

COMPRESSED MATCHED FILTER FOR NON-GAUSSIAN NOISE

Jakob Vovnoboy and Ami Wiesel

The Rachel and Selim Benin School of Computer Science and Engineering,
The Hebrew University of Jerusalem

ABSTRACT

We consider estimation of a deterministic unknown parameter vector in a linear model with non-Gaussian noise. In the Gaussian case, dimensionality reduction via a linear matched filter provides a simple low dimensional sufficient statistic which can be easily communicated and/or stored for future inference. Such a statistic is usually unknown in the general non-Gaussian case. Instead, we propose a hybrid matched filter coupled with a randomized compressed sensing procedure, which together create a low dimensional statistic. We also derive a complementary algorithm for robust reconstruction given this statistic. Our recovery method is based on the fast iterative shrinkage and thresholding algorithm which is used for outlier rejection given the compressed data. We demonstrate the advantages of the proposed framework using synthetic simulations.

Index Terms— Matched filter, robust regression, compressed sensing, JMAP-ML.

1. INTRODUCTION

One of the most fundamental concepts in parameter estimation is sufficient statistics. These are functions of the observations that summarize all the information associated with the parameter of interest [6]. Sufficient statistics minimize the required storage and communication resources. They are task independent and are useful when the data has to be compressed for a future specific use. Their computation usually involves simple and low complexity operations that are suitable for high rate processing. A sufficient statistic for estimation of a deterministic unknown parameter vector in a linear model with Gaussian noise is the well known matched filter. This simple linear operation is a core ingredient of most radar and communication systems.

Many of the modern physical systems are better modeled as linear systems with non-Gaussian noise rather than Gaussian, this mainly due to impulsive noise phenomena [3–5, 10–12, 15, 19]. Typical noise characteristics include generalized Gaussian distributions, mixture distributions, impulsive models, and heavy tailed models. In such scenarios, a low dimensional sufficient statistic is usually unknown, hence the classical matched filter is generally sub-optimal and more complicated non-linear operations are required [14].

A common estimation technique in systems with non-Gaussian noise is to use non-linear element-wise limiters (also known as clipper) prior to the linear filter [12]. However, this method is efficient only when the dynamic range of the data is small compared to the outliers. Another traditional solution in statistics is to resort to robust regression methods which works directly with the observed

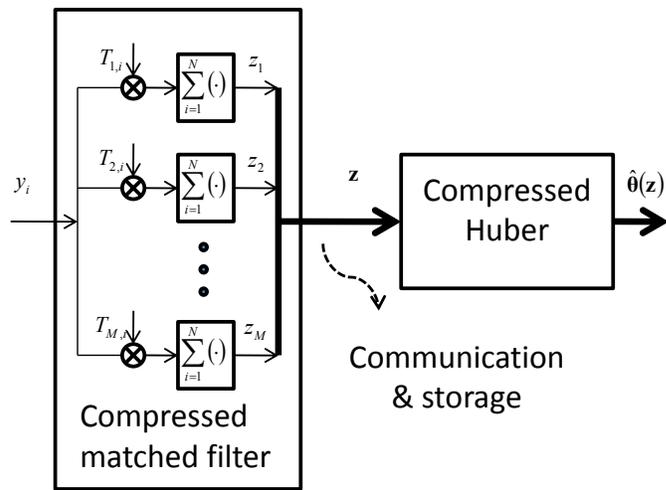


Fig. 1. Block diagram of the CMF and CH.

data, e.g., Huber's technique [10, 11]. This requires all the data to be stored for processing which can be quite expensive in terms of memory. In systems where the data have to be communicated for postprocessing the above is extremely limiting. Instead, the goal of this paper is to propose a compressed matched filter (CMF), namely a bank of approximate (and randomized) matched filters which compress the observations and contain most of the information in the data. Then, the output of the CMF can be easily stored or communicated for future use. In addition, we derive a compressed Huber (CH) estimator which allows to reconstruct the unknown parameter vector using this compressed data. A block diagram of CMF and CH is provided in Fig. 1. It is important to note that recovering the system parameters from the compressed data may still have the same complexity as in the uncompressed case. Therefore our main contribution is in lowering the amount of data transmitted from the receiver to the postprocessing unit.

Our framework is motivated by the theories of sparse recovery and compressed sensing (CS) [1, 8]. Compressed sensing is a technique for reconstructing a high dimensional but sparse unknown vector from a small number of linear measurements. The basic approach is to use linear measurements with randomized coefficients and a recovery algorithm based on convex optimization with an L_1 norm. Sparse impulsive noise (or outliers) can be dealt with using the same framework [13, 17]. A more advanced method known as LASSO extends the setting to noisy measurements [18]. Recently, the problem was generalized to the estimation of a vector which is

This work was supported by the Israeli Smart Grid consortium, and also in part by ISF Grant 786/11. Special gratitude to Amir Globerson

only partially sparse [16]. All of these methods consider sparse signals and simple Gaussian noise. Interestingly, Fuchs (in [7]) showed that Huber's robust regression can be expressed as the solution to a partially sparse model. The result is quite intuitive and can be interpreted as Gaussian noise contaminated by additional sparse and deterministic outliers. Fuchs had also suggested a generalization of the robust regression to the correlated noise case but did not present farther results. Our proposed systems is based on these ideas. CMF complements the classical matched filter with a few additional CS filters. CH estimates both the signal and the outliers by searching for a partially sparse solution that is consistent with the compressed data. Note that our frameworks is different than classical CS in two aspects: First, our sensing procedure is still mostly based on the matched filter. The additional randomized filter assist in detecting and eliminating the outliers. Second, unlike CS, our desired signal is dense. The sparsity is associated with the nuisance outliers.

The paper is organized as follows. We begin in Section 2 by introducing the problem formulation. In Section 3 we consider the inherent performance limitations due to the compression. These bounds are computed assuming a clairvoyant estimator that can only be approximated in practice. In Section 4, we address the choice of CMF using the theory of CS matrix design. In Section 5, we derive the CH algorithm which allows to reconstruct the unknown signal and as byproduct also part of the noise. This optimization is based on the joint Maximum a Posteriori and Maximum Likelihood (JMAP-ML) estimator [20] and ideas from Huber's regression [10]. The input to CH is low dimensional, but it processes high dimensional vectors in its internal computations. Thus, we also provide an efficient implementation of CH based on the fast iterative shrinkage and thresholding algorithm (FISTA) by [2]. Finally, in Section 6 we illustrate the performance of our proposed methods using numerical simulations.

The following notation is used. The sets \mathbb{R}^n and $\mathbb{R}^{n \times m}$ denote the set of length n vectors and the set of size $n \times m$ matrices. The operator $\|\cdot\|_p$ denotes the L_p norm. The superscript \mathbf{X}^T and \mathbf{X}^{-1} denotes the transpose and inverse operations. The subscript x_i denote the i 's element in the vector \mathbf{x} . The Moore Penrose pseudoinverse of a matrix \mathbf{T} is denoted by \mathbf{T}^\dagger . We denote the multivariate Gaussian distribution by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ where $\boldsymbol{\mu}$ and $\boldsymbol{\Sigma}$ are the mean vector and the covariance matrix.

2. PROBLEM FORMULATION

Consider a linear model

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \mathbf{n} \quad (1)$$

where $\mathbf{H} \in \mathbb{R}^{N \times K}$ is a known matrix with $N \gg K$, $\boldsymbol{\theta} \in \mathbb{R}^K$ is an unknown deterministic vector and $\mathbf{n} \in \mathbb{R}^N$ is a random vector with independent elements. In the Gaussian case i.e.

$$n_i \sim \mathcal{N}(0, \sigma_1^2) \quad i = 1, \dots, N \quad (2)$$

It is well known that all the information in \mathbf{y} about $\boldsymbol{\theta}$ can be compressed by a linear matched filter

$$\mathbf{z} = \mathbf{T}\mathbf{y} \quad (3)$$

where $\mathbf{T} = \mathbf{H}^T$. i.e. \mathbf{z} is a sufficient statistic of \mathbf{y} . Remarkably, the dimension of \mathbf{z} is K which is much smaller than N , and hence the compression. This is even true in the extreme continuous case in which N is infinite but the dimension of \mathbf{z} is still K and depends only on the number of unknowns. Using \mathbf{z} we can infer whatever we need about $\boldsymbol{\theta}$ without storing \mathbf{y} .

Our goal is to obtain a similar linear compression in the non-Gaussian case. Specifically, we assume that the marginal distribution of each element in \mathbf{n} is an ϵ -contaminated Gaussian model

$$n_i \sim (1 - \epsilon) \mathcal{N}(0, \sigma_1^2) + \epsilon \mathcal{G} \quad i = 1, \dots, N \quad (4)$$

where $\epsilon > 0$ is a known small contamination ratio parameter, and \mathcal{G} is some symmetric distribution, typically unknown and referred to as outlier distribution. In this case, a low dimensional sufficient statistic is usually unknown. Thus, we seek an approximate compression procedure. We will design a compression matrix \mathbf{T} as in (3) of size $M \times N$ where $K \leq M \ll N$, that summarizes as much information on $\boldsymbol{\theta}$ as possible. Then, given the compressed \mathbf{z} we will derive a computationally efficient algorithm for estimating the unknown parameter $\boldsymbol{\theta}$.

3. PERFORMANCE BOUNDS

It is instructive to begin with two simplified problems which provide inherent performance bounds and explains our methodology. For this purpose, we assume a Gaussian mixture noise distribution meaning that the outliers are distributed normally, i.e. $\mathcal{G} = \mathcal{N}(0, \sigma_2^2)$ where $\sigma_2^2 (\gg \sigma_1^2)$ is a known constant, and consider oracle estimators which somehow know the locations of the outliers in \mathbf{n} . First, we consider the uncompressed case in which $\mathbf{T} = \mathbf{I}$. Under this assumption, the conditional distribution of the observations is

$$\mathbf{z} \sim \mathcal{N}(\mathbf{H}\boldsymbol{\theta}, \mathbf{D}) \quad (5)$$

where \mathbf{D} is a diagonal matrix with the variances of \mathbf{n} . Roughly, $(1 - \epsilon)N$ of its diagonal elements are equal to σ_1^2 and the other ϵN elements are equal to σ_2^2 . This is a simple Gaussian linear model and the optimal estimator is a Weighted Least Squares (WLS) [12]

$$\hat{\boldsymbol{\theta}}_{\text{no compression}} = \arg \min_{\boldsymbol{\theta}} \|\mathbf{y} - \mathbf{H}\boldsymbol{\theta}\|_{\mathbf{D}^{-1}}^2 \quad (6)$$

$$= \left(\mathbf{H}^T \mathbf{D}^{-1} \mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{D}^{-1} \mathbf{z} \quad (7)$$

where we use a weighted norm defined as

$$\|\mathbf{x}\|_{\mathbf{W}}^2 = \mathbf{x}^T \mathbf{W} \mathbf{x}. \quad (8)$$

Its mean squared error is then given by

$$\text{MSE}_{\text{no compression}} = \mathbb{E} \left\{ \left(\mathbf{H}^T \mathbf{D}^{-1} \mathbf{H} \right)^{-1} \right\} \quad (9)$$

where the expectation is with respect to the randomness in \mathbf{D} . This is not a general performance bound, as we have assumed a specific noise distribution but it is quite close if σ_2^2 is indeed the variance of the outliers. Any compression will probably increase the error, and our goal is to get as close as possible to this error with the smallest possible value of M .

In the compressed case (again, with known locations of the outliers and Gaussian outliers), the distribution of the observations is

$$\mathbf{z} \sim \mathcal{N}(\mathbf{T}\mathbf{H}\boldsymbol{\theta}, \mathbf{T}\mathbf{D}\mathbf{T}^T) \quad (10)$$

where we condition on both \mathbf{T} and \mathbf{D} which are statistically independent. This too is a simple Gaussian linear model solved via a WLS

$$\hat{\boldsymbol{\theta}}_{\text{oracle}} = \left(\mathbf{H}^T \mathbf{T}^T \left(\mathbf{T}\mathbf{D}\mathbf{T}^T \right)^{-1} \mathbf{T}\mathbf{H} \right)^{-1} \mathbf{H}^T \mathbf{T}^T \left(\mathbf{T}\mathbf{D}\mathbf{T}^T \right)^{-1} \mathbf{z} \quad (11)$$

Its mean squared error is given by

$$\text{MSE}_{\text{oracle}} = \mathbb{E} \left\{ \left(\mathbf{H}^T \mathbf{T}^T \left(\mathbf{T} \mathbf{D} \mathbf{T}^T \right)^{-1} \mathbf{T} \mathbf{H} \right)^{-1} \right\} \quad (12)$$

where the expectation is with respect to the randomness in both \mathbf{H} and \mathbf{D} . In practice, it is impossible to implement the above oracle. However, it suggests a natural two step approach: first, detect the location of the outliers, then use an approximate oracle assuming these locations are exact. Furthermore, these MSEs are reasonable performance bounds that any practical estimator should be compared to.

4. COMPRESSED MATCHED FILTER FOR NON-GAUSSIAN NOISE

The first part of our design is the choice of the sensing matrix \mathbf{T} which defines CMF. Unfortunately, it is not completely clear what is the optimal criterion for the design and/or how to numerically optimize it. On the one hand, from the signal perspective, we would like \mathbf{T} to be close to the matched filter. At the least we need to ensure that its columns span the columns of \mathbf{H} . On the other hand, we need \mathbf{T} to give some response to the noise shape. Here we can use some insights from the compressed sensing field by looking at a sparse model which is close to ours. Specifically, we can look at a linear system with Gaussian noise and deterministic outliers. Hence, setting $\mathbf{n} = \boldsymbol{\nu} + \mathbf{u}$ where $\boldsymbol{\nu}$ is a random vector with independent $\mathcal{N}(0, \sigma_1^2)$ variables, and \mathbf{u} is a deterministic sparse vector into (1) results in the following model

$$\mathbf{y} = \mathbf{H}\boldsymbol{\theta} + \boldsymbol{\nu} + \mathbf{u} \quad (13)$$

CS theory hints that we can use a random matrix to encode the sparsity of \mathbf{u} . Therefore, to address both criteria we propose the following simple structure:

$$\mathbf{T} = \begin{bmatrix} \mathbf{H}^T \\ \mathbf{W}\mathbf{P} \end{bmatrix} \quad (14)$$

where $\mathbf{W} \in \mathbb{R}^{M-K \times N}$ is a matrix with independent and identically distributed (i.i.d.) $\mathcal{N}(0, 1)$ elements and

$$\mathbf{P} = \mathbf{I} - \mathbf{H} \left(\mathbf{H}^T \mathbf{H} \right)^{-1} \mathbf{H}^T \quad (15)$$

is a projection matrix onto the null space of \mathbf{H} . This choice guarantees that we will always be better or equal to the naive Gaussian matched filter which is exactly the first K rows. The rest of the rows randomly span as much as possible from the remaining space.

5. COMPRESSED HUBER

The second part of our framework is the CH algorithm which estimates $\boldsymbol{\theta}$ given \mathbf{z} for a fixed \mathbf{T} . Ideally, we would like to find the parameter $\boldsymbol{\theta}$ that maximizes the likelihood of \mathbf{z} . But this vector is a high dimensional mixture of many non-Gaussian random variables, and its distribution is hard to analyze. Instead, we propose to estimate both $\boldsymbol{\theta}$ and \mathbf{n} simultaneously. Statistically speaking, we jointly seek for $\boldsymbol{\theta}$ via a maximum likelihood approach and for \mathbf{n} via a maximum a posteriori approach (see [20] for more details on JMAP-ML estimation):

$$\begin{aligned} \min_{\boldsymbol{\theta}, \mathbf{n}} \quad & \sum_{i=1}^N \phi(n_i) \\ \text{s.t.} \quad & \mathbf{z} = \mathbf{T}(\mathbf{H}\boldsymbol{\theta} + \mathbf{n}) \end{aligned} \quad (16)$$

where $\phi(\cdot)$ is the negative-log-posterior distribution of n_i as described in (4). Because the distributions of \mathcal{G} and as a result the distribution of \mathbf{n} are generally unknown, we can not calculate $\phi(\cdot)$ directly. Hence, we have to use some robust objective function which will be indifferent to the specific distribution of the outliers. Such is the Huber's loss function which was proven to be optimal in the uncompressed case (in the minimax sense) [11]. Together, our reconstruction algorithm is the solution to

$$\begin{aligned} \min_{\boldsymbol{\theta}, \mathbf{n}} \quad & \sum_{i=1}^N \rho_h(n_i) \\ \text{s.t.} \quad & \mathbf{z} = \mathbf{T}(\mathbf{H}\boldsymbol{\theta} + \mathbf{n}) \end{aligned} \quad (17)$$

where

$$\rho_h(n) = \begin{cases} n^2 & |n| < h \\ 2h|n| - h^2 & |n| > h \end{cases} \quad (18)$$

and h is calculated from

$$\frac{\sigma}{h} \psi\left(\frac{h}{\sigma}\right) - Q\left(\frac{h}{\sigma}\right) = \frac{\epsilon}{2(1-\epsilon)} \quad (19)$$

with $\psi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ and $Q(t) = \frac{1}{\sqrt{2\pi}} \int_t^\infty e^{-\frac{x^2}{2}} dx$. The above is a convex minimization that can be efficiently solved using off-the-shelf optimization packages, e.g., CVX [9].

Remarkably, [7] showed that Huber's function can be expressed as:

$$\rho_h(n) = \min_u (u - n)^2 + 2h|u| \quad (20)$$

Plugging this expression into (17) yields

$$\begin{aligned} \min_{\boldsymbol{\theta}, \mathbf{n}, \mathbf{u}} \quad & \|\mathbf{n} - \mathbf{u}\|^2 + 2h\|\mathbf{u}\|_1 \\ \text{s.t.} \quad & \mathbf{z} = \mathbf{T}(\mathbf{H}\boldsymbol{\theta} + \mathbf{n}) \end{aligned} \quad (21)$$

Then by solving explicitly for \mathbf{n} we can derive the following equivalent problem

$$\min_{\boldsymbol{\theta}, \mathbf{u}} \|\mathbf{z} - \mathbf{T}[\mathbf{H}\boldsymbol{\theta} + \mathbf{u}]\|_{(\mathbf{T}\mathbf{T}^T)^{-1}}^2 + 2h\|\mathbf{u}\|_1 \quad (22)$$

This formulation provides an interesting observation. The solution of (22) can be interpreted as an estimator to the deterministic sparse outlier model in (13). It can be seen that the first term in (22) is a standard WLS objective whereas the second term penalizes vectors \mathbf{u} which are not sparse.

The above formulation is also useful from a numerical perspective. Note that \mathbf{z} is compressed and low dimensional, but the internal variable \mathbf{u} is of length N and therefore the optimization requires a large scale numerical algorithm. For this purpose, we utilize the well known FISTA solver due to [2]. First, we notice that $\hat{\boldsymbol{\theta}}$ can be solved explicitly

$$\hat{\boldsymbol{\theta}} = \mathbf{Q}(\mathbf{z} - \mathbf{T}\mathbf{u}) \quad (23)$$

where $\mathbf{Q} = (\mathbf{H}^T \mathbf{T}^\dagger \mathbf{T} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{T}^\dagger$. Then by substituting it to (22) we get a classical LASSO problem

$$\min_{\mathbf{u}} \|(\mathbf{I} - \mathbf{T}\mathbf{H}\mathbf{Q})(\mathbf{z} - \mathbf{T}\mathbf{u})\|_{(\mathbf{T}\mathbf{T}^T)^{-1}}^2 + 2h\|\mathbf{u}\|_1 \quad (24)$$

which can be solved efficiently by FISTA. For convenience of notation, we also define the shrinkage and thresholding operator

$$\mathcal{T}_\alpha(x) = \max\{|x| - \alpha, 0\} \text{sign}\{x\} \quad (25)$$

Algorithm 1: FISTA implementation of CH

Input: $\mathbf{T}, \mathbf{H}, \mathbf{z}$
Output: $\hat{\boldsymbol{\theta}}, \hat{\mathbf{u}}$

$$\mathbf{S} = (\mathbf{I} - \mathbf{H}\mathbf{Q})^T (\mathbf{T}\mathbf{T}^T)^{-1} (\mathbf{I} - \mathbf{H}\mathbf{Q})$$

$$L = 2\lambda_{\max}(\mathbf{T}^T \mathbf{S} \mathbf{T})$$

 Step 0: Take $t_1 = 1$ and

$$\mathbf{u}_1 = \tilde{\mathbf{u}} = \mathbf{0}$$

 Step k : ($k \geq 0$) Compute

$$\mathbf{e}_k = \mathbf{z} - \mathbf{T}\tilde{\mathbf{u}}_k$$

$$\mathbf{u}_k = \mathcal{T}_{\frac{2h}{L}}(\tilde{\mathbf{u}}_k - \frac{2}{L}\mathbf{T}^T \mathbf{S} \mathbf{e}_k)$$

$$t_{k+1} = \frac{1 + \sqrt{1 + 4t_k^2}}{2}$$

$$\tilde{\mathbf{u}}_{k+1} = \mathbf{u}_k + \frac{t_k - 1}{t_{k+1}}(\mathbf{u}_k - \mathbf{u}_{k-1})$$

 Return: $\hat{\mathbf{u}} = \mathbf{u}_k$ and $\hat{\boldsymbol{\theta}} = \mathbf{Q}(\mathbf{z} - \mathbf{T}\mathbf{u}_k)$

Summing the above, a pseudocode for solving CH in (22) using FISTA is provided in Algorithm 1.

After computing CH, we propose to fine tune the estimate. By examining the optimal \mathbf{u} we (approximately) detect the locations of the outliers

$$\hat{\mathcal{I}}_{out} = \{i \mid |\hat{u}_i| > \sigma_1\} \quad (26)$$

Then we estimate their variance

$$\hat{\sigma}_2^2 = \frac{1}{|\hat{\mathcal{I}}_{out}|} \sum_{i \in \hat{\mathcal{I}}_{out}} \hat{u}_i^2 \quad (27)$$

recover the diagonal covariance matrix of \mathbf{n} denoted by $\hat{\mathbf{D}}$ and finally compute $\hat{\boldsymbol{\theta}}_{oracle}$ in (11) replacing the true \mathbf{D} with its estimate $\hat{\mathbf{D}}$. We denote this second phase as AWLS for approximate WLS.

6. NUMERICAL RESULTS

To demonstrate the advantages of CMF and CH we present simulation results in a simple signal processing application. We consider the estimation of amplitudes and phases of $K/2$ sinusoids with known frequencies contaminated by a Gaussian mixture noise. Specifically, we define $K = 10$, $N = 500$, $\epsilon = 1\%$ and $\sigma_1^2 = 1$. We express the sinusoids in linear form by defining $\mathbf{H} = [\mathbf{H}^c \ \mathbf{H}^s]$ where $\mathbf{H}_{i,n}^c = \cos(2\pi f_i(n-1))$, $\mathbf{H}_{i,n}^s = \sin(2\pi f_i(n-1))$, $i = 1, \dots, K/2$ and $n = 1, \dots, N$. The frequencies are $f_1 = .1$, $f_2 = .2$, $f_3 = .3$, $f_4 = .35$, and $f_5 = .4$. The true amplitudes are all unit and $\mathcal{G}_i = \mathcal{N}(0, \sigma_2^2)$ with $\sigma_2^2 = 500$. The data is compressed using CMF and the system parameters are estimated using the suggested algorithms.

Fig. 2 presents the estimation mean squared errors averaged over the realizations of the noise \mathbf{n} as a function of compression ratio M/N . For comparison, the errors are bounded below by computing (9) (NO COMPRESSION) and above by computing (12) with $\mathbf{T} = \mathbf{H}^T$ (FULL COMPRESSION). In between is the ORACLE using the proposed CMF with randomized versions of \mathbf{T} . Our proposed estimators are denoted by CH and AWLS. It is easy to see the advantages of CH which closes 90% the performance gap with only a quarter of the complete measurements (i.e. four fold compression). AWLS is even better and achieves the same performance with a higher compression. On the downside, the simulation suggest that there may still be room for improvement. Neither CH nor AWLS succeed in achieving the ORACLE performance that knows the locations of the outliers.

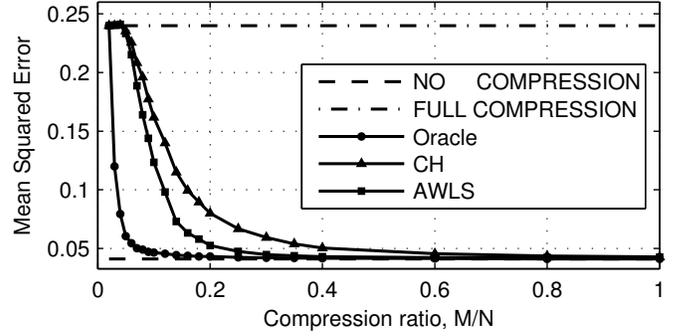


Fig. 2. Estimation mean squared error as a function of compression ratio M/N .

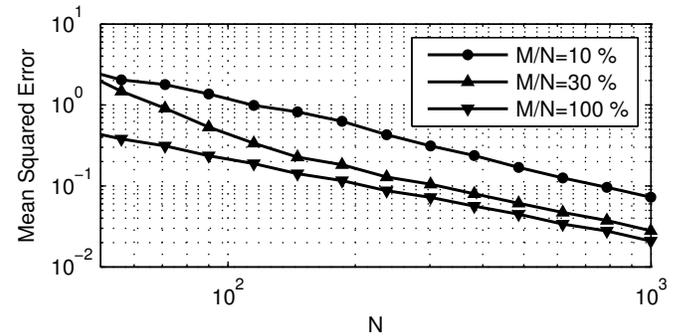


Fig. 3. Estimation (using CH) mean squared error as a function of $N = 50, \dots, 1000$.

Fig. 3 presents the estimation (using CH) mean squared errors averaged over the realizations of the noise \mathbf{n} as a function of N for several compression ratios. As can be seen the estimation error is monotonic and asymptotically tends to zero inversely proportional to N ($MSE \propto N^{-1}$). Thus, suggesting asymptotic consistency of the estimator. It is also worth noting that asymptotically the ratio between estimation MSE's for different compression is constant.

Results similar to Fig. 2 and Fig. 3 were obtained for Laplace distributed outliers with the same variance (i.e. $\mathcal{G}_i = Laplace(0, \frac{\sigma_2}{\sqrt{2}})$) and are not shown here due to space constraints.

7. CONCLUSIONS AND FUTURE WORK

In this paper we have presented a simple compression scheme for a linear system without major information loss. We have also developed a fast recovery scheme for the compressed data. Combination of the two methods were shown, by simulations, to recover the system parameters using approximately four fold compression with no significant loss in MSE. Additional research is needed to optimize the compression matrix and finding more efficient recovery algorithms or providing a tighter lower bound for them.

8. REFERENCES

- [1] R. G. Baraniuk. Compressive sensing [lecture notes]. *Signal Processing Magazine, IEEE*, 24(4):118–121, 2007.

- [2] A. Beck and M. Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.
- [3] K. L. Blackard, T. S. Rappaport, and C. W. Bostian. Measurements and models of radio frequency impulsive noise for indoor wireless communications. *IEEE Journal on Selected Areas in Communications*, 11(7):991–1001, 1993.
- [4] R. S. Blum, R. J. Koziak, and B. M. Sadler. An adaptive spatial diversity receiver for non-Gaussian interference and noise. *IEEE Transactions on Signal Processing*, 47(8):2100–2111, 1999.
- [5] P. L. Brockett, M. Hinich, and G. R. Wilson. Nonlinear and non-Gaussian ocean noise. *The Journal of the Acoustical Society of America*, 82:1386, 1987.
- [6] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 222:309–368, 1922.
- [7] J. J. Fuchs. An inverse problem approach to robust regression. In *Proceedings., 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, 1999.*, volume 4, pages 1809–1812. IEEE, 1999.
- [8] G.B. Giannakis, G. Mateos, S. Farahmand, V. Kekatos, and H. Zhu. USPACOR: Universal sparsity-controlling outlier rejection. pages 1952–1955, 2011.
- [9] M. Grant, S. Boyd, and Y. Ye. CVX: Matlab software for disciplined convex programming, 2008.
- [10] P. J. Huber. Robust estimation of a location parameter. *The Annals of Mathematical Statistics*, 35(1):73–101, 1964.
- [11] P. J. Huber. *Robust statistics*. Springer, 2011.
- [12] S. Kay. Fundamentals of statistical processing, vol. i: Estimation theory. *America: Prentice Hall PTR*, 1993.
- [13] J.N. Laska, M.A. Davenport, and R.G. Baraniuk. Exact signal recovery from sparsely corrupted measurements through the pursuit of justice. pages 1556–1560, 2009.
- [14] J. Lindenlaub and K. Chen. Performance of matched filter receivers in non-Gaussian noise environments. *Communication Technology, IEEE Transactions on*, 13(4):545–547, 1965.
- [15] D. Middleton. Man-made noise in urban environments and transportation systems: Models and measurements. *Communications, IEEE Transactions on*, 21(11):1232–1241, 1973.
- [16] T. Rautenberg, Y. C. Eldar, and L. Tong. Maximum likelihood estimation under partial sparsity constraints. In *Proceedings., 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2013*. IEEE, 2013.
- [17] C. Studer, P. Kuppinger, G. Pope, and H. Bolcskei. Sparse signal recovery from sparsely corrupted measurements. In *Information Theory Proceedings (ISIT), 2011 IEEE International Symposium on*, pages 1422–1426, 2011.
- [18] R. Tibshirani. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 267–288, 1996.
- [19] X. Wang and H. V. Poor. Robust multiuser detection in non-Gaussian channels. *IEEE Transactions on Signal Processing*, 47(2):289–305, 1999.
- [20] A. Yeredor. The joint MAP-ML criterion and its relation to ML and to extended least-squares. *IEEE Transactions on Signal Processing*, 48(12):3484–3492, 2000.