathrm{mmse}\left(\|\mathbf{X}\|\mid\mathbf{Y}\right)\\&=\mathrm{mmse}\left(\|\mathbf{X}\|\mid\|\mathbf{Y}\|\right).\end{aligned}\tag{21}$$

*Moreover,*

$$\begin{aligned}&\mathrm{mmse}\left(\|\mathbf{X}\|\mid\|\mathbf{Y}\|\right)\\&=n-\frac{2\sigma^{2+n}}{1+\sigma^2}\sum_{k=0}^{\infty}\sum_{m=0}^{\infty}\frac{a_k a_m}{(\sigma^2+2)^{k+m+\frac{n}{2}}}\frac{\Gamma(\frac{n}{2}+k+m)}{\Gamma(\frac{n}{2})},\end{aligned}$$

*where*

$$a_k=\frac{\Gamma\left(k+\frac{n+1}{2}\right)}{k!\Gamma\left(k+\frac{n}{2}\right)},$$

*and $\Gamma(\cdot)$ being the gamma function. In addition,*

$$\begin{aligned}&\mathbb{E}\left[\|\mathbf{X}\|\mid\|\mathbf{Y}\|=\|\mathbf{y}\|\right]=\mathbb{E}\left[\|\mathbf{X}\|\mid\mathbf{Y}=\mathbf{y}\right]\\&=\frac{\Gamma(\frac{n+1}{2})\sqrt{2\sigma^2}\mathrm{e}^{-\frac{\|\mathbf{y}\|^2}{2\sigma^2(1+\sigma^2)}}}{\Gamma(\frac{n}{2})\sqrt{1+\sigma^2}}F_{1,1}\left(\frac{n+1}{2},\frac{n}{2};\frac{\|\mathbf{y}\|^2}{2\sigma^2(1+\sigma^2)}\right),\end{aligned}$$

*where $F_{1,1}(\cdot,\cdot;\cdot)$ is the confluent hypergeometric function defined in Definition 2.*

*Proof:* We here prove the bound in (21), and delegate the computations of $\mathrm{mmse}\left(\|\mathbf{X}\|\mid\|\mathbf{Y}\|\right)$ and $\mathbb{E}\left[\|\mathbf{X}\|\mid\|\mathbf{Y}\|=\|\mathbf{y}\|\right]$ to [11, Appendix C]. We have

$$\begin{aligned}\mathrm{mmse}\left(\vec{\mathbf{X}}\mid\vec{\mathbf{Y}}\right)&=\mathbb{E}\left[\|\vec{\mathbf{X}}\|^2\right]-\mathbb{E}\left[\left\|\mathbb{E}\left[\vec{\mathbf{X}}\mid\vec{\mathbf{Y}}\right]\right\|^2\right]\\&\overset{(a)}{=}\mathbb{E}\left[\|\mathbf{X}\|^2\right]-\mathbb{E}\left[\left\|\mathbb{E}\left[\vec{\mathbf{X}}\mid\vec{\mathbf{Y}}\right]\right\|^2\right]\\&\overset{(b)}{\geq}\mathbb{E}\left[\|\mathbf{X}\|^2\right]-\mathbb{E}\left[\left(\mathbb{E}\left[\|\vec{\mathbf{X}}\|\mid\vec{\mathbf{Y}}\right]\right)^2\right]\\&\overset{(c)}{=}\mathbb{E}\left[\|\mathbf{X}\|^2\right]-\mathbb{E}\left[\left(\mathbb{E}\left[\|\mathbf{X}\|\mid\|\mathbf{Y}\|\right]\right)^2\right]\\&\overset{(d)}{=}\mathbb{E}\left[\|\mathbf{X}\|^2\right]-\mathbb{E}\left[\left(\mathbb{E}\left[\|\mathbf{X}\|\mid\mathbf{Y}\right]\right)^2\right]\\&=\mathrm{mmse}\left(\|\mathbf{X}\|\mid\mathbf{Y}\right),\end{aligned}\tag{22}$$

where the labeled (in)-equalities follow from: (a) since $\|\mathbf{X}\|=\|\vec{\mathbf{X}}\|$; (b) using modulus inequality (i.e., $\|\mathbb{E}[\mathbf{U}]\|\leq\mathbb{E}[\|\mathbf{U}\|]$ for any random vector $\mathbf{U}$); (c) and (d) using the fact that

$$\mathbb{E}\left[\|\vec{\mathbf{X}}\|\mid\vec{\mathbf{Y}}=\vec{\mathbf{r}}\right]=\mathbb{E}\left[\|\mathbf{X}\|\mid\|\mathbf{Y}\|=\|\vec{\mathbf{r}}\|\right]=\mathbb{E}\left[\|\mathbf{X}\|\mid\mathbf{Y}=\vec{\mathbf{r}}\right],$$

which is formally proved in [11, Lemma 8], and holds under the assumption of $\mathbf{X}\sim\mathcal{N}(\mathbf{0},\mathbf{I}_n)$. ∎

Observe that the approach taken in Theorem 4 is an unconventional one. Indeed, instead of taking a usual approach, such as finding a Bayesian Cramer-Rao lower bound, Theorem 4 produces a lower bound on the MMSE by finding a closed-form expression for the MMSE of an 'easier' problem, namely estimating the norm of $\mathbf{X}$ from the noisy observation $\mathbf{Y}$. While we are interested in finding lower bounds on $\mathrm{mmse}\left(\vec{\mathbf{X}}\mid\vec{\mathbf{Y}}\right)$, the result in Theorem 4 might be of an independent interest.

## V. Conclusions

In this paper, we have presented an estimation framework to study the performance of the sorting function over perturbed data. The main contribution of our work is three-fold: (1) we have analyzed the optimal MMSE estimator and showed that, under certain conditions, its structure depends on the estimators on the unsorted data; (2) we have proposed a suboptimal estimator, which offers guarantees with respect to the optimal MMSE; and (3) for an i.i.d. Gaussian input, we have derived a lower bound on the desired MMSE, which also provides the solution to a novel estimation problem that consists of estimating the norm of an unsorted vector from a noisy observation of it.

## References


[1] M. Braverman and E. Mossel, "Noisy sorting without resampling," in *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, 2008, pp. 268–276.

[2] A. Pananjady, M. J. Wainwright, and T. A. Courtade, "Denoising linear models with permuted data," in *Proceedings of the International Symposium on Information Theory (ISIT)*, 2017, pp. 446–450.

[3] E. Lloyd, "Least-squares estimation of location and scale parameters using order statistics," *Biometrika*, vol. 39, no. 1/2, pp. 88–95, 1952.

[4] H. L. Harter and N. Balakrishnan, *CRC Handbook of Tables for the Use of Order Statistics in Estimation*. CRC press, 1996.

[5] H. A. David and H. N. Nagaraja, *Order Statistics*. Wiley Online Library, 2004, vol. 9.

[6] A. S. Goldberger, "Best linear unbiased prediction in the generalized linear regression model," *Journal of the American Statistical Association*, vol. 57, no. 298, pp. 369–375, 1962.

[7] N. Doganaksoy and N. Balakrishnan, "A useful property of best linear unbiased predictors with applications to life-testing," *The American Statistician*, vol. 51, no. 1, pp. 22–28, 1997.

[8] L. Bain, *Statistical Analysis of Reliability and Life-Testing Models: Theory and Methods*. Routledge, 2017.

[9] V. Barnett and T. Lewis, *Outliers in Statistical Data*. Wiley, 1974.

[10] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," *Biometrika*, vol. 52, no. 3/4, pp. 591–611, 1965.

[11] A. Dytso, M. Cardone, M. S. Veedu, and H. V. Poor, "On estimation under noisy order statistics," *arXiv preprint arXiv:1901.06294*, 2019.