Signal Estimation with Low Infinity-Norm Error by Minimizing the Mean p-Norm Error

Jin Tan and Dror Baron
Department of Electrical and Computer Engineering
North Carolina State University
Raleigh, NC 27695, USA
Email: {jtan,barondror}@ncsu.edu

Liyi Dai
Computing Sciences Division
US Army Research Office
Research Triangle Park, NC 27709, USA
Email: liyi.dai.civ@mail.mil

***Abstract*—We consider the problem of estimating an input signal from noisy measurements in both parallel scalar Gaussian channels and linear mixing systems. The performance of the estimation process is quantified by the ℓ_∞-norm error metric (worst case error). Our previous results have shown for independent and identically distributed (i.i.d.) Gaussian mixture input signals that, when the input signal dimension goes to infinity, the Wiener filter minimizes the ℓ_∞-norm error. However, the input signal dimension is finite in practice. In this paper, we estimate the finite dimensional input signal by minimizing the mean ℓ_p-norm error. Numerical results show that the ℓ_p-norm minimizer outperforms the Wiener filter, provided that the value of p is properly chosen. Our results further suggest that the optimal value of p increases with the signal dimension, and that for i.i.d. Bernoulli-Gaussian input signals, the optimal p increases with the percentage of nonzeros.**

***Index Terms*—Gaussian mixture, ℓ_∞-norm error, linear mixing systems, parallel scalar Gaussian channels, Wiener filters.**

I. Introduction

A. Motivation

The Gaussian distribution is widely used to describe the probability densities of various types of data, owing to its mathematical advantages [2]. It has been shown that non-Gaussian distributions can often be sufficiently approximated by an infinite mixture of Gaussians [3], so that the mathematical advantages of the Gaussian distribution can be leveraged when discussing non-Gaussian data models [4, 5]. A set of parallel scalar Gaussian channels with a Gaussian mixture input signal has been used to model image denoising problems [4–6], while linear mixing systems are popular models used in many settings such as compressed sensing [7, 8], regression [9, 10], and multiuser detection [11].

Signal reconstruction from noisy measurements is prevalent in the literature, but the minimization of the ℓ_∞-norm error has received less attention. Our interest in the ℓ_∞-norm error is motivated by applications such as wireless communications [12], group testing [13], and trajectory planning in control systems [14], where we want to decrease the worst-case sensitivity to noise.

This work was supported in part by the National Science Foundation under Grant CCF-1217749 and in part by the U.S. Army Research Office under Grant W911NF-04-D-0003. Portions of the work appeared at the Information Theory and Applications workshop (ITA), San Diego, CA, Feb. 2013 [1].

B. Problem setting

Let us describe *parallel scalar Gaussian channels* and *linear mixing systems*. In both settings, the input signal $\mathbf{x}$ is generated by an independent and identically distributed (i.i.d.) Gaussian mixture source,

$$x_i \sim \sum_{k=1}^{K} s_k \cdot \mathcal{N}(\mu_k, \sigma_k^2) = \sum_{k=1}^{K} \frac{s_k}{\sqrt{2\pi\sigma_k^2}} e^{-\frac{(x_i-\mu_k)^2}{2\sigma_k^2}}, \quad (1)$$

where the subscript $(\cdot)_i$ denotes the i-th component of a sequence (or a vector), $\mu_1, \mu_2, \ldots, \mu_K$ (respectively, $\sigma_1^2, \sigma_2^2, \ldots, \sigma_K^2$) are the means (respectively, variances) of the Gaussian components, and $0 < s_1, s_2, \ldots, s_K < 1$ are the probabilities of the K Gaussian components. Note that $\sum_{k=1}^{K} s_k = 1$.

A special case of the Gaussian mixture is Bernoulli-Gaussian (sparse Gaussian),

$$x_i \sim s \cdot \mathcal{N}(\mu_x, \sigma_x^2) + (1-s) \cdot \delta(x_i),$$

for some $0 < s < 1$, μ_x, and σ_x^2, where $\delta(\cdot)$ is the delta function [2]. The zero-mean Bernoulli-Gaussian model is often used in sparse signal processing [7, 8, 11, 15–18].

In *parallel scalar Gaussian channels* [5, 6], we consider

$$\mathbf{r} = \mathbf{x} + \mathbf{z}, \quad (2)$$

where $\mathbf{r}, \mathbf{x}, \mathbf{z} \in \mathbb{R}^N$ are the output signal, the input signal, and the additive white Gaussian noise (AWGN), respectively. The AWGN channel can be described by the conditional distribution

$$f_{\mathbf{R}|\mathbf{X}}(\mathbf{r}|\mathbf{x}) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma_z^2}} \exp\left(-\frac{(r_i - x_i)^2}{2\sigma_z^2}\right), \quad (3)$$

where σ_z^2 is the variance of the Gaussian noise.

In a linear mixing system [7, 8, 11, 16], we consider

$$\mathbf{w} = \mathbf{\Phi}\mathbf{x}, \quad (4)$$

where the measurement matrix $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ is sparse and its entries are i.i.d. (detailed assumptions on $\mathbf{\Phi}$ appear in Rangan [16]). Because each component of the measurement vector $\mathbf{w} \in \mathbb{R}^M$ is a linear combination of the components of $\mathbf{x}$, we call the system (4) a *linear mixing system*. The measurements $\mathbf{w}$ are passed through a bank of separable scalar channels characterized by conditional distributions

$$f_{\mathbf{Y}|\mathbf{W}}(\mathbf{y}|\mathbf{w}) = \prod_{i=1}^{M} f_{Y|W}(y_i|w_i), \tag{5}$$

where $\mathbf{y} \in \mathbb{R}^M$ are the channel outputs. However, unlike the parallel scalar Gaussian channels (3), the channels (5) of the linear mixing system are not restricted to Gaussian [15, 16].

Our goal is to estimate the original input signal $\mathbf{x}$ either from the parallel scalar Gaussian channel outputs $\mathbf{r}$ (2) or from the linear mixing system outputs $\mathbf{y}$ and the measurement matrix $\mathbf{\Phi}$ (4, 5). To evaluate how accurate the estimation process is, we quantify the ℓ_∞-norm error between $\mathbf{x}$ and its estimate $\widehat{\mathbf{x}}$,

$$\|\widehat{\mathbf{x}} - \mathbf{x}\|_\infty = \max_{i \in \{1,\dots,N\}} |\widehat{x}_i - x_i|;$$

this error metric penalizes significant errors during the estimation process. The estimator that minimizes the mean value of $\|\widehat{\mathbf{x}} - \mathbf{x}\|_\infty$ is called the *minimum mean ℓ_∞-norm error estimator*. We denote this estimator by $\widehat{\mathbf{x}}_{\ell_\infty}$, which is

$$\widehat{\mathbf{x}}_{\ell_\infty} = \arg\min_{\widehat{\mathbf{x}}} E\left[\|\widehat{\mathbf{x}} - \mathbf{x}\|_\infty\right]. \tag{6}$$

C. Related work

In our previous work [19], we dealt with an additive error metric defined as

$$D(\widehat{\mathbf{x}}, \mathbf{x}) = \sum_{i=1}^{N} d(\widehat{x}_i, x_i), \tag{7}$$

where $d(\cdot,\cdot)$ is a scalar function. For example, when $d(\widehat{x}_i, x_i) = (\widehat{x}_i - x_i)^2$, the value of $D(\widehat{\mathbf{x}}, \mathbf{x})$ quantifies the square error between $\widehat{\mathbf{x}}$ and $\mathbf{x}$.

We proposed a metric-optimal reconstruction algorithm that minimizes the expected value of error metrics of the form (7) [19], where the reconstruction process is performed component-wise, i.e., for each $i \in \{1, 2, ..., N\}$, $d(\widehat{x}_i, x_i)$ is minimized separately. However, in contrast to (7), the ℓ_∞-norm error is not additive, because it only considers the largest component of the vector $(\widehat{\mathbf{x}} - \mathbf{x})$, and thus it is not straightforward to extend the algorithm [19] to minimize the ℓ_∞-norm error.

In a companion paper [20], we proved that, when the input signal is i.i.d. Gaussian mixture, the Wiener filter provides asymptotically optimal mean ℓ_∞-norm error in parallel scalar Gaussian channels. The Wiener filter can also be applied to linear mixing systems (4, 5) to minimize the ℓ_∞-norm error. This extension to linear mixing systems is based on settings where the measurement matrix $\mathbf{\Phi}$ is sparse and i.i.d., and a linear mixing system can be decoupled to parallel Gaussian channels [21–23]. The current paper complements the asymptotic analyses with numerical results for a heuristic algorithm. A survey of other works involving the ℓ_∞-norm error appears in our companion paper [20].

D. Contributions

While our previous results [20] are asymptotic in nature, the results provided in this paper are practical in nature. In order to deal with signals of finite dimension N in practice, we apply the ℓ_p-norm minimizer using the metric-optimal procedure by Tan et al. [19]. The numerical results show that, with a finite signal length N, the ℓ_p-norm minimizer [19] has lower ℓ_∞-norm error than the Wiener filter.

We also apply the ℓ_p-norm minimizer to linear mixing systems (4, 5). Let us highlight that the channels (5) in the linear mixing systems are not restricted to Gaussian. Again, for a finite N, the ℓ_p-norm minimizer outperforms the Wiener filter for ℓ_∞-norm error.

The remainder of the paper is arranged as follows. We review background material in Section II, and propose our new algorithm in Section III. Numerical results and a discussion appear in Section IV.

II. Background

A. Asymptotic optimality of the Wiener filter

For parallel scalar Gaussian channels (2), the *minimum mean squared error estimator*, denoted by $\widehat{\mathbf{x}}_{\ell_2}$, is obtained by the conditional expectation $E[\mathbf{x}|\mathbf{r}]$. If the input signal $\mathbf{x}$ is i.i.d. Gaussian (not Gaussian mixture), i.e., $x_i \sim \mathcal{N}(\mu_x, \sigma_x^2)$, then the Wiener filter achieves the minimum mean squared error [24]. It has been shown by Sherman [25, 26] that, besides the ℓ_2-norm error, the Wiener filter is also optimal for all ℓ_p-norm errors ($p \geq 1$), including the ℓ_∞-norm error. Surprisingly, we have found that, if the input signal is generated by an i.i.d. Gaussian mixture source, then the Wiener filter asymptotically minimizes the ℓ_∞-norm error. That is, the largest absolute error between $\mathbf{x}$ and $\widehat{\mathbf{x}}$ lies in an index that *corresponds to the Gaussian mixture component with greatest variance*.

Theorem 1. *[20] In parallel scalar Gaussian channels* (2)*, if the input signal* $\mathbf{x}$ *is generated by an i.i.d. Gaussian mixture source* (1)*, then the Wiener filter*

$$\widehat{\mathbf{x}}_{W,GM} = \frac{\sigma_m^2}{\sigma_m^2 + \sigma_z^2}(\mathbf{r} - \mu_m) + \mu_m \tag{8}$$

asymptotically minimizes the ℓ_∞-norm error, where $m = \arg\max_{k \in \{1,2,\dots,K\}} \sigma_k^2$*. More specifically,*

$$\lim_{N\to\infty} \frac{E\left[\|\mathbf{x} - \widehat{\mathbf{x}}_{W,GM}\|_\infty\right]}{E\left[\|\mathbf{x} - \widehat{\mathbf{x}}_{\ell_\infty}\|_\infty\right]} = 1, \tag{9}$$

where $\widehat{\mathbf{x}}_{\ell_\infty}$ *satisfies* (6).

Note that $E\left[\|\mathbf{x} - \widehat{\mathbf{x}}_{\ell_\infty}\|_\infty\right]$ is proportional to $\sqrt{\ln(N)}$. Therefore, equation (9) implies that the constant in front of $\sqrt{\ln(N)}$ converges to the correct constant.

B. Relaxed belief propagation

The set of parallel scalar Gaussian channels (2) has a simple structure, because each channel $r_i = x_i + z_i$ is separable from other scalar channels, and thus scalar signal estimation is optimal (vector estimation will not be better).

However, the estimation process in linear mixing systems is more complicated. Relaxed belief propagation (BP) [16] is an algorithm that deals with linear mixing systems. An important result of relaxed BP is that, after a sufficient number of iterations, relaxed BP calculates a vector $\mathbf{q} = [q_1, q_2, \ldots, q_N]^T \in \mathbb{R}^N$, such that estimating each input signal entry x_i from the corresponding q_i is asymptotically statistically equivalent to estimating the input signals $\mathbf{x}$ from the outputs $\mathbf{y}$. In particular, q_i is regarded as the output of a scalar Gaussian channel:

$$q_i = x_i + v_i \tag{10}$$

for $i \in \{1, \ldots, N\}$, where each channel's additive noise v_i is Gaussian distributed $\mathcal{N}(0, \mu_v)$, and μ_v satisfies Tanaka's fixed point equation [21–23]. The value of μ_v can also be obtained from relaxed BP [16]. We note in passing that when we discuss $q_i = x_i + v_i$, these are parallel scalar channels resulting from relaxed BP, and when we discuss $r_i = x_i + z_i$, these are true parallel scalar Gaussian channels (2).

After a linear mixing system has been decoupled to scalar Gaussian channels (10), we can apply the Wiener filter (8) to $\mathbf{q}$. Theorem 1 indicates that application of the Wiener filter to $\mathbf{q}$ asymptotically minimizes the mean ℓ_∞-norm error.

III. The mean ℓ_p-norm minimizer

A. Parallel scalar Gaussian channels

The channels $r_i = x_i + z_i$ for $i \in \{1, 2, \ldots, N\}$ are separable from each other. If the error metric is additive (7) and thus also separable, then estimating the input signal $\mathbf{x}$ from the channel outputs $\mathbf{r}$ can be reduced to scalar estimation (estimating x_i from r_i), which is simple to implement in a computationally efficient manner. However, the ℓ_∞-norm error only considers the component with greatest absolute value, and does not have an additive form (7). Recall that the definition of the ℓ_p-norm error between $\widehat{\mathbf{x}}$ and $\mathbf{x}$ is

$$\|\widehat{\mathbf{x}} - \mathbf{x}\|_p = \left(\sum_{i \in \{1, \ldots, N\}} |\widehat{x}_i - x_i|^p \right)^{1/p}.$$

This type of error is closely related to our definition of the additive error metric (7). We define

$$D_p(\widehat{\mathbf{x}}, \mathbf{x}) = \sum_{i=1}^{N} |\widehat{x}_i - x_i|^p = \|\widehat{\mathbf{x}} - \mathbf{x}\|_p^p, \tag{11}$$

and let $\widehat{\mathbf{x}}_p$ denote the estimate that minimizes the conditional expectation of $D_p(\widehat{\mathbf{x}}, \mathbf{x})$, i.e.,

$$\begin{aligned} \widehat{\mathbf{x}}_p &= \arg\min_{\widehat{\mathbf{x}}} E[D_p(\widehat{\mathbf{x}}, \mathbf{x}) | \mathbf{r}] \\ &= \arg\min_{\widehat{\mathbf{x}}} E[\|\widehat{\mathbf{x}} - \mathbf{x}\|_p^p | \mathbf{r}]. \end{aligned} \tag{12}$$

Because scalar channels are separable, equation (12) reduces to a scalar estimation,

$$\begin{aligned} \widehat{x}_{p,i} &= \arg\min_{\widehat{x}_i} E[|\widehat{x}_i - x_i|^p | r_i] \\ &= \arg\min_{\widehat{x}_i} \int |\widehat{x}_i - x_i|^p f(x_i | r_i) dx_i \end{aligned} \tag{13}$$

for $i \in \{1, 2, ..., N\}$, where $f(x_i|r_i)$ can be calculated using Bayes' rule,

$$f(x_i | r_i) = \frac{f(r_i | x_i) f(x_i)}{\int f(r_i | x_i) f(x_i) dx_i}, \tag{14}$$

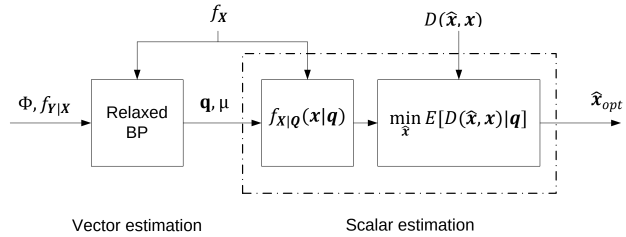


Figure 1: *The structure of the metric-optimal estimation algorithm.*

and $f(r_i|x_i)$ is obtained from (3). Although $\widehat{\mathbf{x}}_p$ minimizes the $(\ell_p)^p$ error, rather than the ℓ_p-norm error, we call $\widehat{\mathbf{x}}_p$ the *ℓ_p-norm minimizer* for simplicity. Because it can be shown that

$$\lim_{p \to \infty} \|\widehat{\mathbf{x}} - \mathbf{x}\|_p = \|\widehat{\mathbf{x}} - \mathbf{x}\|_\infty,$$

it is reasonable to expect that if we set p to a large value, then running our ℓ_p-norm minimizer (12) will give a solution that converges to an estimate that minimizes the ℓ_∞-norm error.

B. Linear mixing systems

In our previous work [19], we utilized the outputs of relaxed BP [16], and introduced a general metric-optimal estimation algorithm that deals with additive error metrics. (There are different versions of relaxed BP [16, 27, 28]; our algorithm [19] is based on the one by Rangan [16], specifically the software package "GAMP" [29].) Figure 1 illustrates the structure of our metric-optimal algorithm (dashed box). The algorithm is essentially a scalar estimation process, whereas relaxed BP decouples the vector estimation problem to scalar estimation problems.

We first compute the conditional probability density function $f_{\mathbf{X}|\mathbf{Q}}(\mathbf{x}|\mathbf{q})$ using Bayes' rule (similar to (14)), where $\mathbf{q}$ is obtained from relaxed BP. Then, given an additive error metric $D(\widehat{\mathbf{x}}, \mathbf{x})$, the optimal estimate $\widehat{\mathbf{x}}_{\text{opt}}$ is generated by minimizing the conditional expectation of the error metric $E[D(\widehat{\mathbf{x}}, \mathbf{x})|\mathbf{q}]$:

$$\widehat{\mathbf{x}}_{\text{opt}} = \arg\min_{\widehat{\mathbf{x}}} E[D(\widehat{\mathbf{x}}, \mathbf{x}) | \mathbf{q}]. \tag{15}$$

In the large system limit, the estimate satisfying (15) is asymptotically optimal, because it minimizes the conditional expectation of the error metric. Similar to (12) and (13), the estimate $\widehat{\mathbf{x}}_{\text{opt}}$ is solved in a component-wise fashion:

$$\widehat{x}_{\text{opt},i} = \arg\min_{\widehat{x}_i} \int D(\widehat{x}_i, x_i) f(x_i | q_i) dx_i, \tag{16}$$

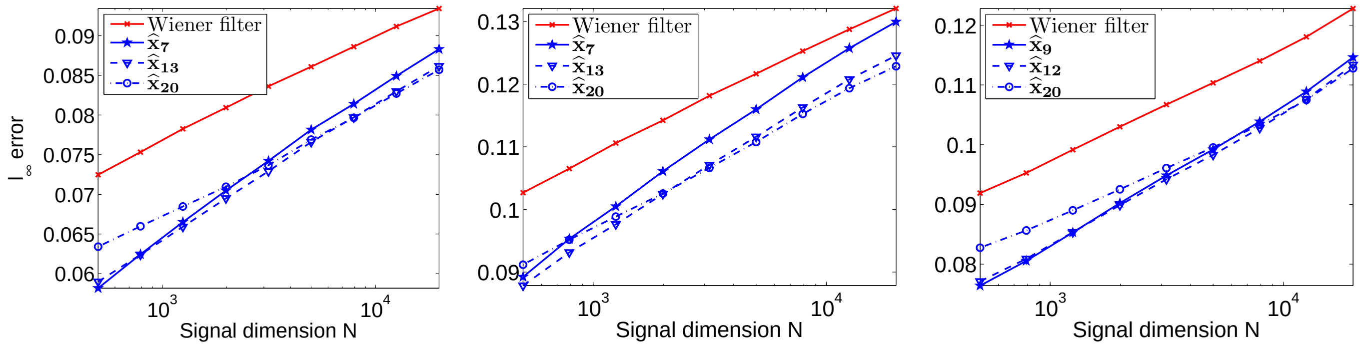
for each x_i, $i \in \{1, 2, \ldots, N\}$. Finally, the ℓ_p-norm minimizer for linear mixing systems is

$$\widehat{x}'_{p,i} = \arg\min_{\widehat{x}_i} \int |\widehat{x}_i - x_i|^p f(x_i | q_i) dx_i \tag{17}$$

for $i \in \{1, 2, \ldots, N\}$.

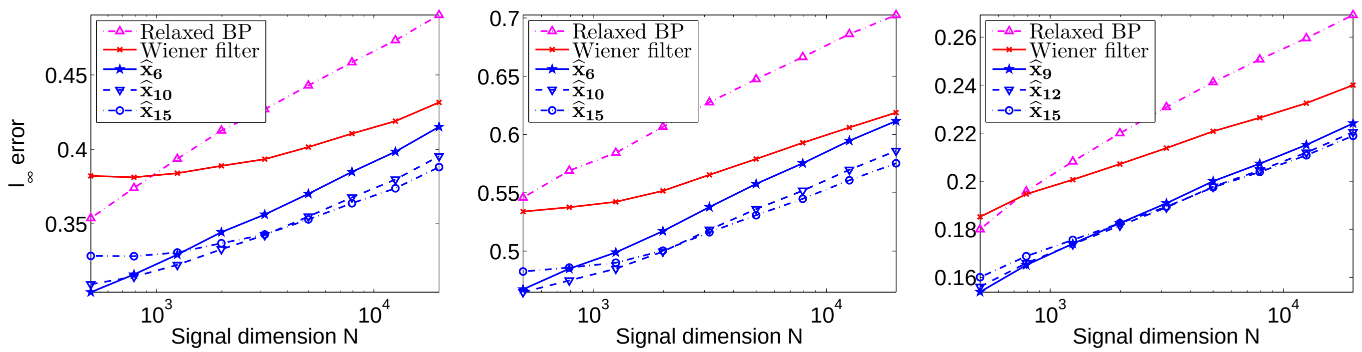
IV. Numerical Results

Recall that the Wiener filter (8) is asymptotically optimal, but its performance for finite N is not clear. On the other hand, the ℓ_p-norm minimizer is a heuristic for finite N. Let us compare the two approaches numerically for both parallel scalar Gaussian channels (2) and linear mixing systems (4, 5).



(a) Case 1 (b) Case 2 (c) Case 3

Figure 2: *The performance of the Wiener filter and the ℓ_{p_1}, ℓ_{p_2}, ℓ_{p_3}-norm minimizers in terms of ℓ_∞-norm error in parallel scalar Gaussian channels. The optimal p_{opt} increases as N increases. SNR is 20 dB. Signals involved in Cases 1, 2, and 3 are described in Section IV-A. (a)* $p_1 = 7$, $p_2 = 13$, $p_3 = 20$. *(b)* $p_1 = 7$, $p_2 = 13$, $p_3 = 20$. *(c)* $p_1 = 9$, $p_2 = 12$, $p_3 = 20$.



(a) Case 1 (b) Case 2 (c) Case 3

Figure 3: *The performance of the Wiener filter and the ℓ_{p_1}, ℓ_{p_2}, ℓ_{p_3}-norm minimizers in terms of ℓ_∞-norm error in linear mixing systems. The optimal p_{opt} increases as N increases. SNR is 20 dB. Signals involved in Cases 1, 2, and 3 are described in Section IV-A. (a)* $p_1 = 6$, $p_2 = 10$, $p_3 = 15$, $M/N = 0.3$. *(b)* $p_1 = 6$, $p_2 = 10$, $p_3 = 15$, $M/N = 0.4$. *(c)* $p_1 = 9$, $p_2 = 12$, $p_3 = 15$, $M/N = 0.4$.

A. Parallel scalar Gaussian channels

We first test for the parallel scalar Gaussian channels $\mathbf{r} = \mathbf{x} + \mathbf{z}$ (2), where the input signal $\mathbf{x}$ is generated by the following 3 sources,

- Case 1: i.i.d. sparse Gaussian with sparsity rate 0.05, $x_i \sim 0.05 \cdot \mathcal{N}(0,1) + 0.95 \cdot \delta(x_i)$.
- Case 2: i.i.d. sparse Gaussian with sparsity rate 0.1, $x_i \sim 0.1 \cdot \mathcal{N}(0,1) + 0.9 \cdot \delta(x_i)$.
- Case 3: i.i.d. Gaussian mixture, $x_i \sim 0.01 \cdot \mathcal{N}(0,2) + 0.03 \cdot \mathcal{N}(0,1) + 0.06 \cdot \mathcal{N}(0,0.5) + 0.9 \cdot \delta(x_i)$.

The noise is $z_i \sim \mathcal{N}(0, \sigma_z^2)$, where the noise variance σ_z^2 is set such that the signal to noise ratio (SNR) satisfies

$$\text{SNR} = \frac{E\left[\sum_{i=1}^N x_i^2\right]}{E\left[\sum_{i=1}^N z_i^2\right]} = 100 = 20 \text{ dB}.$$

The Wiener filter is calculated by equation (8). For each case, we choose 3 values of p (p_1, p_2, and p_3), and obtain the ℓ_p-norm minimizers by equation (12). Figure 2 illustrates the ℓ_∞-norm errors for Cases 1–3. In each subfigure, we compare the ℓ_∞-norm errors achieved by the Wiener filter (solid line with cross markers) and by the ℓ_p-norm error minimizers, $\widehat{\mathbf{x}}_{p_1}$ (solid line with pentagram markers), $\widehat{\mathbf{x}}_{p_2}$ (dashed with inverse triangles), and $\widehat{\mathbf{x}}_{p_3}$ (dash-dot with circles). The horizontal axis represents the signal dimension N varying from 500–20,000, and the vertical axis represents the mean ℓ_∞-norm error. The values shown are averages over 15,000 repeated tests.

It is shown in Figure 2 that the curves corresponding to the Wiener filter increase more slowly as a function of N than the curves corresponding to the ℓ_p-norm minimizers. Therefore, we can see that the Wiener filter minimizes the mean ℓ_∞-norm error when $N \to \infty$.

At the same time, the ℓ_p-norm error minimizers (12) with different values of p indeed outperform the Wiener filter for finite N. For each N, there exists an optimal value of p such that $\widehat{\mathbf{x}}_p$ outperforms all the other estimators; and the optimal p increases as N increases. An intuitive explanation is that as N increases, the probability that larger errors occur also increases, and thus a larger p in (11) is used to suppress larger outliers.

Figures 2(a) and 2(b) correspond to different sparsity rates, $s = 0.05$ and 0.1. The figures suggest that the value of p_{opt} for a fixed signal dimension N is related to the sparsity rate of the input signal $\mathbf{x}$.

B. Linear mixing systems

We perform simulations for linear mixing systems (4, 5) using the software package "GAMP" [29] and our metric-optimal algorithm [30]. Our metric-optimal software package [30] automatically computes equation (16) where the distortion function (7) is given as the input of the algorithm.

Again, we test for settings where input signals are generated by the 3 different sources (Cases 1–3 in Section IV-A). The measurement matrices $\mathbf{\Phi}$ are i.i.d. zero-mean Gaussian and are normalized to have unit-norm columns. The channels (5) are AWGN, and the noise variance is set such that the SNR is 20 dB. The number of measurements M in the matrices $\mathbf{\Phi}$ varies as follows: (*i*) in Case 1, $M/N = 0.3$; (*ii*) in Case 2, $M/N = 0.4$; and (*iii*) in Case 3, $M/N = 0.4$. (Cases 2 and 3 are more complicated signals, and so more measurements are necessary for reasonable performance.) As before, the signal dimension N ranges from 500 to $20,000$.

We run the Wiener filter approach (8), relaxed BP [16, 29], and our ℓ_p-norm minimizers (17) with different values of p. Figure 3 shows results for the 3 cases. It can be seen from Figure 3 that the Wiener filter outperforms relaxed BP (dash-dot line with triangle markers) for most N, whereas $\widehat{\mathbf{x}}_p$'s outperform the Wiener filter. Moreover, as N increases, p_{opt} increases.

To summarize, the heuristic approach of the ℓ_p-norm minimizers complements the asymptotic optimality of the Wiener filter. On the other hand, it seems unlikely that this heuristic is truly optimal, and we leave the design of signal estimation algorithms with optimal ℓ_∞-norm error for future work.

ACKNOWLEDGMENTS

We thank Nikhil Krishnan for useful discussions, which greatly helped us improve this manuscript.

REFERENCES

[1] J. Tan and D. Baron, "Signal reconstruction in linear mixing systems with different error metrics," in *Inf. Theory Appl. Workshop*, Feb. 2013.

[2] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. McGraw Hill Book Co., 1991.

[3] T. Alecu, S. Voloshynovskiy, and T. Pun, "The Gaussian transform of distributions: Definition, computation and application," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2976–2985, Aug. 2006.

[4] ——, "Denoising with infinite mixture of Gaussians," in *Eur. Signal Process. Conf.*, Sep. 2005.

[5] A. Bijaoui, "Wavelets, Gaussian mixtures and Wiener filtering," *Signal Process.*, vol. 82, no. 4, pp. 709–712, Apr. 2002.

[6] M. Tabuchi, N. Yamane, and Y. Morikawa, "Adaptive Wiener filter based on Gaussian mixture model for denoising chest X-ray CT image," in *Proc. IEEE SICE Annu. Conf.*, Sep. 2007, pp. 682–689.

[7] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[8] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[9] P. Huber, "Robust regression: Asymptotics, conjectures and Monte Carlo," *Ann. Stat.*, vol. 1, no. 5, pp. 799–821, 1973.

[10] D. O'Leary, "Robust regression computation using iteratively reweighted least squares," *SIAM J. Matrix Anal. Appl.*, vol. 11, no. 3, pp. 466–480, July 1990.

[11] D. Guo and C. C. Wang, "Multiuser detection of sparsely spread CDMA," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 3, pp. 421–431, Apr. 2008.

[12] C. Studer, W. Yin, and R. G. Baraniuk, "Signal representations with minimum ℓ_∞-norm," in *Proc. 50th Allerton Conf. Commun., Control, Comput.*, Oct. 2012.

[13] A. C. Gilbert, B. Hemenway, A. Rudra, M. J. Strauss, and M. Wootters, "Recovering simple signals," in *Inf. Theory Appl. Workshop*, Feb. 2012, pp. 382–391.

[14] M. Egerstedt and C. F. Martin, "Trajectory planning in the infinity norm for linear control systems," *Int. J. Control*, vol. 72, no. 13, pp. 1139–1146, 1999.

[15] D. Guo and C. C. Wang, "Random sparse linear systems observed via arbitrary channels: A decoupling principle," in *Proc. IEEE Int. Symp. Inf. Theory*, June 2007, pp. 946–950.

[16] S. Rangan, "Estimation with random linear mixing, belief propagation and compressed sensing," *ArXiv preprint arXiv:1001.2228*, Jan. 2010.

[17] J. Starck, F. Murtagh, and J. Fadili, *Sparse image and signal processing: Wavelets, curvelets, morphological diversity*. Cambridge Univ. Press, 2010.

[18] J. Vila and P. Schniter, "Expectation-maximization Bernoulli-Gaussian approximate message passing," in *Proc. IEEE 45th Asilomar Conf. Signals, Syst., and Comput.*, Nov. 2011.

[19] J. Tan, D. Carmon, and D. Baron, "Signal estimation with additive error metrics in compressed sensing," *IEEE Trans. Inf. Theory*, vol. 60, no. 1, pp. 150–158, Jan. 2014.

[20] J. Tan, D. Baron, and L. Dai, "Wiener filters in Gaussian mixture signal estimation with ℓ_∞-norm error," *submitted*, Jan. 2014.

[21] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," *IEEE Trans. Inf. Theory*, vol. 51, no. 6, pp. 1983–2010, June 2005.

[22] D. Guo, D. Baron, and S. Shamai, "A single-letter characterization of optimal noisy compressed sensing," in *Proc. 47th Allerton Conf. Commun., Control, Comput.*, Sep. 2009, pp. 52–59.

[23] T. Tanaka, "A statistical-mechanics approach to large-system analysis of CDMA multiuser detectors," *IEEE Trans. Inf. Theory*, vol. 48, no. 11, pp. 2888–2910, Nov. 2002.

[24] N. Wiener, *Extrapolation, interpolation, and smoothing of stationary time series with engineering applications*. MIT press, 1949.

[25] S. Sherman, "A theorem on convex sets with applications," *Ann. Math. Stat.*, vol. 26, no. 4, pp. 763–767, Dec. 1955.

[26] ——, "Non-mean-square error criteria," *IEEE Trans. Inf. Theory*, vol. 4, no. 3, pp. 125–126, Sep. 1958.

[27] D. Guo and C. Wang, "Asymptotic mean-square optimality of belief propagation for sparse linear systems," in *IEEE Inf. Theory Workshop*, Oct. 2006, pp. 194–198.

[28] D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, pp. 269–280, Jan. 2010.

[29] S. Rangan, A. Fletcher, V. Goyal, U. Kamilov, J. Parker, and P. Schniter, "GAMP," http://gampmatlab.wikia.com/wiki/Generalized_Approximate_Message_Passing/.

[30] J. Tan, D. Carmon, and D. Baron, "Arbitrary Error Metrics Software," http://people.engr.ncsu.edu/dzbaron/software/arb_metric/.