OPTIMAL ESTIMATION WITH ARBITRARY ERROR METRICS IN COMPRESSED SENSING

Jin Tan, Danielle Carmon, and Dror Baron

Department of Electrical and Computer Engineering
North Carolina State University, Raleigh, USA
Email: {jtan, dmcarmon, dzbaron}@ncsu.edu

ABSTRACT

Noisy compressed sensing deals with the estimation of a system input from its noise-corrupted linear measurements. The performance of the estimation is usually quantified by some standard error metric such as squared error or support error. In this paper, we consider a noisy compressed sensing problem with any arbitrary error metric. We propose a simple, fast, and general algorithm that estimates the original signal by minimizing an arbitrary error metric defined by the user. We verify that, owing to the decoupling principle, our algorithm is optimal, and we describe a general method to compute the fundamental information-theoretic performance limit for any well-defined error metric. We provide an example where the metric is absolute error and give the theoretical performance limit for it. The experimental results show that our algorithm outperforms methods such as relaxed belief propagation, and reaches the suggested theoretical limit for our example error metric.

Index Terms— Belief propagation, compressed sensing, error metric, estimation theory

1. INTRODUCTION

Consider a linear system,

$$\mathbf{w} = \mathbf{\Phi}\mathbf{x}, \tag{1}$$

where the system input $\mathbf{x} \in \mathbb{R}^N$ is independent and identically distributed (i.i.d.), and the random linear mixing matrix [1] $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$ is known, where typically $M < N$. The vector $\mathbf{w} \in \mathbb{R}^M$ is called the measurement of $\mathbf{x}$, and is passed through a bank of separable channels characterized by the conditional distributions,

$$f_{\mathbf{Y}|\mathbf{W}}(\mathbf{y}|\mathbf{w}) = \prod_{i=1}^{M} f_{\mathbf{Y}|\mathbf{W}}(y_i|w_i). \tag{2}$$

Note that the channels are general and are not restricted to Gaussian. We observe the channel output $\mathbf{y}$, and want to estimate the original input signal $\mathbf{x}$ from $\mathbf{y}$ and $\mathbf{\Phi}$.

The performance of the estimation is often characterized by some error metric that quantifies the distance between the estimated and the original signals. For a signal $\mathbf{x}$ and its estimate $\widehat{\mathbf{x}}$, both of length N, the error between them is the summation over the component-wise errors,

$$D(\widehat{\mathbf{x}}, \mathbf{x}) = \sum_{i=1}^{N} d(\widehat{x}_i, x_i). \tag{3}$$

Squared error is most commonly used as the error metric in estimation problems with the same or similar model described by (1) and (2). Mean-square optimal analysis and algorithms were introduced in [2–6] to estimate the original signal from Gaussian-noise corrupted measurements; in [1, 7, 8], further discussions were made given the circumstances where the output channel was arbitrary, while, again, the minimum mean square error (MMSE) estimator was put forth. Support recovery error is another metric of great importance, for example it relates to properties of the measurement matrices [9]. The authors of [9, 10] discussed the support error rate when recovering a sparse signal from its noisy measurements; support-related performance metrics were applied in the derivations of theoretical bounds on the sampling rate for signal recovery [11]. The readers may notice that previous work only paid attention to limited types of error metrics. What if absolute error, cubic error, or other non-standard metrics are required in a certain application?

In this paper (*i*) we suggest an estimation algorithm that minimizes an arbitrary error metric; and (*ii*) we prove that the algorithm is optimal and study the best possible performance of an estimator for a given error metric. This algorithm applies the relaxed belief propagation (BP) method [1, 3] and the decoupling principle [2, 4, 5, 7]. Furthermore, it is simple and fast, and it reconstructs the original signal based on minimizing the conditional expected error metric required by the users. This is convenient for users who desire to recover the original signal with an arbitrary non-standard error metric. Some simulation results show that our method outperforms algorithms such as the relaxed BP described in [1], which is optimal for squared error. Moreover, we compare our algorithm with the suggested theoretical limit for minimum mean absolute error (MMAE), and illustrate that our algorithm is optimal.

2. REVIEW OF RELAXED BELIEF PROPAGATION

Before describing the estimation algorithm, a review of the relaxed BP method [1, 3] is helpful.

Belief Propagation (BP) [12] is an iterative method used to compute the marginals of a Bayesian network. The method is based on a bipartite graph, called *Tanner* or *factor* graph, which consists of nodes and edges that represent the random variables and their relations. The marginals of the variables are computed by passing messages through nodes. In relaxed BP [3], means and variances of the variables serve as the messages passed through nodes in the Tanner graph. As a result, the mean and variance of $\mathbf{x}$ conditioned on $\mathbf{y}$ and $\mathbf{\Phi}$ are obtained. In [1, 7], this method was extended to a more general case where the channel is not necessarily Gaussian.

An important result for BP in compressed sensing is that it decouples the linear mixed channels (2) to a series of parallel scalar Gaussian channels [2, 4, 5, 7]. In other words, the input signal $\mathbf{x}$ passing through the linear mixing system (1), (2) is equivalent to each input entry x_j passing through a scalar Gaussian channel:

$$q_j = x_j + v_j, \quad \text{for } i = 1, 2, 3, ..., N, \tag{4}$$

where each channel's additive Gaussian noise v_j is $\mathcal{N}(0, \mu)$ distributed, and μ satisfies the fixed point equation discussed in [2, 4, 5, 8, 13, 14].

3. ESTIMATION ALGORITHM

3.1. Algorithm

We use four terms in our algorithm: (*i*) a distribution function $f_{\mathbf{X}}(\mathbf{x})$, the prior of the original input $\mathbf{x}$; (*ii*) a vector $\mathbf{q} = (q_1, q_2, ..., q_N)$, the outputs of the scalar Gaussian channels; (*iii*) a scalar μ, the variance of the Gaussian noise in (4); and (*iv*) an error metric function $D(\widehat{\mathbf{x}}, \mathbf{x})$ specified by the user.

Now that we know that the scalar channels have additive Gaussian noise, and that the variances of the noise are μ, we can compute the conditional probability density function $f_{\mathbf{X}|\mathbf{Q}}(\mathbf{x}|\mathbf{q})$ immediately from Bayes' rule.

Given an error metric $D(\widehat{\mathbf{x}}, \mathbf{x})$, the optimal estimation $\widehat{\mathbf{x}}_{\text{opt}}$ is generated by minimizing the conditional expectation of the error metric $E[D(\widehat{\mathbf{x}}, \mathbf{x})|\mathbf{q}]$, which is easy to compute using $f_{\mathbf{X}|\mathbf{Q}}(\mathbf{x}|\mathbf{q})$:

$$E[D(\widehat{\mathbf{x}}, \mathbf{x})|\mathbf{q}] = \int D(\widehat{\mathbf{x}}, \mathbf{x}) f_{\mathbf{X}|\mathbf{Q}}(\mathbf{x}|\mathbf{q}) d\mathbf{x}$$
$$= \int D(\widehat{\mathbf{x}}, \mathbf{x}) \frac{1}{\sqrt{(2\pi\mu)^N}} \exp\left(-\frac{\|\mathbf{q}-\mathbf{x}\|^2}{2\mu}\right) d\mathbf{x}. \tag{5}$$

Then,

$$\widehat{\mathbf{x}}_{\text{opt}} = \arg\min_{\widehat{\mathbf{x}}} \int D(\widehat{\mathbf{x}}, \mathbf{x}) f_{\mathbf{X}|\mathbf{Q}}(\mathbf{x}|\mathbf{q}) d\mathbf{x}. \tag{6}$$

Under a condition that $D(\widehat{\mathbf{x}}, \mathbf{x})$ is continuous and differentiable almost everywhere, we simply take the derivative of (5), set it equal to zero, and solve for $\widehat{\mathbf{x}}_{\text{opt}}$,

$$\left.\frac{\partial \int D(\widehat{\mathbf{x}}, \mathbf{x}) f_{\mathbf{X}|\mathbf{Q}}(\mathbf{x}|\mathbf{q}) d\mathbf{x}}{\partial \widehat{\mathbf{x}}}\right|_{\widehat{\mathbf{x}}=\widehat{\mathbf{x}}_{\text{opt}}} = \mathbf{0}. \tag{7}$$

Since both the error metric function $D(\widehat{\mathbf{x}}, \mathbf{x})$ and the conditional probability $f_{\mathbf{X}|\mathbf{Q}}(\mathbf{x}|\mathbf{q})$ are separable, the problem reduces to scalar estimation [15].

3.2. Theoretical results

Having discussed the algorithm, we now give a theoretical justification for its performance. Note that the decoupling principle that we use in our proofs is based on the replica method in statistical physics, and is not rigorous [2]. Therefore, our results are given as claims.

Claim 1. *Given the system model described by* (1), (2) *and an error metric* $D(\widehat{\mathbf{x}}, \mathbf{x})$ *of the form defined by* (3)*, the optimal estimation of the input signal is given by*

$$\widehat{\mathbf{x}}_{opt} = \arg\min_{\widehat{\mathbf{x}}} E\left[D(\widehat{\mathbf{x}}, \mathbf{x})|\mathbf{q}\right], \tag{8}$$

where the vector $\mathbf{q}$ *is the output of the decoupled Gaussian scalar channel* (4)*.*

Proof. From the main result in [2], linear mixed channels can be decoupled to a bank of scalar Gaussian channels $q_i = x_i + v_i$ where $v_i \sim \mathcal{N}(0, \mu)$ for $i \in \{1, 2, ..., N\}$. Then, $f_{X_i|\mathbf{Y}}(x_i|\mathbf{y})$ gives the same distribution as $f_{X_i|Q_i}(x_i|q_i)$. In other words, once we know the value of μ, estimating each x_i from all channel outputs $\mathbf{y} = (y_1, y_2, ..., y_M)$ is equivalent to estimating x_i from the corresponding scalar channel output q_i. Therefore, an estimator based on minimizing the conditional expectation of the error metric, $E\left(D(\widehat{\mathbf{x}}, \mathbf{x})|\mathbf{q}\right)$, gives the best possible result. □

Following Claim 1, we immediately get the optimal performance limit for any error metric $D(\widehat{\mathbf{x}}_{\text{opt}}, \mathbf{x})$.

Claim 2. *With the optimal estimator* $\widehat{\mathbf{x}}_{opt}$ *determined by* (6)*, the minimum mean user-defined error (MMUE) is given by the form shown in* (9)*, where* $R(\cdot)$ *represents the range of a variable.*

$$\text{MMUE}(f_{\mathbf{X}}, \mu) = E[D(\widehat{\mathbf{x}}_{\text{opt}}, \mathbf{x})] = \int_{R(\mathbf{Q})} \left(\int_{R(\mathbf{X})} D(\widehat{\mathbf{x}}_{\text{opt}}, \mathbf{x}) \left(\frac{1}{\sqrt{(2\pi\mu)^N}} \exp\left(-\frac{\|\mathbf{q}-\mathbf{x}\|^2}{2\mu} \right) \right) f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} \right) d\mathbf{q}. \tag{9}$$

Proof.

$$\begin{aligned} & \text{MMUE}(f_{\mathbf{X}}, \mu) = E[D(\widehat{\mathbf{x}}_{\text{opt}}, \mathbf{x})] \\ = & \int_{R(\mathbf{Q})} E_{\mathbf{Q}}\big[E[D(\widehat{\mathbf{x}}_{\text{opt}}, \mathbf{x})|\mathbf{q}]\big] d\mathbf{q} \\ = & \int_{R(\mathbf{Q})} E[D(\widehat{\mathbf{x}}_{\text{opt}}, \mathbf{x})|\mathbf{q}] f(\mathbf{q}) d\mathbf{q} \\ = & \int_{R(\mathbf{Q})} \left(\int_{R(\mathbf{X})} D(\widehat{\mathbf{x}}_{\text{opt}}, \mathbf{x}) f(\mathbf{x}|\mathbf{q}) d\mathbf{x} \right) f(\mathbf{q}) d\mathbf{q} \\ = & \int_{R(\mathbf{Q})} \left(\int_{R(\mathbf{X})} D(\widehat{\mathbf{x}}_{\text{opt}}, \mathbf{x}) \frac{f(\mathbf{q}|\mathbf{x}) f(\mathbf{x})}{f(\mathbf{q})} d\mathbf{x} \right) f(\mathbf{q}) d\mathbf{q} \\ = & \int_{R(\mathbf{Q})} \int_{R(\mathbf{X})} D(\widehat{\mathbf{x}}_{\text{opt}}, \mathbf{x}) \frac{1}{\sqrt{(2\pi\mu)^N}} \exp\left(-\frac{(\mathbf{q}-\mathbf{x})^2}{2\mu} \right) \\ & f_{\mathbf{X}}(\mathbf{x}) d\mathbf{x} d\mathbf{q}. \end{aligned}$$

□

3.3. Example

We now provide an example of absolute error, in order to illustrate how our approach can be utilized for specific error metrics.

Since the minimum mean square error (MMSE) estimator is the mean of the conditional distribution, the outliers in the set of data may corrupt the estimation, and in this case the minimum mean absolute error (MMAE), which is the median of the data, is a good alternative.

The error metric function (3) for absolute error (AE) is defined as

$$d_{\text{AE}}(\widehat{x_i}, x_i) = |\widehat{x_i} - x_i|. \tag{10}$$

The estimator $\widehat{x}_{i,\text{MMAE}}$ that minimizes $E\left[d_{\text{AE}}(\widehat{x_i}, x_i)|q_i\right]$ is such that

$$\int_{-\infty}^{\widehat{x}_{i,\text{MMAE}}} f(x_i|q_i) dx_i = \int_{\widehat{x}_{i,\text{MMAE}}}^{+\infty} f(x_i|q_i) dx_i = \frac{1}{2}.$$

Then, the conditional mean absolute error is,

$$\begin{aligned} & E[|\widehat{x}_{i,\text{MMAE}} - x_i||\, q_i] \\ = & \int_{-\infty}^{+\infty} |\widehat{x}_{i,\text{MMAE}} - x_i| f(x_i|q_i) dx_i \\ = & \int_{-\infty}^{\widehat{x}_{i,\text{MMAE}}} (-x_i) f(x_i|q_i) dx_i + \int_{\widehat{x}_{i,\text{MMAE}}}^{+\infty} x_i f(x_i|q_i) dx_i. \end{aligned}$$

Therefore, the MMAE for location i, $\text{MMAE}_i(f_{X_i}, \mu)$, is

$$\begin{aligned} & \text{MMAE}_i(f_{X_i}, \mu) = E[|\widehat{x}_{i,\text{MMAE}} - x_i|] \\ = & \int_{-\infty}^{+\infty} E[|\widehat{x}_{i,\text{MMAE}} - x_i||\, q_i] f(q_i) dq_i \\ = & \int_{-\infty}^{+\infty} \int_{-\infty}^{\widehat{x}_{i,\text{MMAE}}} (-x_i) f(x_i|q_i) dx_i f(q_i) dq_i \\ & + \int_{-\infty}^{+\infty} \int_{\widehat{x}_{i,\text{MMAE}}}^{+\infty} x_i f(x_i|q_i) dx_i f(q_i) dq_i, \end{aligned} \tag{11}$$

where the integrations are evaluated numerically.

Since the input $\mathbf{x}$ is i.i.d., and the decoupled scalar channels have the same parameter μ, the values of MMAE_i for all $i \in \{1, 2, ..., N\}$ are the same, and the overall MMAE is

$$\text{MMAE}(f_{\mathbf{X}}, \mu) = N \cdot \text{MMAE}_i(f_{X_i}, \mu). \tag{12}$$

4. NUMERICAL SIMULATIONS

Some experimental results are shown in this section in order to illustrate the performance of our estimation algorithm when minimizing a user-defined error metric.

We test our estimation algorithm on a linear system modeled by (1) and (2), where the input is sparse Gaussian and the channel is Gaussian. The input's length N is 10,000, and its sparsity rate is 3%, meaning that the entries of the input vector are non-zero with probability 3%, and are zero otherwise. The matrix $\mathbf{\Phi}$ we use is Bernoulli-0.5 distributed, and is normalized to have unit-norm rows. The non-zero input entries are $\mathcal{N}(0, 1)$ distributed, and the Gaussian noise is $\mathcal{N}(0, 0.003)$ distributed, i.e., the signal to noise ratio (SNR) is 20dB. Keep in mind that the channels (2) are in general form, and we provide a Gaussian channel example for brevity.

In order to illustrate that our estimation algorithm is suitable for reasonable error metrics, we considered absolute error and two other non-standard metrics:

$$\text{Error}_p = \sum_{j=1}^{N} |\widehat{x}_j - x_j|^p, \tag{13}$$

where $p = 0.5$ or 1.5.

In Figure 1, lines marked with "metric-optimal" present the error of our estimation algorithm, and lines marked with "Relaxed BP" show the error of the relaxed BP algorithm [1]. Each point in the figure is an average of 40 experiments with the same parameters. It is clear that our algorithm outperforms relaxed BP, especially for mean absolute error and $\text{Error}_{0.5}$.

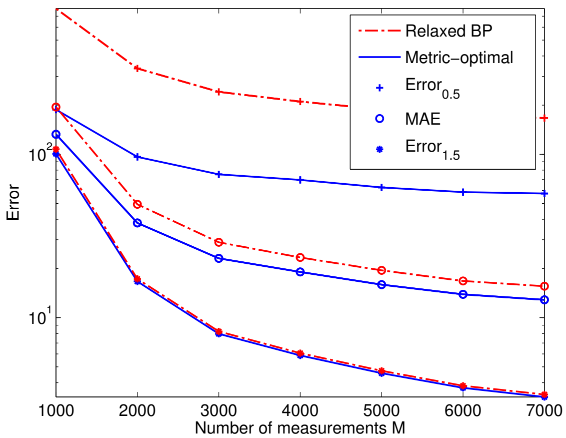

Fig. 1. Comparison of the metric-optimal estimation algorithm and the relaxed BP algorithm.

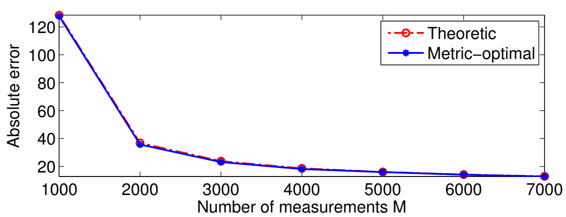

Fig. 2. Comparison of our minimum mean absolute error (MMAE) estimator and the theoretical limit (12).

To demonstrate the theoretical analysis of our algorithm in Section 3, we compare our MMAE estimator results with the theoretical performance limit (12) in Figure 2, where the integrations (11) are computed numerically. In the figure, each point on the "metric-optimal" line is also generated by averaging 40 experiments. It is shown that our estimation algorithm reaches the limit and is thus optimal.

Acknowledgment

We thank Dr. Sundeep Rangan for kindly providing the Matlab code [16] of the relaxed BP algorithm [1, 8].

5. REFERENCES

[1] S. Rangan, "Estimation with random linear mixing, belief propagation and compressed sensing," in *44th Annu. Conf. Inform. Sci. and Syst. (CISS)*. IEEE, 2010, pp. 1–6.

[2] D. Guo and S. Verdú, "Randomly spread CDMA: Asymptotics via statistical physics," vol. 51, no. 6, pp. 1983–2010, 2005.

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

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

[5] S. Rangan, A.K. Fletcher, and V.K. Goyal, "Asymptotic analysis of MAP estimation via the replica method and applications to compressed sensing," *Arxiv preprint arXiv:0906.3234*, 2009.

[6] D. Baron, S. Sarvotham, and R.G. Baraniuk, "Bayesian compressive sensing via belief propagation," vol. 58, no. 1, pp. 269–280, 2010.

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

[8] S. Rangan, "Generalized approximate message passing for estimation with random linear mixing," in *IEEE Int. Symp. Inform. Theory*. IEEE, 2011, pp. 2168–2172.

[9] W. Wang, M.J. Wainwright, and K. Ramchandran, "Information-theoretic limits on sparse signal recovery: Dense versus sparse measurement matrices," vol. 56, no. 6, pp. 2967–2979, 2010.

[10] A. Tulino, G. Caire, S. Shamai, and S. Verdú, "Support recovery with sparsely sampled free random matrices," in *IEEE Int. Symp. Inform. Theory*. IEEE, 2011, pp. 2328–2332.

[11] M. Akçakaya and V. Tarokh, "Shannon-theoretic limits on noisy compressive sampling," vol. 56, no. 1, pp. 492–504, 2010.

[12] C.M. Bishop, *Pattern recognition and machine learning*, vol. 4, Springer New York, 2006.

[13] A. Montanari and D. Tse, "Analysis of belief propagation for non-linear problems: The example of CDMA (or: How to prove Tanaka's formula)," in *Proc. IEEE Inform. Theory Workshop*. IEEE, 2006, pp. 160–164.

[14] D.L. Donoho, A. Maleki, and A. Montanari, "Message-passing algorithms for compressed sensing," *Proc. Nat. Academy of Sci.*, vol. 106, no. 45, pp. 18914, 2009.

[15] B.C. Levy, *Principles of signal detection and parameter estimation*, Springer Verlag, 2008.

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