Compressed Sensing with Linear Correlation Between Signal and Measurement Noise

Arildsen, Thomas; Larsen, Torben

Compressed Sensing with Linear Correlation Between Signal and Measurement Noise

Thomas Arildsen[a,*] and Torben Larsen[b]

Aalborg University, Faculty of Engineering and Science
Department of Electronic Systems
Postal address:
Niels Jernes Vej 12, DK-9220 Aalborg, Denmark

—

[a] *e-mail: tha@es.aau.dk*   [b] *e-mail: tl@es.aau.dk*
phone: +45 99409844
ORCID: 0000-0003-3254-3790
*(Corresponding author, EURASIP member)

Abstract

Existing convex relaxation-based approaches to reconstruction in compressed sensing assume that noise in the measurements is independent of the signal of interest. We consider the case of noise being linearly correlated with the signal and introduce a simple technique for improving compressed sensing reconstruction from such measurements. The technique is based on a linear model of the correlation of additive noise with the signal. The modification of the reconstruction algorithm based on this model is very simple and has negligible additional computational cost compared to standard reconstruction algorithms, but is not known in existing literature. The proposed technique reduces reconstruction error considerably in the case of linearly correlated measurements and noise. Numerical experiments confirm the efficacy of the technique. The technique is demonstrated with application to low-rate quantization of compressed measurements, which is known to intro-

duce correlated noise, and improvements in reconstruction error compared to ordinary Basis Pursuit De-Noising of up to approximately 7 dB are observed for 1 bit/sample quantization. Furthermore, the proposed method is compared to Binary Iterative Hard Thresholding which it is demonstrated to outperform in terms of reconstruction error for sparse signals with a number of non-zero coefficients greater than approximately ⅟₁₀th of the number of compressed measurements.

1. Introduction

In the recently emerged field of compressed sensing, one considers linear measurements $\mathbf{y}$ of a sparse vector $\mathbf{x}$, possibly affected by noise as:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{n}, \tag{1}$$

where the measurements $\mathbf{y} \in \mathbb{R}^{M \times 1}$, the sparse vector $\mathbf{x} \in \mathbb{R}^{N \times 1}$, the additive noise $\mathbf{n} \in \mathbb{R}^{M \times 1}$, the system matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$, and $M \ll N$ [1, 2, 3]. $\mathbf{A}$ is generally the product of a measurement matrix and a dictionary matrix: $\mathbf{A} = \mathbf{\Phi}\mathbf{\Psi}$, where $\mathbf{\Phi} \in \mathbb{C}^{M \times N}$, $\mathbf{\Psi} \in \mathbb{C}^{N \times N}$. For simplicity, we assume that $\mathbf{\Psi}$ is an orthonormal basis although more general dictionaries are indeed possible [4].

The essence of compressed sensing, as Donoho, Candès, Romberg, and Tao show in [1, 2], is that the under-determined equation system (1) can be solved provided that:

1. The vector $\mathbf{x}$ is sparse; i.e., only few (K) elements in $\mathbf{x}$ are non-zero.

$$K = |\{x_i | x_i \neq 0, i = 1, \ldots, N\}| \tag{2}$$

   $\mathbf{x}$ can also be approximated sparsely if it is compressible [3, Sec. 3.3], meaning that its coefficients sorted by magnitude decay rapidly to zero.

2. The system matrix $\mathbf{A}$ obeys the Restricted Isometry Property (RIP) with isometry constant $\delta_K > 0$, defined as follows:

$$(1 - \delta_K) \|\mathbf{x}\|_{\ell_2}^2 \leq \|\mathbf{A}\mathbf{x}\|_{\ell_2}^2 \leq (1 + \delta_K) \|\mathbf{x}\|_{\ell_2}^2, \tag{3}$$

   for any at most K-sparse vector $\mathbf{x}$ such that [5]:

$$\delta_K + \delta_{2K} + \delta_{3K} < 1. \tag{4}$$

   This holds with high probability when $\mathbf{\Phi}$ is generated with zero-mean independent identically distributed (i.i.d.) Gaussian entries with variance $\frac{1}{M}$. Note that (3) and (4) are sufficient but not necessary conditions, and rather conservative conditions indeed, as shown in [6]. Conditions (3) and (4) lead to the following sufficient amount of measurements M for Gaussian measurement matrices $\mathbf{\Phi}$ [7]:

$$M \geq CK \log\left(\frac{N}{M}\right), \tag{5}$$

   where C is a fairly small constant which can be calculated as a function of $\frac{M}{N}$ [5].

Given the measurements $\mathbf{y}$, the unknown sparse vector $\mathbf{x}$ can be reconstructed by solving the following convex optimization problem [3, Sec. 4]:

$$\hat{\mathbf{x}} = \underset{\mathbf{u}:\ \|\mathbf{y}-\mathbf{A}\mathbf{u}\|_2 \leq \epsilon}{\operatorname{argmin}} \|\mathbf{u}\|_1, \tag{6}$$

where the fidelity constraint $\|\mathbf{y} - \mathbf{Au}\|_2 \leq \epsilon$ ensures consistency with the observed measurements to within some margin of error, ϵ, which is chosen sufficiently large to accommodate the error $\mathbf{n}$ and/or approximation error in the case of compressible signals. The form of the optimization problem in (6) is known as Least Absolute Shrinkage and Selection Operator (LASSO) [8] or Basis Pursuit De-Noising (BPDN) [9] and also comes in other variants such as the Dantzig selector [10]. In addition to the convex optimization approach to reconstruction in compressed sensing, there exist several iterative/greedy algorithms such as Iterative Hard Thresholding (IHT) [11], or Subspace Pursuit (SP) [12] and Compressive Sampling Matching Pursuit (CoSaMP) [13] as well as the more generalized incarnation of the two latter, Two-Stage Thresholding (TST) [14]. We generally refer to such convex or greedy approaches as reconstruction algorithms. The reconstruction algorithms generally assume the noise to be white and independent of the measurements before noise $\bar{\mathbf{y}} = \mathbf{Ax}$. In particular, to the best of the authors' knowledge, the case of measurement noise being linearly correlated with the measurements has not been treated in the existing literature. Such correlation arises in for example the case of low-resolution quantization. As we demonstrate in Section 2, this case poses a problem for the accuracy of the found solution $\hat{\mathbf{x}}$. More special cases of correlated noise arising from Poisson measurements or quantisation of measurements has, however, been treated in for example [15, 16, 17].

In this paper, we propose a simple yet efficient approach to alleviating the problem of linear correlation between the measurements before noise $\bar{\mathbf{y}}$ and the noise $\mathbf{n}$. Our proposal boils down to a simple scaling of the solution $\hat{\mathbf{x}}$. Through numerical experiments we demonstrate how linearly

4

correlated measurements and noise adversely affect the reconstruction error
and demonstrate how our proposal improves the estimates considerably.

As an application example, we demonstrate the proposed approach in
the case of low-rate scalar quantization of the measurements $\bar{\mathbf{y}}$ which can be
observed to introduce the mentioned linearly correlated measurement noise.
We demonstrate how a well-known linear model used for modeling such cor-
relation in scalar quantization is equivalent to the model of correlated mea-
surement noise considered in this work.

The article is structured as follows: Section 2 introduces the considered
model of linear correlation between compressed measurements and noise and
proposes a solution to enhance reconstruction under these conditions, Sec-
tion 3 describes simulations conducted to evaluate the performance of the
proposed approach compared to a traditional approach, Section 4 presents
the results of these numerical simulations, Section 5 provides discussions of
some of the presented results, and Section 6 concludes the article.

2. Methodology

2.1. Correlated Measurements and Noise

We consider additive measurement noise $\mathbf{n}$ which is correlated with the
measurements before noise $\bar{\mathbf{y}}$. We model the correlation by the linear model:

$$\mathbf{y} = \alpha \mathbf{A}\mathbf{x} + \mathbf{w}, \tag{7}$$

where $\mathbf{w}$ is assumed an additive white noise uncorrelated with $\mathbf{x}$ and $0 <$
$\alpha \leq 1$ where $\alpha = 1$ covers the ordinary case of uncorrelated measurement
noise. $\mathbf{A}$ is the product of a measurement matrix $\boldsymbol{\Phi}$ with i.i.d. Gaussian

entries $\sim \mathcal{N}\left(0, \frac{1}{M}\right)$ and an orthonormal dictionary matrix $\mathbf{\Psi}$. The model (7) results in the following additive noise term:

$$\mathbf{n} = \mathbf{y} - \bar{\mathbf{y}} = \alpha \mathbf{A}\mathbf{x} + \mathbf{w} - \mathbf{A}\mathbf{x} = (\alpha - 1)\mathbf{A}\mathbf{x} + \mathbf{w} \tag{8}$$

We define $\bar{\mathbf{y}} = \mathbf{A}\mathbf{x}$ to signify the measurements before introduction of additive noise. It is readily seen from (8) that $\mathbf{n}$ is correlated with $\mathbf{x}$. The noise variance is

$$\sigma_n^2 = \frac{1}{M} \mathbb{E}\left[\mathbf{n}^\mathsf{T}\mathbf{n}\right] = \frac{1}{M}\left((\alpha - 1)^2 \mathbb{E}\left[\bar{\mathbf{y}}^\mathsf{T}\bar{\mathbf{y}}\right] + \mathbb{E}\left[\mathbf{w}^\mathsf{T}\mathbf{w}\right]\right), \tag{9}$$

which can be calculated by assuming that $\sigma_{\bar{y}}^2 = \frac{1}{M} \mathbb{E}\left[\bar{\mathbf{y}}^\mathsf{T}\bar{\mathbf{y}}\right]$ and $\sigma_w^2 = \frac{1}{M} \mathbb{E}\left[\mathbf{w}^\mathsf{T}\mathbf{w}\right]$ are known or can be estimated. For example, we show an example for σ_w^2 in the case of quantization in Section 2.5, (21).

The specific problem caused by correlated measurements and noise as modeled by (7) is that the noise itself is partly sparse in the same dictionary as the signal of interest, $\mathbf{x}$. Intuitively, this causes a solution $\hat{\mathbf{x}}$ as given by, e.g., (6) to adapt to part of the noise *as well as* the signal of interest, unless steps are taken to mitigate this effect.

2.2. Proposed Approach

Using the model in (7), we propose the following reconstruction of the sparse vector $\mathbf{x}$ instead of the standard approach in (6). Equation (7) motivates replacing the system matrix $\mathbf{A}$ by its scaled version $\alpha \mathbf{A}$. We exemplify this approach by applying it in the BPDN reconstruction formulation as below. Replacing $\mathbf{A}$ by $\alpha \mathbf{A}$ in the standard approach (6), we arrive at

$$\hat{\mathbf{x}}_1 = \underset{\mathbf{u}:\ \|\mathbf{y}-\alpha\mathbf{A}\mathbf{u}\|_2 \leq \epsilon}{\operatorname{argmin}} \|\mathbf{u}\|_1. \tag{10}$$

Since ϵ should be chosen to accommodate the level of noise in the measurements $\mathbf{y}$, we can see that, one choice could be to set

$$\epsilon = \|\mathbf{n}\|_2 \tag{11}$$

$$\epsilon = \|\mathbf{w}\|_2 \tag{12}$$

in (6) or (10), respectively. Since the noise terms $\mathbf{n}$ and $\mathbf{w}$ are assumed unknown, (11) and (12) are not realistic choices of ϵ. The optimal choice of ϵ is dependent on the true solution $\mathbf{x}$, and is therefore difficult to obtain in practice as exemplified for more general inverse problems in, e.g., [18]. For this reason, various rules of thumb exist for the selection of ϵ. One such choice is found in [19, Sec. 5.3]:

$$\epsilon = \sqrt{M + 2\sqrt{2M}}\sigma, \tag{13}$$

where σ is the noise level (standard deviation) of the stochastic error $\mathbf{n}$ or $\mathbf{w}$ in (1) or (7), respectively.

2.3. Additional Insight on the Proposed Approach

As outlined in Section 2.2, the model of the correlation between $\mathbf{n}$ and $\bar{\mathbf{y}}$ suggests scaling $\mathbf{A}$ in the constraint of (10). In fact, as we show here, an equivalent solution can be obtained simply by scaling the solution found by the optimization formulation (6).

Proposition 1. *The following optimization formulation is equivalent to the formulation (10) in the sense that they produce solutions of comparable precision:*

$$\hat{\mathbf{x}}_2 = \frac{1}{\alpha} \operatorname*{argmin}_{\mathbf{v}:\ \|\mathbf{y}-\mathbf{A}\mathbf{v}\|_2 \leq \epsilon} \|\mathbf{v}\|_1. \tag{14}$$

7

To see why (14) is equivalent to (10), consider the optimization problem over the variable $\mathbf{v}$, in which we introduce a change of variable $\mathbf{v} \curvearrowright \mathbf{u}$:

$$
\begin{aligned}
\hat{\mathbf{x}}_2 \in \hat{X}_2 &= \underset{\mathbf{v}:\ \|\mathbf{y}-\mathbf{A}\mathbf{v}\|_2 \leq \epsilon}{\operatorname{Argmin}} \left\| \frac{1}{\alpha}\mathbf{v} \right\|_1 \\
&= \underset{\mathbf{u}:\ \|\mathbf{y}-\alpha\mathbf{A}\mathbf{u}\|_2 \leq \epsilon}{\operatorname{Argmin}} \|\mathbf{u}\|_1, \ \mathbf{v} = \alpha\mathbf{u} \\
&= \hat{X}_1 \ni \hat{\mathbf{x}}_1
\end{aligned}
\tag{15}
$$

In (15) we use the notation $\hat{X} = \operatorname{Argmin}\ldots$ to denote the set of solutions to the stated optimization problem since this is generally not one unique solution [20, Ch. 5]. $\hat{\mathbf{x}} \in \hat{X}$ is used to emphasize that $\hat{\mathbf{x}}$ is any feasible minimizer of the problem. It can generally not be guaranteed that algorithms used to obtain solutions to the two optimization problems (10) and (15) return the same solution, but they are subject to the same guarantees of reconstruction accuracy (stability) as given by [20, Theorem 5.3].

According to the above, down-scaling the solution to the optimization in (14) by α results in a solution $\hat{\mathbf{x}}_2$ of comparable accuracy to the solution $\hat{\mathbf{x}}_1$ to (10). Please note that all constraints in (10), (14) and (15) use the same value of ϵ given by (13) with $\sigma = \sigma_w$, the standard deviation of the entries in $\mathbf{w}$ in (7).

In short, Proposition 1 says that for compressed measurements with noise correlated with the measurements according to the model (7), given the correlation parameter α, when the signal $\mathbf{x}$ is reconstructed using BPDN, (6), the obtained solution should be scaled by the factor $\frac{1}{\alpha}$ to account for the effect of the correlation.

2.4. Optimality of the Proposed Approach

In relation to the method proposed in Sections 2.2 and 2.3, it is of course interesting to investigate whether the corrective scaling by α in the reconstruction of $\mathbf{x}$ is indeed optimal. To investigate this, consider the following optimization formulation:

$$\hat{\mathbf{x}}_\beta = \frac{1}{\beta} \underset{\mathbf{u}:\ \|\mathbf{y}-\mathbf{A}\mathbf{u}\|_2 \leq \epsilon}{\operatorname{argmin}} \|\mathbf{u}\|_1, \tag{16}$$

where ϵ is given by (13) and the optimization problem is evaluated for a number of values of $\beta \in [\alpha - \beta_1, \alpha + \beta_2]$ for a given value of α used in the correlated noise model (7) and a suitable choice of β_1 and β_2. The numerical results of this investigation can be found in Section 4.3. $\beta = \alpha$ intuitively seems a suitable choice, but numerical experiments indicate that it is in fact not optimal. An explanation of this observation is offered in Section 5.

2.5. An Application: Quantization

As a practical example where the introduced measurement noise is correlated with the measurements, we investigate low-rate scalar quantization of the individual compressed measurements in $\mathbf{y}$. Quantization is usually modeled by an additive noise model [21]:

$$y = Q(\bar{y}) = \bar{y} + q, \tag{17}$$

where $\bar{y}$ is the original value before quantization, which we consider as $\bar{y} \in \mathbb{R}$. $Q(\cdot)$ is the (non-linear) operation of scalar quantization, mapping $\bar{y}$ to an index i representing a quantized value y

$$Q:\ \bar{y} \to y_i,\ \text{if } \bar{y} \in R_i, i \in \{1, \ldots, L\}, \tag{18}$$

where the range of input values is partitioned into L regions R_i, $i \in \{1, \ldots, L\}$ and any value $\bar{y} \in R_i$ is quantized to the point $y_i \in R_i$. For input $\bar{y}$ with unbounded support, the regions R_i can be defined as follows:

$$R_i = \begin{cases} (p_{i-1}, p_i], & \text{for } i = 1, \ldots, L-1 \\ (p_{i-1}, p_i), & \text{for } i = L, \end{cases} \tag{19}$$

where $p_0 = -\infty \wedge p_L = \infty$. The additive noise $q = y - \bar{y}$ represents the error introduced by quantizing $\bar{y}$ to the value y.

Various modeling assumptions are typically made about q. One type of quantizers has centroid codebooks, i.e. quantizers where the reconstruction points y_i are calculated as the respective centroids of the distribution of the input y in each of the regions R_i, e.g., Lloyd-Max quantizers [22, 23]. For quantizers with centroid codebooks, q is correlated with the input x. A model of this correlation used in the literature is the so-called gain-plus-additive-noise model [24, Sec. II]:

$$y = Q(\bar{y}) = \alpha \bar{y} + r, \tag{20}$$

where $\alpha \in [0, 1]$ and r is an additive noise, assumed uncorrelated with $\bar{y}$. The variance of r is

$$\sigma_r^2 = \alpha(1 - \alpha)\sigma_{\bar{y}}^2. \tag{21}$$

The variance of q is

$$\sigma_q^2 = (1 - \alpha)\sigma_{\bar{y}}^2, \tag{22}$$

which is easily seen by inserting (21) in $\sigma_q^2 = (\alpha - 1)^2 \sigma_{\bar{y}} + \sigma_r^2$.

The parameter α can be computed for a specific quantizer. One way to do this is to estimate it numerically by Monte-Carlo simulation. From [24,

Eq. (8)] we have

$$\alpha = 1 - \frac{\sigma_q^2}{\sigma_{\bar{y}}^2}. \tag{23}$$

The procedure is to generate a random test sequence $\bar{y}$, quantize it with the given quantizer Q designed[1] for the probability density function (p.d.f.) of $\bar{y}$, estimate the variances σ_q^2 and $\sigma_{\bar{y}}^2$ from the realizations of $\bar{y}$ and $q = \bar{y} - y$, and use these to calculate (23).

The model (20) of the quantizer corresponds to the proposed model of correlated measurements and noise described by (7), where $r = w$. Please note that the model, (20), considers scalar quantization. In the case of quantization of a vector $\mathbf{v}$, we use $Q(\mathbf{v})$ to signify scalar quantization of the individual elements of the vector $\mathbf{v}$.

We consider quantization of compressed measurements $\mathbf{y}$ of the signal $\mathbf{x}$:

$$\mathbf{y} = Q(\mathbf{Ax}) \tag{24}$$

$$= \mathbf{Ax} + \mathbf{q} \tag{25}$$

$$\approx \alpha\mathbf{Ax} + \mathbf{w}, \tag{26}$$

where

$$\mathbb{E}\left[\mathbf{q}\mathbf{q}^{\mathrm{T}}\right] = \sigma_q^2\mathbf{I}, \qquad\qquad \mathbb{E}\left[\bar{\mathbf{y}}\bar{\mathbf{y}}^{\mathrm{T}}\right] = \sigma_{\bar{y}}^2\mathbf{I},$$

and $\mathbf{I}$ is the $M \times M$ identity matrix.

Approximating the quantization operation by the noise model in (26), we propose using the reconstruction technique (14) to improve reconstruction

---

[1]The quantizer can for example be trained on test data representing $\bar{y}$ or calculated based on the known or assumed probability density function (p.d.f.) of $\bar{y}$.

with scalar quantized compressed measurements, (24), as an example of noise correlated with the measurements.

Noise variance estimates given by (21) and (22) can be obtained from a known $\sigma_{\bar{y}}^2$. In hardware implementations, $\sigma_{\bar{y}}^2$ can be considered known through the use of automatic gain control prior to quantization or by other means of estimating signal variance prior to quantization.

3. Simulation Framework

In this section we present the numerical simulation set-up used to evaluate the reconstruction method proposed in (14).

Donoho & Tanner have shown in [6] that compressed sensing problems can be divided into two "phases" according to their probability of correct recovery by the method (6). When evaluating the probability of correct reconstruction of a sparse vector $\mathbf{x}$ over the parameter space defined by $\delta = \frac{M}{N} \in [0, 1]$ and $\rho = \frac{K}{M} \in [0, 1]$, a given problem can be proven to fall into one of two phases where the probability of correct reconstruction is close to 1 (feasible) and 0 (infeasible), respectively. These two phases are divided by a sharp phase transition around the correct reconstruction probability of 50% as drawn in Fig. 1 (—). The feasible phase lies below the transition and the infeasible phase lies above. Compressed sensing is utilized most efficiently when operating close to the phase transition in the feasible phase since $\mathbf{x}$ can be reconstructed with the highest possible number of non-zero elements K, given N and M, here. This phase transition occurs in the case of noiseless measurements, in the limit of $N \to \infty$. The theory still holds for finite N, but the phase transition is shifted downwards with respect to ρ in the (δ, ρ)-

12

parameter space, see Fig. 1 (- - -). It has also been shown that a similar transition occurs at the same location in the noisy case, i.e. (1) [25]. In the noisy case, mean squared reconstruction error, $\mathbb{E}\left[\|\hat{\mathbf{x}} - \mathbf{x}\|_2^2/N\right]$ relative to the measurement noise variance σ_n^2 is bounded in the feasible region and unbounded in the infeasible region.

[Figure 1 about here.]

In all simulations, we apply the proposed approach to test signals generated randomly according to the following specifications: size of $\mathbf{x}$ vector $N = 1000$; number of compressed measurements $M \in \{200, 300, 400, 500, 600, 700, 800, 900, 1000\}$. The non-zero elements of $\mathbf{x}$ are i.i.d. $\sim \mathcal{N}(0, 1)$; the number of non-zero elements K is selected for each value of M. This is done by calculating the largest possible K for each M according to the lower bound on the 99% phase transition for finite $N = 1000$ by the formula given in [6, Sec. IV, Theorem 2], drawn in Fig. 1 (- - -). The resulting values are $K \in \{1, 17, 41, 73, 115, 167, 235, 330, 542\}$. The corresponding (δ, ρ)-points are plotted in Fig. 1 ($\times$).

The measurement matrix $\mathbf{\Phi}$ has i.i.d. entries $\sim \mathcal{N}(0, \frac{1}{M})$ and we use the dictionary $\mathbf{\Psi} = \mathbf{I}$, so that $\mathbf{A} = \mathbf{\Phi}$. We repeat the experiment $T = 1000$ times for randomly generated $\mathbf{x}$ and $\mathbf{\Phi}$ in each repetition and average the reconstructed signal Normalized Mean Squared Error (NMSE), $\mathcal{P}$, over all solution instances $\hat{\mathbf{x}}_i, \ i \in \{1, \ldots, T\}$:

$$\mathcal{P} = \frac{1}{T} \sum_{i=1}^{T} \frac{\|\hat{\mathbf{x}}_i - \mathbf{x}_i\|_2^2}{\|\mathbf{x}_i\|_2^2}. \tag{27}$$

To enable assessment of the quality of the obtained results, we plot the simulated figures with error bars signifying their 99% confidence intervals com-

puted under the assumption of a Gaussian distributed mean of the NMSE, see e.g. [26, Sec. 7.3.1]. The simulations were conducted for reconstruction using regular BPDN (6) vs. our proposed approach (14) (denoted "BPDN-scale" in result plots). The numerical optimization problems were solved using the SPGL1[2] software package [27].

Regarding the choice of ϵ, for regular BPDN (6), we chose ϵ according to (13), with $\sigma = \sqrt{\sigma_q^2}$ from (22). For our proposed approach (14), we chose ϵ according to (13), with $\sigma = \sqrt{\sigma_r^2}$ from (21). For both compared approaches, we consider $\sigma_{\tilde{y}}^2$ known. As demonstrated in Section 4.3, ϵ could be chosen better from empirical observations to provide smaller error in the reconstruction, i.e. $\|\hat{\mathbf{x}} - \mathbf{x}\|$. We chose the values (13) as practically useful values for fairness of the evaluation of our proposed method.

As we have chosen low-rate scalar quantization to demonstrate the proposed approach to noise correlated with the measurements, we additionally performed simulations to compare the proposed method to a state-of-the-art reconstruction algorithm for 1-bit compressed sensing, Binary Iterative Hard Thresholding (BIHT) [17]. This simulation was performed by evaluating both our proposed method and BIHT over the phase space $\delta, \rho \in [0, 1]$ where we discretized the range [0,1] in steps of 0.01 for both δ and ρ. In each point (δ, ρ) we evaluated $\mathcal{P}$ according to (27) over $T = 1000$ repetitions with different $\mathbf{x}$ and $\mathbf{A}$ in each instance. For each value $\delta \in \{0.01, 0.02, \ldots, 1\}$ we evaluate each of the methods from $\rho = 0.01$ until $\mathcal{P} > 1$. For BIHT, we generated sparse signals $\mathbf{x}$ normalized to $\|\mathbf{x}\|_2 = 1$ which is assumed by BIHT

---

[2]SPGL1: A solver for large-scale sparse reconstruction (http://www.cs.ubc.ca/labs/scl/spgl1).

14

and other 1-bit compressed sensing reconstruction algorithms in general. In BIHT, estimates $\hat{\mathbf{x}}$ are re-normalised after reconstruction which is not the case in our proposed method.

All scripts required to reproduce the simulation results are openly accessible[3].

4. Numerical Simulation Results

In this section we present results of the numerical simulations conducted according to Section 3. Firstly, we evaluate the proposed method under artificial correlated measurement noise generated according to (7). Secondly, we evaluate the method under correlated measurement noise incurred by scalar quantization of the compressed measurements. These results are shown in Section 4.1. Furthermore, in Section 4.2 we present results of simulations comparing the proposed method to BIHT. Finally, in Section 4.3 we present results of simulations to shed light on how the choices of the parameters β and ϵ in (16) affect the main results.

4.1. Main Results

In this section, noise variance and correlation parameters are first set equal to the corresponding parameters estimated for the Lloyd-Max quantizer used later in this section, for comparability. The parameter values for α are listed in Table 1.

[Table 1 about here.]

---

[3]http://github.com/ThomasA/cs-correlated-noise

The listed values of α (Lloyd-Max) are used together with $\sigma_{\mathbf{r}}^2$ calculated from (21) to generate correlated measurement noise according to (7). In the conducted simulations, BPDN is used to reconstruct $\hat{\mathbf{x}}_2$ from the compressed measurements $\mathbf{y}$. We compare the standard (correlation-unaware) reconstruction, (6), of the signal (denoted "BPDN" in Fig. 2) to the reconstruction obtained by our proposed method, (14), of scaling the reconstructed signal to account for correlation (denoted "BPDN-scale" in Fig. 2). Selected results for equivalent quantizer resolutions 1 bit/sample, 3 bit/sample, and 5 bit/sample are shown in Fig. 2. The proposed method is observed to improve the reconstruction error $\mathcal{P}$ by 7.3 dB to 1.3 dB (for increasing ρ) at 1 bit/sample, 3.1 dB to 0.26 dB (for increasing ρ) at 3 bit/sample, and 0.86 dB to 0.059 dB (for increasing ρ) at 5 bit/sample.

[Figure 2 about here.]

The experiments for quantized measurements are conducted exactly as above, with the exception that the measurements $\mathbf{y}$ are quantized using a *Lloyd-Max quantizer* [23, 22]. The Lloyd-Max quantizer is designed for the Gaussian distribution of the entries of $\bar{\mathbf{y}}$ which results from the use of a measurement matrix containing i.i.d. zero-mean Gaussian entries. The correlated noise model uses the values of α (Lloyd-Max) for the selected quantizer resolutions listed in Table 1.

Selected results for quantizer resolutions 1 bit/sample, 3 bit/sample, and 5 bit/sample with Lloyd-Max quantization are shown in Fig. 3. It can be observed that the reconstruction error figures $\mathcal{P}$ agree well with those simulated with artificially generated correlated noise in Fig. 2. The observed

improvements by the proposed method are almost identical to those observed for artificial noise: $7.6\,\mathrm{dB}$ to $1.3\,\mathrm{dB}$ at $1\,\mathrm{bit/sample}$, $3.1\,\mathrm{dB}$ to $0.26\,\mathrm{dB}$ at $3\,\mathrm{bit/sample}$, and $0.80\,\mathrm{dB}$ to $0.028\,\mathrm{dB}$ at $5\,\mathrm{bit/sample}$.

[Figure 3 about here.]

To evaluate our proposed approach for a more practical quantization scheme than the non-uniform Lloyd-Max quantizer, we additionally simulated results where the measurements $\mathbf{y}$ are quantized using a uniform quantizer with mid-point quantization points, optimized for minimum mean squared error (MMSE) of the quantized measurements. The uniform quantizer is designed for the Gaussian distribution of the entries of $\bar{\mathbf{y}}$. This serves to evaluate how well the proposed approach performs for a more practical quantizer type that does not theoretically obey the quantization noise model (20) due to the fact that its reconstruction points are not the centroids of the input signal's p.d.f. in the quatizer's input regions. The correlated noise model uses the values of α (uniform) from Table 1.

Selected results for quantizer resolutions $1\,\mathrm{bit/sample}$, $3\,\mathrm{bit/sample}$, and $5\,\mathrm{bit/sample}$ with uniform quantization are shown in Fig. 4. The observed improvements by the proposed method are close to those observed for artificial noise: $7.6\,\mathrm{dB}$ to $1.3\,\mathrm{dB}$ at $1\,\mathrm{bit/sample}$, $3.2\,\mathrm{dB}$ to $0.28\,\mathrm{dB}$ at $3\,\mathrm{bit/sample}$, and $0.89\,\mathrm{dB}$ to $0.073\,\mathrm{dB}$ at $5\,\mathrm{bit/sample}$.

[Figure 4 about here.]

The results in Fig. 3a and 4a are identical due to the fact that the 2-level Lloyd-Max quantizer is a uniform 2-level quantizer optimized for MMSE of the quantized values. It can also be observed that the uniform quantizer for

17

3 bit/sample and 5 bit/sample results in slightly larger reconstruction error while the improvement by our proposed method is preserved.

4.2. Comparison to Binary Iterative Hard Thresholding (BIHT)

In this section, we provide results comparing our proposed method to BIHT. Results for our proposed method were computed in the same manner as for the results regarding 1-bit quantization in Section 4.1. The simulated NMSE of our proposed method and BIHT are shown in Fig. 5. The white regions of the phase space are un-tested as they lie beyond $\mathcal{P} > 1$; a threshold we selected to define the region we wished to investigate. The bold contour lines mark the boundary where the NMSEs of our proposed method and BIHT are equal. As the numbered contour lines show, "BPDN-scale" exhibits lower NMSE than BIHT in the majority (upper left region) of the phase space, whereas the NMSE of BIHT is lower along the bottom of the phase space – up to around $\rho = 0.1$ – and in the upper right-hand corner – towards $(\delta, \rho) = (1, 1)$.

[Figure 5 about here.]

4.3. Empirical Investigation of Scaling Factors and Regularization Parameters

In order to assess the optimality of the proposed approach as described in Section 2.4, we conducted simulations for values of β in (16) using artificial pseudo-random noise generated according to the model (20). Since the reconstruction error performance is also affected by the choice of ϵ in (16), we similarly performed the simulations over different values ϵ. Preliminary simulations indicated that $\mathcal{P}$ (see (27)) evolves in a quasi-convex manner

18

over β and ϵ. Based on this observation, we have used the Nelder-Mead simplex algorithm [28] to find the (β, ϵ)-optimal error figures $\mathcal{P}$ for each of the points (M, K) listed in Section 3. The results for all (M, K) with correlated noise generated according to each of the values α (Lloyd-Max) in Table 1 are shown in tables 2 to 4. The optimal regularization parameter values for ordinary BPDN are denoted ϵ_1 – with resulting error figure $\mathcal{P}_1$, while the optimal scaling and regularization parameter values for the proposed method are denoted β_2 and ϵ_2 – with resulting error figure $\mathcal{P}_2$. The error figures from our proposed method as reported in Fig. 2 are included in tables 2 to 4 as $\mathcal{P}_\alpha$ to facilitate comparison.

[Table 2 about here.]

[Table 3 about here.]

[Table 4 about here.]

It was expected that α would be the optimal choice of β, i.e. $\beta = \alpha$. However, it turns out that the (empirically observed) optimal value of β_2 is typically slightly smaller than α with observed values $\beta_2 \in [0.74\alpha, 0.98\alpha]$, depending on (M, K). An exception is seen in Table 2, where $\beta_2 \in [1.0\alpha, 1.2\alpha]$.

The optimal values of the regularization parameter ϵ are similarly found to be lower than the values given by (13). For the baseline method (6), the (empirically observed) optimal values are observed as $\epsilon_1 \in [0.41\epsilon, 0.84\epsilon]$, depending on (M, K), where ϵ denotes the values given by (13) as described in Section 3. For our proposed method (14), the optimal values are generally closer to the values given by (13) with observed values $\epsilon_2 \in [0.37\epsilon, 1.1\epsilon]$, depending on (M, K).

19

It is important to note that the demonstrated advantage of our proposed approach in Section 4.1 is not merely a result of a particularly lucky choice of ϵ, as these experiments testify. The observed NMSE of our proposed method, $\mathcal{P}_2$, consistently outperforms the baseline approach, $\mathcal{P}_1$. The improvement is consistent across different correlation parameters α as seen in tables 2 to 4 where $\mathcal{P}_2$ is smaller than $\mathcal{P}_1$ by $13\,\mathrm{dB}$, $11\,\mathrm{dB}$ and $8.4\,\mathrm{dB}$ in tables 2 to 4, respectively, for $(M, K) = (200, 1)$. At the other extreme of $(M, K) = (1000, 542)$, $\mathcal{P}_2$ is smaller than $\mathcal{P}_1$ by $0.12\,\mathrm{dB}$, $0.090\,\mathrm{dB}$ and $0.20\,\mathrm{dB}$, respectively. Additionally, the observed NMSEs $\mathcal{P}_2$ are generally around an order of magnitude lower than $\mathcal{P}_\alpha$ arising from our proposed choices of $\beta = \alpha$ and ϵ according to (13). However, note that β_2 and ϵ_2 optimized through simulations are not practically useful.

5. Discussion

As seen from the experimental results in Section 4.3, the correlation parameter α from (7) may in fact not be the optimal choice of scaling parameter, as expressed by β in (16). The generally smaller values found in Section 4.3 to be optimal for BPDN reconstruction according to (16) can be explained by the fact that they scale the estimate $\hat{\mathbf{x}}_\beta$ larger. It is well-known in the literature that the ℓ_1-norm minimization approach represented by, e.g., (6) tends to penalize larger coefficients of $\mathbf{x}$ more than smaller coefficients [29], thus estimating the former relatively too small. Therefore, it is possible to choose a scaling parameter $\beta < \alpha$ in (16) that improves the estimate $\hat{\mathbf{x}}_\beta$, i.e. yields smaller $\|\hat{\mathbf{x}}_\beta - \mathbf{x}\|$ compared to $\|\hat{\mathbf{x}}_\alpha - \mathbf{x}\|$. At this time, we cannot quantify the optimal β analytically and it depends on the indeterminacy and/or

20

measurement density of the performed compressed sensing.

Regarding the comparison of the proposed method to BIHT, the two methods require two different kinds of prior information. BIHT requires knowing that the sparse vector $\mathbf{x}$ is unit-norm: $\|\mathbf{x}\|_2 = 1$. Our proposed method requires knowing the variance of the unquantized measurements $\bar{y}$ – the elements of $\bar{\mathbf{y}}$. It may depend on the specific application which quantity is more realistic to know about the signal. At least, the variance assumed known in our proposed method does not require any knowledge (such as norm) of the sparse representation $\mathbf{x}$ of the observed signal.

6. Conclusion

We proposed a simple technique to model correlation between measurements and an additive noise in compressed sensing signal reconstruction. The technique is based on a linear model of the correlation between the measurements and noise. It consists of scaling signals reconstructed by a well-known ℓ_1-norm convex optimization method according to the model and comes at negligible computational cost. We provided practical expressions for computing the scaling parameter and the reconstruction regularization parameter.

We performed numerical simulations to demonstrate the obtainable reconstruction error improvement by the proposed method compared to ordinary ℓ_1-norm convex optimization reconstruction for noise generated according to the model. We further demonstrated as an example that the model applies well to low-rate scalar quantization of the measurements; both Lloyd-Max quantization that complies accurately with the correlation model, as well as the more practical uniform quantization. For example, simulations indi-

cated that the proposed method offers improvements on the order of $1\,\mathrm{dB}$ to $7\,\mathrm{dB}$ for $1\,\mathrm{bit/sample}$ quantization, depending on the indeterminacy of the performed compressed sensing.

We compared the proposed approach to a state-of-the-art reconstruction method, Binary Iterative Hard Thresholding (BIHT), for the special case of $1\,\mathrm{bit/sample}$ quantization. This comparison showed that the proposed approach reconstructs signals with smaller error than BIHT when the signals contain more non-zero elements than an approximate fraction of 0.1 of the number of measurements. This indicated that the proposed method is able to reconstruct less sparse signals from 1-bit quantized measurements than BIHT is capable of.

We conducted numerical simulations to evaluate the validity of our results which confirmed that the improvements offered by the proposed method are not merely a coincidental result of the suggested practical choices of scaling and optimization regularization parameters. These results further indicated that the proposed method is robust to the choice of scaling and optimization regularization parameter in the sense that a suboptimal choice still leads to considerable improvements over the ordinary convex optimization reconstruction method.

Acknowledgements

References

[1] D. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (4) (2006) 1289–1306. `doi:10.1109/TIT.2006.871582`.

[2] E. J. Candès, J. Romberg, T. Tao, Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information, IEEE Trans. Inf. Theory 52 (2) (2006) 489–509. `doi:10.1109/TIT.2005.862083`.

[3] E. J. Candès, Compressive sampling, in: Proceedings of the International Congress of Mathematicians, Vol. 3, 2006, pp. 1433–1452.

[4] E. J. Candès, Y. C. Eldar, D. Needell, P. Randall, Compressed sensing with coherent and redundant dictionaries, Applied and Computational Harmonic Analysis 31 (1) (2011) 59–73. `doi:10.1016/j.acha.2010.10.002`.

[5] E. J. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inf. Theory 51 (12) (2005) 4203–4215. `doi:10.1109/TIT.2005.858979`.

[6] D. Donoho, J. Tanner, Precise undersampling theorems, Proc. IEEE 98 (6) (2010) 913–924. `doi:10.1109/JPROC.2010.2045630`.

[7] E. J. Candès, T. Tao, Near-optimal signal recovery from random projections: Universal encoding strategies?, IEEE Trans. Inf. Theory 52 (12) (2006) 5406–5425. `doi:10.1109/TIT.2006.885507`.

[8] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal

of the Royal Statistical Society. Series B (Methodological) 58 (1) (1996) 267–288.

[9] S. S. Chen, D. L. Donoho, M. A. Saunders, Atomic decomposition by basis pursuit, SIAM J. Sci. Comput. 20 (1) (1998) 33–61. `doi:10.1137/S1064827596304010`.

[10] E. Candes, T. Tao, The Dantzig selector: Statistical estimation when p is much larger than n, Annals of Statistics 35 (6) (2007) 2313–2351. `doi:10.1214/009053606000001523`.

[11] T. Blumensath, M. Davies, Normalized iterative hard thresholding: Guaranteed stability and performance, IEEE J. Sel. Topics Signal Process. 4 (2) (2010) 298–309. `doi:10.1109/JSTSP.2010.2042411`.

[12] W. Dai, O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, IEEE Trans. Inf. Theory 55 (5) (2009) 2230–2249. `doi:10.1109/TIT.2009.2016006`.

[13] D. Needell, J. A. Tropp, CoSaMP: Iterative signal recovery from incomplete and inaccurate samples, Applied and Computational Harmonic Analysis 26 (3) (2009) 301–321. `doi:10.1016/j.acha.2008.07.002`.

[14] A. Maleki, D. Donoho, Optimally tuned iterative reconstruction algorithms for compressed sensing, IEEE J. Sel. Topics Signal Process. 4 (2) (2010) 330–341. `doi:10.1109/JSTSP.2009.2039176`.

[15] M. Raginsky, R. Willett, Z. Harmany, R. Marcia, Compressed sensing performance bounds under poisson noise, Signal Processing, IEEE

Transactions on 58 (8) (2010) 3990–4002. `doi:10.1109/TSP.2010.2049997`.

[16] L. Jacques, D. Hammond, J. Fadili, Dequantizing compressed sensing: When oversampling and non-gaussian constraints combine, Information Theory, IEEE Transactions on 57 (1) (2011) 559–571. `doi:10.1109/TIT.2010.2093310`.

[17] L. Jacques, J. Laska, P. Boufounos, R. Baraniuk, Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors, IEEE Trans. Inf. Theory 59 (4) (2013) 2082–2102. `doi:10.1109/TIT.2012.2234823`.

[18] P. C. Hansen, Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion (Monographs on Mathematical Modeling and Computation), Society for Industrial Mathematics, Philadelphia, 1987.

[19] S. Becker, J. Bobin, E. J. Candes, NESTA: A fast and accurate first-order method for sparse recovery, SIAM Journal on Imaging Sciences 4 (1) (2011) 1–39. `doi:10.1137/090756855`.

[20] M. Elad, Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing, Springer, New York, 2010. `doi:10.1007/978-1-4419-7011-4`.

[21] N. S. Jayant, P. Noll, Digital Coding of Waveforms - Principles and Applications to Speech and Video, Prentice Hall, Englewood Cliffs, New Jersey, 1984.

[22] S. Lloyd, Least squares quantization in PCM, IEEE Trans. Inf. Theory 28 (2) (1982) 129–137. `doi:10.1109/TIT.1982.1056489`.

[23] J. Max, Quantizing for minimum distortion, IEEE Trans. Inf. Theory 6 (1) (1960) 7–12. `doi:10.1109/TIT.1960.1057548`.

[24] P. H. Westerink, J. Biemond, D. E. Boekee, Scalar quantization error analysis for image subband coding using qmfs, IEEE Trans. Signal Process. 40 (2) (1992) 421–428. `doi:10.1109/78.124952`.

[25] D. Donoho, A. Maleki, A. Montanari, The noise-sensitivity phase transition in compressed sensing, IEEE Trans. Inf. Theory 57 (10) (2011) 6920–6941. `doi:10.1109/TIT.2011.2165823`.

[26] S. M. Ross, Introduction to Probability and Statistics for Engineers and Scientists, 2nd Edition, Academic Press, San Diego, 2000.

[27] E. van den Berg, M. P. Friedlander, Probing the Pareto frontier for basis pursuit solutions, SIAM J. Sci. Comput. 31 (2) (2008) 890–912. `doi:10.1137/080714488`.

[28] J. A. Nelder, R. Mead, A simplex method for function minimization, The Computer Journal 7 (4) (1965) 308–313. `doi:10.1093/comjnl/7.4.308`.

[29] E. Candès, M. Wakin, S. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, Journal of Fourier Analysis and Applications 14 (5) (2008) 877–905. `doi:10.1007/s00041-008-9045-x`.

List of Figures

27
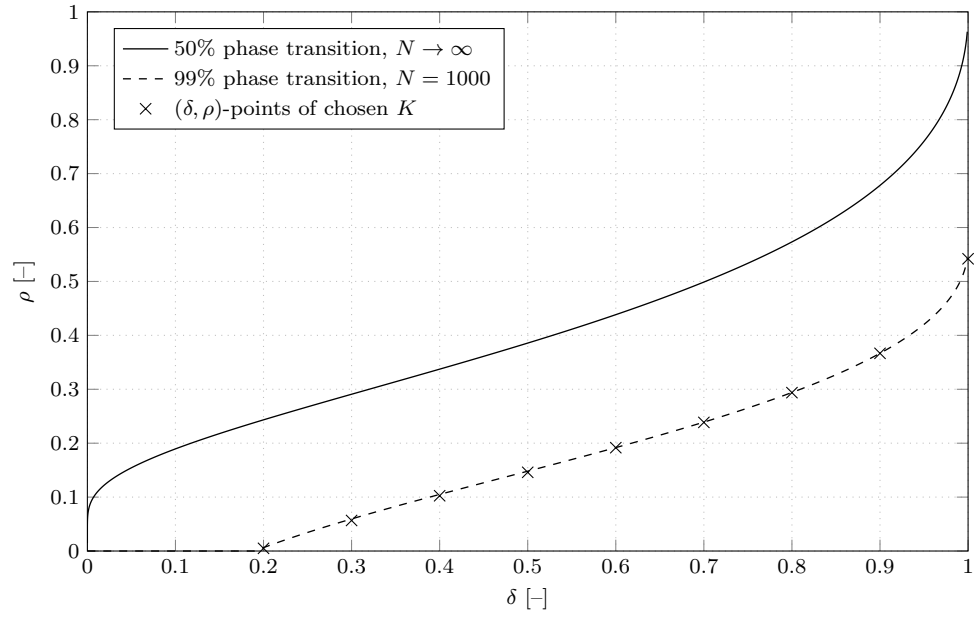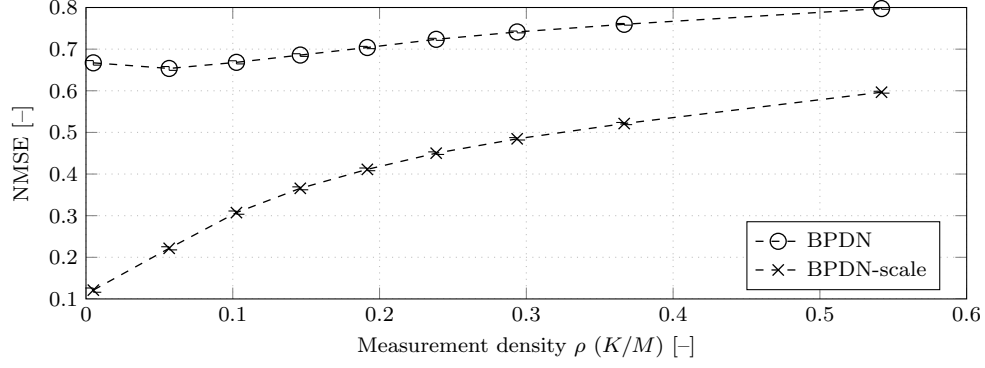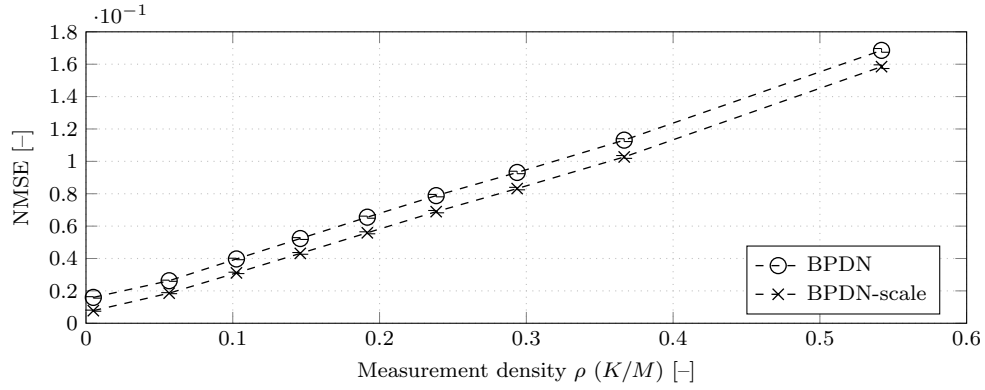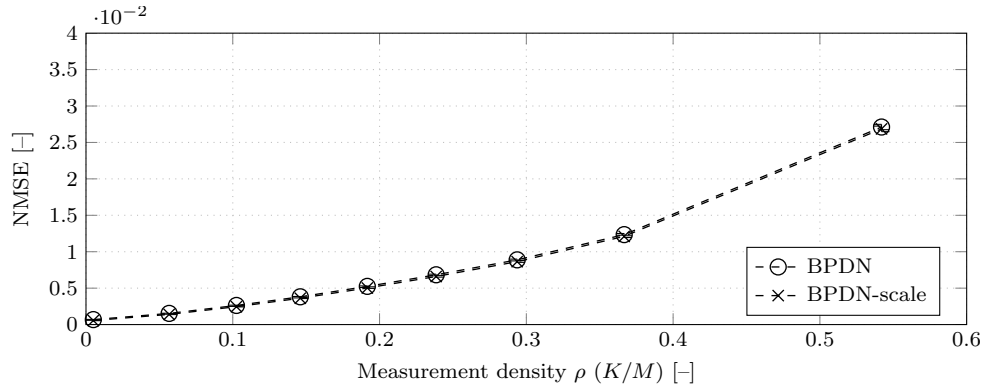
Figure 1: The theoretical Donoho-Tanner phase transition for $N \to \infty$, lower bound for $N = 1000$, and points corresponding to the values of K chosen for test signals.
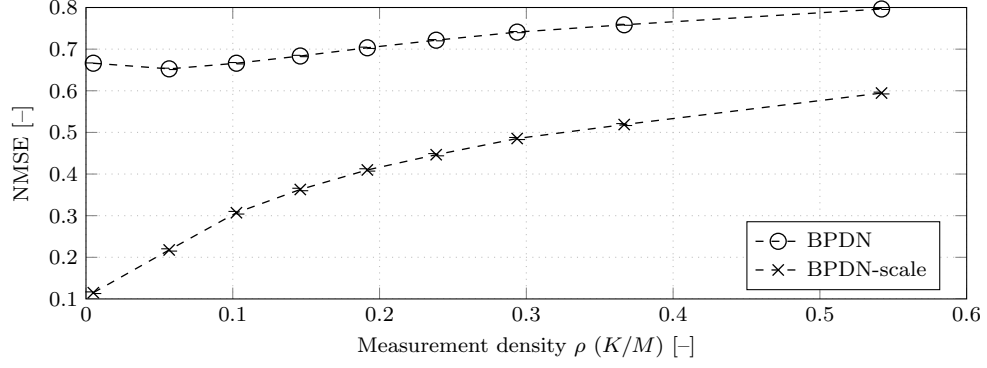
(a) Noise of var. equivalent to $1\,\mathrm{bit/sample}$ quantizer.

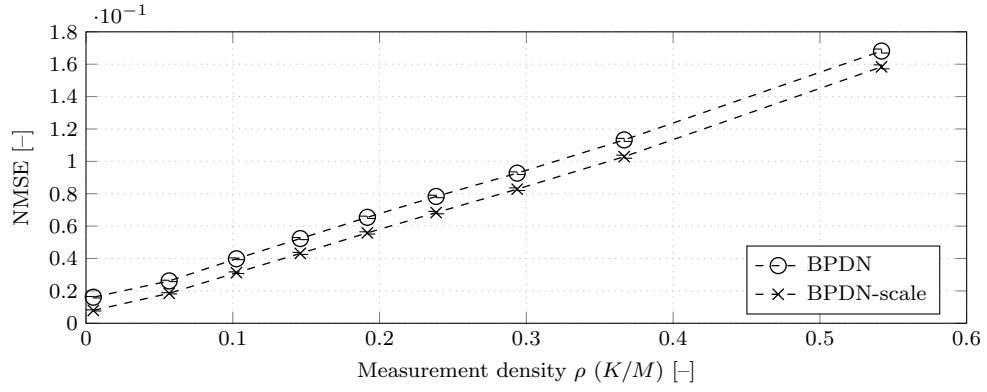(b) Noise of var. equivalent to $3\,\mathrm{bit/sample}$ quantizer.

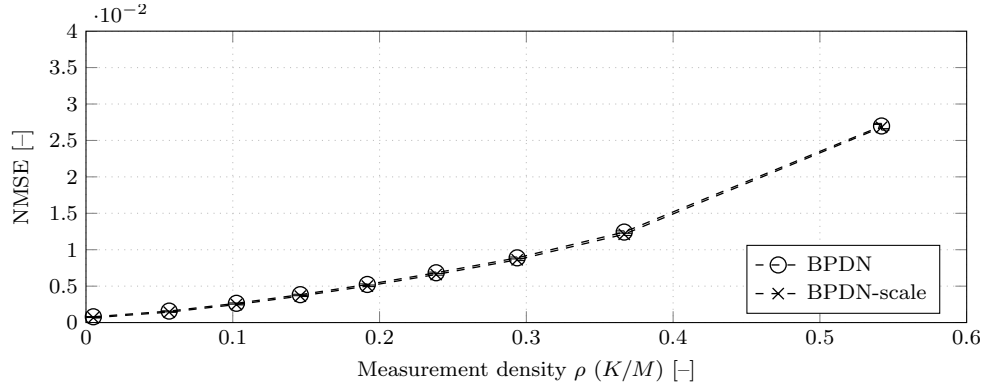(c) Noise of var. equivalent to $5\,\mathrm{bit/sample}$ quantizer.

Figure 2: Simulated NMSE of reconstruction using BPDN vs. relative number of measurements for parameters α and $\sigma_{\mathbf{r}}^2$ equal to corresponding values for Lloyd-Max quantizers.

(a) 1 bit/sample quantizer.

(b) 3 bit/sample quantizer.

(c) 5 bit/sample quantizer.

Figure 3: Simulated NMSE of reconstruction using BPDN vs. relative number of measurements for parameters α and $\sigma_{\mathbf{r}}^2$ equal to corresponding values for Lloyd-Max quantizers.

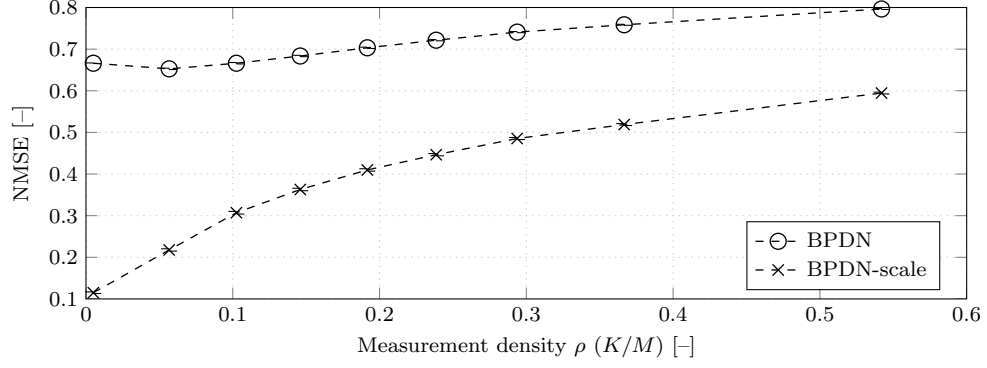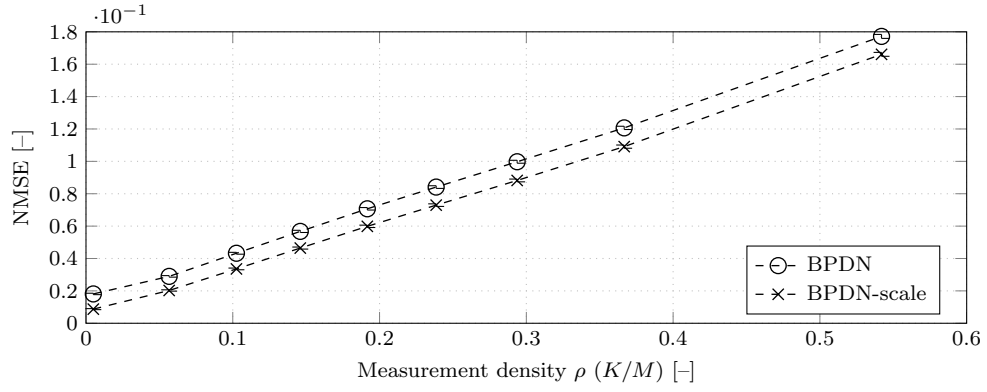(a) 1 bit/sample quantizer.

(b) 3 bit/sample quantizer.
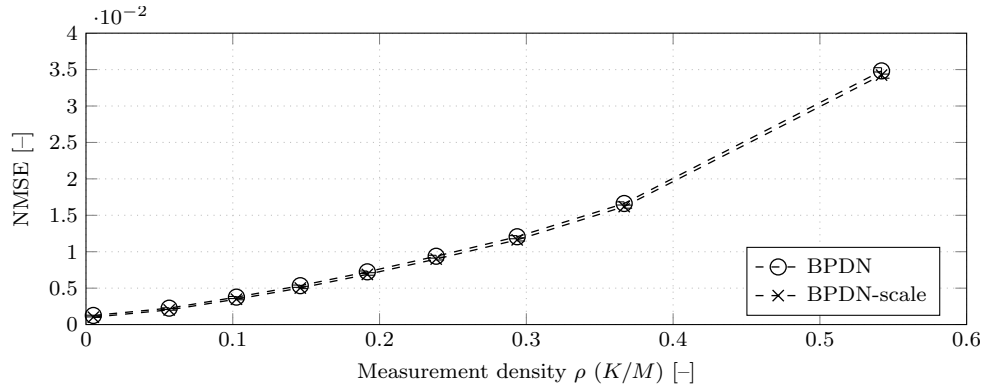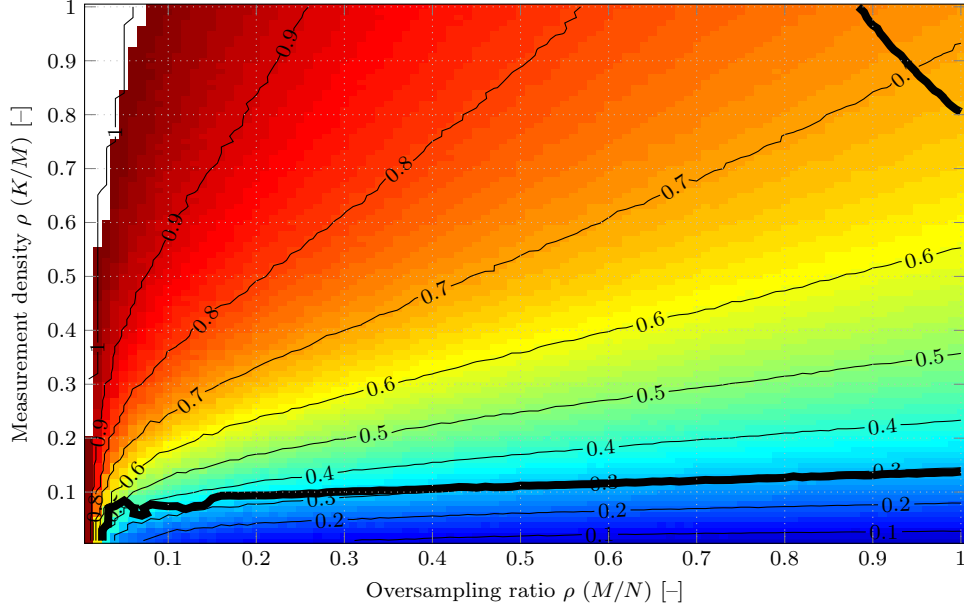

(c) 5 bit/sample quantizer.

Figure 4: Simulated NMSE of reconstruction using BPDN vs. relative number of measurements for parameters α and $\sigma_{\mathbf{r}}^2$ equal to corresponding values for uniform quantizers.

(a) Proposed method (BPDN-scale).
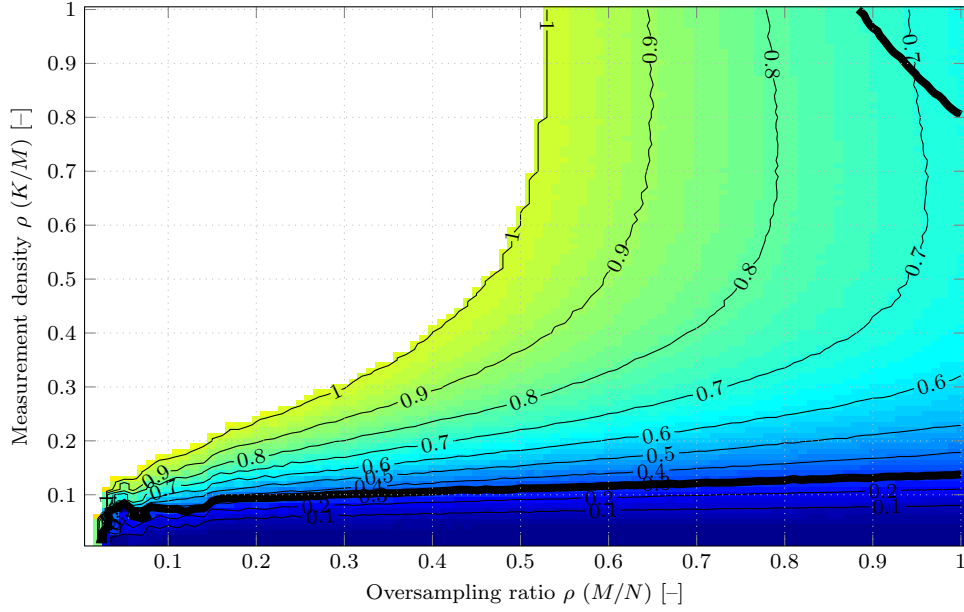
(b) BIHT.

Figure 5: Simulated NMSE of reconstruction from 1-bit quantized measurements. The numbered (—0.1— etc.) contour lines trace equal NMSE levels. The bold contour lines (——) mark the boundary where the NMSE levels of the proposed method and BIHT, respectively, are equal. ("Wiggly" contour lines are caused by interpolation in Matlab).

List of Tables

Table 1: Correlation parameter values used in Figs. 2–4.

Equiv. quantizer resolution	α (Lloyd-Max)	α (uniform)
$1\,\mathrm{bit/sample}$	0.6366	0.6366
$3\,\mathrm{bit/sample}$	0.9655	0.9626
$5\,\mathrm{bit/sample}$	0.9975	0.9965

Table 2: Simulated NMSE at empirically optimal parameter values β and ϵ. Noise equivalent to $1\,\mathrm{bit/sample}$ quantizer.

(M, K)	ϵ_1/ϵ	$\mathcal{P}_1$	β_2/α	ϵ_2/ϵ	$\mathcal{P}_2$	$\mathcal{P}_\alpha$
$(200, 1)$	0.60	$2.0 \cdot 10^{-2}$	0.84	0.92	$1.0 \cdot 10^{-3}$	$1.2 \cdot 10^{-1}$
$(300, 17)$	0.50	$3.3 \cdot 10^{-2}$	0.74	0.93	$1.4 \cdot 10^{-2}$	$2.2 \cdot 10^{-1}$
$(400, 41)$	0.51	$3.9 \cdot 10^{-2}$	0.83	0.87	$2.2 \cdot 10^{-2}$	$3.1 \cdot 10^{-1}$
$(500, 73)$	0.46	$4.3 \cdot 10^{-2}$	0.77	0.93	$2.9 \cdot 10^{-2}$	$3.7 \cdot 10^{-1}$
$(600, 115)$	0.51	$4.5 \cdot 10^{-2}$	0.79	0.90	$3.4 \cdot 10^{-2}$	$4.1 \cdot 10^{-1}$
$(700, 167)$	0.47	$4.6 \cdot 10^{-2}$	0.88	0.84	$3.8 \cdot 10^{-2}$	$4.5 \cdot 10^{-1}$
$(800, 235)$	0.50	$4.8 \cdot 10^{-2}$	1.00	0.74	$4.2 \cdot 10^{-2}$	$4.8 \cdot 10^{-1}$
$(900, 330)$	0.46	$4.9 \cdot 10^{-2}$	1.0	0.73	$4.5 \cdot 10^{-2}$	$5.2 \cdot 10^{-1}$
$(1000, 542)$	0.46	$5.3 \cdot 10^{-2}$	1.2	0.65	$5.1 \cdot 10^{-2}$	$6.0 \cdot 10^{-1}$

Table 3: Simulated NMSE at empirically optimal parameter values β and ϵ. Noise equivalent to 3 bit/sample quantizer.

(M, K)	ϵ_1/ϵ	$\mathcal{P}_1$	β_2/α	ϵ_2/ϵ	$\mathcal{P}_2$	$\mathcal{P}_\alpha$
$(200, 1)$	0.84	$5.0 \cdot 10^{-4}$	0.90	1.1	$3.8 \cdot 10^{-5}$	$7.8 \cdot 10^{-3}$
$(300, 17)$	0.75	$2.0 \cdot 10^{-3}$	0.91	0.93	$8.9 \cdot 10^{-4}$	$1.9 \cdot 10^{-2}$
$(400, 41)$	0.64	$3.0 \cdot 10^{-3}$	0.88	0.96	$1.8 \cdot 10^{-3}$	$3.1 \cdot 10^{-2}$
$(500, 73)$	0.65	$3.9 \cdot 10^{-3}$	0.87	0.93	$2.7 \cdot 10^{-3}$	$4.3 \cdot 10^{-2}$
$(600, 115)$	0.65	$4.8 \cdot 10^{-3}$	0.92	0.74	$3.6 \cdot 10^{-3}$	$5.6 \cdot 10^{-2}$
$(700, 167)$	0.61	$5.6 \cdot 10^{-3}$	0.90	0.73	$4.5 \cdot 10^{-3}$	$6.9 \cdot 10^{-2}$
$(800, 235)$	0.55	$6.4 \cdot 10^{-3}$	0.92	0.67	$5.4 \cdot 10^{-3}$	$8.3 \cdot 10^{-2}$
$(900, 330)$	0.50	$7.4 \cdot 10^{-3}$	0.93	0.63	$6.8 \cdot 10^{-3}$	$1.0 \cdot 10^{-1}$
$(1000, 542)$	0.49	$1.0 \cdot 10^{-2}$	0.98	0.51	$1.0 \cdot 10^{-2}$	$1.6 \cdot 10^{-1}$

Table 4: Simulated NMSE at empirically optimal parameter values β and ϵ. Noise equivalent to $5\,\text{bit/sample}$ quantizer.

(M, K)	ϵ_1/ϵ	$\mathcal{P}_1$	β_2/α	ϵ_2/ϵ	$\mathcal{P}_2$	$\mathcal{P}_\alpha$
$(200, 1)$	0.83	$1.8 \cdot 10^{-5}$	0.97	1.0	$2.6 \cdot 10^{-6}$	$5.6 \cdot 10^{-4}$
$(300, 17)$	0.70	$1.2 \cdot 10^{-4}$	0.97	0.94	$6.3 \cdot 10^{-5}$	$1.4 \cdot 10^{-3}$
$(400, 41)$	0.65	$2.0 \cdot 10^{-4}$	0.97	0.96	$1.3 \cdot 10^{-4}$	$2.5 \cdot 10^{-3}$
$(500, 73)$	0.68	$2.8 \cdot 10^{-4}$	0.97	0.85	$1.9 \cdot 10^{-4}$	$3.6 \cdot 10^{-3}$
$(600, 115)$	0.63	$3.8 \cdot 10^{-4}$	0.97	0.72	$2.7 \cdot 10^{-4}$	$5.0 \cdot 10^{-3}$
$(700, 167)$	0.50	$4.8 \cdot 10^{-4}$	0.96	0.72	$3.6 \cdot 10^{-4}$	$6.6 \cdot 10^{-3}$
$(800, 235)$	0.56	$5.8 \cdot 10^{-4}$	0.97	0.66	$4.8 \cdot 10^{-4}$	$8.6 \cdot 10^{-3}$
$(900, 330)$	0.45	$7.3 \cdot 10^{-4}$	0.96	0.58	$6.4 \cdot 10^{-4}$	$1.2 \cdot 10^{-2}$
$(1000, 542)$	0.41	$1.3 \cdot 10^{-3}$	0.98	0.37	$1.3 \cdot 10^{-3}$	$2.7 \cdot 10^{-2}$