MISMATCHED SPARSE DENOISER REQUIRES OVERESTIMATING THE SUPPORT LENGTH

Giulio Coluccia[⋆], *Aline Roumy*[†], *Enrico Magli*[⋆]

[⋆] Politecnico di Torino, Italy [†] INRIA, France

ABSTRACT

A well-known result [1, Lemma 3.4] states that, without noise, it is better to overestimate the support of a sparse signal, since, if the estimated support includes the true support, the reconstruction is perfect. In this paper, we investigate whether this result holds also in the presence of noise. First, we derive the covariance matrix of the signal estimate when the observation matrix is Gaussian, generalizing existing results. Then, we show that, even in the noisy case, overestimating the support length is the preferred solution, as the error incurred by missing some signal components dominates the overall error variance. Finally, an upper bound of the estimated support length is provided to avoid excessive noise amplification.

***Index Terms*—** Compressive sensing, sparse approximation, support recovery.

1. INTRODUCTION

We consider the problem of denoising a signal $\mathbf{x} \in \mathbb{R}^N$, where the signal is observed through an encoder $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ and the observation is corrupted by noise $\mathbf{z} \in \mathbb{R}^M$, leading to the observation model $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{z}$. This task is indeed central in signal processing and it has been shown [2] that having a sparse prior on the signal model can help improve the reconstruction of the signal $\mathbf{x}$. In this paper, we focus on a kind of sparse estimator, where first the support in which the sparse signal lives is estimated, and then an orthogonal projection is performed on the subspace. This assumption is not too restrictive as in each type of sparse estimator (optimization, greedy or thresholding based methods [1, Chap. 3]), there exists at least one estimator that can be seen as this two-step process. For instance, orthogonal matching pursuit, basic thresholding and the projected-LARS [2, Sec. 5.3.4], all perform a final orthogonal projection on a previously estimated subspace.

Analyses exist in the case the support is perfectly known [3, 4, 5]. However, support recovery is not always successful [2, Fig. 5.13], [6, 7] and evaluating the effect of an error in the support estimate is of great importance to evaluate the ℓ_2-norm error variance on the signal. This is, for instance, useful to derive the rate–distortion performance [4]. Regarding support mismatch, an interesting result [1, Lemma 3.4] states that, if the estimated support includes the true support then reconstruction is perfect. This result holds without noise and we conclude that it is therefore better to overestimate the support length and therefore have a chance to retrieve at least all the true support.

The goal of this paper is to analyze the tradeoff between the accuracy of the support estimate and the impact of noise on the reconstruction error and, in a way, to generalize [1, Lemma 3.4] to the presence of noise. We assume that the encoder matrix performs dimensionality reduction ($M \ll N$) and that the matrix entries are independent and identically distributed (i.i.d.) Gaussian variables, as for instance, in the compressive sensing (CS) context [8, 9]. The obtained result is that, in the presence of noise, even if there is a tradeoff between model accuracy and noise in the system, it is generally better to overestimate the support and not miss any component of the signal. Nevertheless, the overestimation should avoid the regime where the estimated length gets close to the number of observations.

In the rest of the paper, we use, when possible, the following notation: a random vector $\mathbf{x}$ is denoted using boldface lowercase and a random matrix $\mathbf{M}$ is denoted using boldface uppercase. $\mathbf{M}_i$ refers to the ith column of $\mathbf{M}$ and $\mathbf{M}_S$ is the submatrix extracted from $\mathbf{M}$ while keeping the columns indexed by the elements in th set S. Covariance matrices are denoted using the boldface uppercase greek letter $\mathbf{\Sigma}$, the identity matrix of dimension K is denoted $\mathbf{I}_K$, and the all zero $N \times K$ matrix is denoted $\mathbf{0}_{N \times K}$.

2. SYSTEM MODEL

CS [8, 9] aims at recovering a signal $\mathbf{x} \in \mathbb{R}^N$, having a K–sparse representation in some basis $\mathbf{\Psi} \in \mathbb{R}^{N \times N}$, *i.e.*: $\mathbf{x} = \mathbf{\Psi}\boldsymbol{\theta}$, $\|\boldsymbol{\theta}\|_0 = K$, $K \ll N$, by a smaller vector of noisy linear measurements $\mathbf{y} = \mathbf{\Phi}\mathbf{x} + \mathbf{z}$, $\mathbf{y} \in \mathbb{R}^M$ and $K < M < N$. $\mathbf{\Phi} \in \mathbb{R}^{M \times N}$ is the *sensing matrix* and $\mathbf{z} \in \mathbb{R}^M$ is the vector representing additive noise. If each submatrix consisting of K columns of $\mathbf{\Phi}\mathbf{\Psi}$ is (almost) distance preserving [10, Def. 1.3] [1, Def. 6.1], then reconstruction is possible. The latter condition is the *Restricted Isometry Property* (RIP). Formally, the matrix $\mathbf{\Phi}\mathbf{\Psi}$ satisfies the RIP of order K if $\exists \delta_K \in (0, 1]$ such that, for any $\boldsymbol{\theta}$ with $\|\boldsymbol{\theta}\|_0 \leq K$:

$$(1 - \delta_K) \|\boldsymbol{\theta}\|_2^2 \leq \|\mathbf{\Phi}\mathbf{\Psi}\boldsymbol{\theta}\|_2^2 \leq (1 + \delta_K) \|\boldsymbol{\theta}\|_2^2 , \qquad (1)$$

This work has been supported by the European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013) / ERC Grant agreement n 279848

where δ_K is the RIP constant of order K. First, the value of this constant that guarantees reconstruction depends on the algorithm used [1, Sec. 6.2 to 6.4]. Second, it has been shown in [11] that when $\mathbf{\Phi}$ is an i.i.d. random matrix drawn from any subgaussian distribution and $\mathbf{\Psi}$ is an orthogonal matrix, $\mathbf{\Phi\Psi}$ satisfies the RIP with overwhelming probability provided that $M = \mathcal{O}(K \log(N/K))$. RIP is therefore a nice property that ensures success of CS algorithms. However, the quality of success is given by (1), which is unfortunately not tight. In [4] the exact performance of CS algorithms based on orthogonal projection were derived, under the condition that the support of the sparse component were perfectly known. In this paper, we generalize this result without any restriction on the estimated support.

We consider the measurement vector as

$$\mathbf{y} = \frac{1}{\sqrt{M}}\mathbf{\Phi}\mathbf{x} + \mathbf{z}\,. \tag{2}$$

where $\mathbf{z}$ is a centered random vector with covariance matrix $\mathbf{\Sigma}_z$ and $\mathrm{vec}\{\mathbf{\Phi}\} \sim \mathcal{N}\left(\mathbf{0}, \sigma_\phi^2 \mathbf{I}_{MN}\right)$. (2) complies with the usual form $\mathbf{y} = \mathbf{\Phi}\mathbf{x}+\mathbf{z}$ where the variance σ_ϕ^2 of the elements of $\mathbf{\Phi}$ depends on M. Here, we want to keep σ_ϕ^2 independent of system parameters.

We further assume that $\mathbf{x}$ is directly sparse. This holds without loss of generality since the distribution of the Gaussian matrix Φ is invariant under orthogonal transformation. Referring to the true sparsity support of the signal as Λ, with $|\Lambda| = K$, we can rewrite the above equation as

$$\mathbf{y} = \frac{1}{\sqrt{M}}\mathbf{\Phi}_\Lambda \mathbf{x}_\Lambda + \mathbf{z}\,,$$

where $\mathbf{x}_{\Lambda^c} = \mathbf{0}$, $\mathbf{x}_\Lambda$ is a centered random vector with covariance matrix $\sigma_x^2 \mathbf{I}_K$ and $\mathbf{x}_\Lambda$, $\mathbf{\Phi}$ and $\mathbf{z}$ are independent from each other.

3. MISMATCHED DENOISER PERFORMANCE

Let us assume that a receiver is willing to estimate the signal $\mathbf{x}$ from its CS measurements $\mathbf{y}$. It is provided with an estimation Ω of the true sparsity support Λ, with $|\Omega| = L < M$. The estimated support Ω may for example derive from a run of a greedy algorithm estimating the sparsity support, first, then estimating the signal by pseudo-inverting the sensing matrix, like OMP [12], CoSaMP [13] or Subspace Pursuit [14]. Hence, given Ω, the signal estimation is given by

$$\begin{cases} \widehat{\mathbf{x}}_\Omega & = \sqrt{M}\mathbf{\Phi}_\Omega^\dagger \mathbf{y} \\ \widehat{\mathbf{x}}_{\Omega^c} & = \mathbf{0} \end{cases} \tag{3}$$

Theorem 1. *Let the random sparse signal $\mathbf{x} \in \mathbb{R}^N$ be observed through the noisy linear model* (2)*, with $\mathbf{y}, \mathbf{z} \in \mathbb{R}^M$. Let Λ be the true support of $\mathbf{x}$ and Ω the estimated support, such that $|\Lambda| = K$, $|\Omega| = L$ and $k = |\Lambda \cap \Omega|$ is the number of positions in the support that are correctly estimated. Let $\hat{\mathbf{x}}$ be the orthogonal projection of $\mathbf{y}$ on the subspace spanned by the estimated support* (3)*. Let $\mathbf{x}_\Lambda, \mathbf{z}$ be centered random vectors with covariance matrices $\sigma_x^2\mathbf{I}_K$ and $\sigma_z^2\mathbf{Q}\mathbf{Q}^\mathsf{T}$ respectively, where $\mathbf{Q}$ is an arbitrary orthogonal matrix, and let the entries of the matrix $\mathbf{\Phi}$ be i.i.d Gaussian centered with variance σ_ϕ^2. Then the covariance matrix of the signal estimate is*

$$\mathbb{E}[\widehat{\mathbf{x}}\widehat{\mathbf{x}}^\mathsf{T}] = \sigma_x^2 \begin{pmatrix} \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} + \sigma_x^2 \frac{K-k}{M-L-1}\mathbf{I}_L + \frac{\sigma_z^2}{\sigma_\phi^2}\frac{M}{M-L-1}\mathbf{I}_L$$

More generally, for a centered noise $\mathbf{z}$ with covariance matrix $\mathbf{\Sigma}_z$, the norm of the estimation error $\mathbf{e} = \widehat{\mathbf{x}} - \mathbf{x}$ is

$$\mathbb{E}\left[\|\mathbf{e}\|^2\right] = \sigma_x^2(K-k)\frac{M-1}{M-L-1} + \frac{L}{M-L-1}\frac{\mathrm{Tr}\left(\mathbf{\Sigma}_z\right)}{\sigma_\phi^2}\,,$$

Proof. Let us now assume without loss of generality that the nonzeros of $\mathbf{x}$ are concentrated at the beginning of the vector, *i.e.,* $\Lambda = \{1, 2, \ldots, K\}$ and that Ω is split into two continuous subsets, one corresponding to the beginning of Λ, the other corresponding to the beginning of Λ^c (see Fig. 1). We can then distinguish between the k detected components, where $k = |\Lambda \cap \Omega|$ and the $K - k$ undetected components, where $K - k = |\Lambda \cap \Omega^c|$.

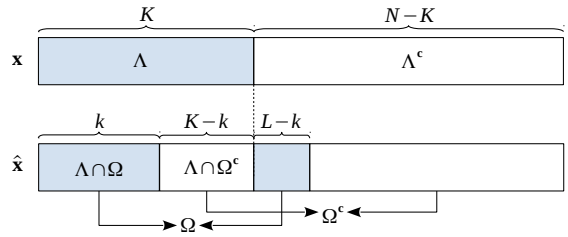


Fig. 1. Relationships between index sets Λ and Ω. Note that Λ^c contains only zeroes.

First note that $\mathbb{E}\left[\widehat{\mathbf{x}}\right] = \mathbf{0}$ due to the independence of $\mathbf{x}$, $\mathbf{\Phi}$, $\mathbf{z}$ and the fact that $\mathbb{E}\left[\mathbf{x}\right] = \mathbf{0}$ and $\mathbb{E}\left[\mathbf{z}\right] = \mathbf{0}$.

As for the covariance matrix $\mathbb{E}\left[\widehat{\mathbf{x}}\widehat{\mathbf{x}}^\mathsf{T}\right]$, we develop (3) in the following way:

$$\begin{aligned} \widehat{\mathbf{x}}_\Omega &= \sqrt{M}\mathbf{\Phi}_\Omega^\dagger\left(\frac{1}{\sqrt{M}}\mathbf{\Phi}_\Lambda\mathbf{x}_\Lambda + \mathbf{z}\right) \\ &= \mathbf{\Phi}_\Omega^\dagger\left(\mathbf{\Phi}_{\Lambda\cap\Omega}\mathbf{x}_{\Lambda\cap\Omega} + \mathbf{\Phi}_{\Lambda\cap\Omega^c}\mathbf{x}_{\Lambda\cap\Omega^c} + \sqrt{M}\mathbf{z}\right) \\ &= \mathbf{\Phi}_\Omega^\dagger\left(\mathbf{\Phi}_{\Lambda\cap\Omega}\mathbf{x}_{\Lambda\cap\Omega} + \mathbf{\Phi}_{\Lambda^c\cap\Omega}\mathbf{x}_{\Lambda^c\cap\Omega}\right. \\ &\quad \left. + \mathbf{\Phi}_{\Lambda\cap\Omega^c}\mathbf{x}_{\Lambda\cap\Omega^c} + \sqrt{M}\mathbf{z}\right) \\ &= \mathbf{\Phi}_\Omega^\dagger\left(\mathbf{\Phi}_\Omega\mathbf{x}_\Omega + \mathbf{\Phi}_{\Lambda\cap\Omega^c}\mathbf{x}_{\Lambda\cap\Omega^c} + \sqrt{M}\mathbf{z}\right) \\ &= \mathbf{x}_\Omega + \mathbf{\Phi}_\Omega^\dagger\left(\mathbf{\Phi}_{\Lambda\cap\Omega^c}\mathbf{x}_{\Lambda\cap\Omega^c} + \sqrt{M}\mathbf{z}\right) \end{aligned} \tag{4}$$

In (4), we can notice the presence of three terms. The first represents the effectively estimated true signal components. The second is an error term accounting for undetected terms.

Finally, the third represents the noise. Each term is uncorrelated from each other, hence the covariance $\mathbb{E}\left[\widehat{\mathbf{x}}_\Omega \widehat{\mathbf{x}}_\Omega^\mathsf{T}\right]$ will be the summation of the covariances of those three components. The first term is given by

$$\mathbb{E}\left[\mathbf{x}_\Omega \mathbf{x}_\Omega^\mathsf{T}\right] = \begin{pmatrix} \sigma_x^2 \mathbf{I}_k & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{(L-k)\times(L-k)} \end{pmatrix} . \tag{5}$$

The second term can be derived as

$$\begin{aligned} &\mathbb{E}\Big[\mathbf{\Phi}_\Omega^\dagger \mathbf{\Phi}_{\Lambda\cap\Omega^c} \underbrace{\mathbf{x}_{\Lambda\cap\Omega^c}\mathbf{x}_{\Lambda\cap\Omega^c}^\mathsf{T}}_{\sigma_x^2\mathbf{I}_{(K-k)}} \mathbf{\Phi}_{\Lambda\cap\Omega^c}^\mathsf{T} (\mathbf{\Phi}_\Omega^\dagger)^\mathsf{T}\Big] \\ &= \sigma_x^2 \mathbb{E}\Big[\left(\mathbf{\Phi}_\Omega^\mathsf{T}\mathbf{\Phi}_\Omega\right)^{-1} \mathbf{\Phi}_\Omega^\mathsf{T} \underbrace{\mathbf{\Phi}_{\Lambda\cap\Omega^c}\mathbf{\Phi}_{\Lambda\cap\Omega^c}^\mathsf{T}}_{\sigma_\phi^2(K-k)\mathbf{I}_M} \mathbf{\Phi}_\Omega \left(\mathbf{\Phi}_\Omega^\mathsf{T}\mathbf{\Phi}_\Omega\right)^{-1}\Big] \\ &= \sigma_x^2\sigma_\phi^2(K-k)\mathbb{E}\Big[\underbrace{\left(\mathbf{\Phi}_\Omega^\mathsf{T}\mathbf{\Phi}_\Omega\right)^{-1}}_{\frac{1}{\sigma_\phi^2(M-L-1)}\mathbf{I}_L}\Big] = \sigma_x^2 \frac{K-k}{M-L-1}\mathbf{I}_L . \end{aligned} \tag{6}$$

The second equality follows from the fact that $\mathbf{\Phi}_{\Lambda\cap\Omega^c}\mathbf{\Phi}_{\Lambda\cap\Omega^c}^\mathsf{T}$ is Wishart distributed with moments given in [15, Sec 3.2.2] and the third equality follows from the fact that $\left(\mathbf{\Phi}_\Omega^\mathsf{T}\mathbf{\Phi}_\Omega\right)^{-1}$ is the inverse of a Wishart matrix, which moment is given in [16]. For what concerns the noise term, we have

$$\mathbb{E}\Big[\left(\mathbf{\Phi}_\Omega^\mathsf{T}\mathbf{\Phi}_\Omega\right)^{-1} \mathbf{\Phi}_\Omega^\mathsf{T} \underbrace{\mathbf{z}M\mathbf{z}^\mathsf{T}}_{M\mathbf{\Sigma}_z} \mathbf{\Phi}_\Omega \left(\mathbf{\Phi}_\Omega^\mathsf{T}\mathbf{\Phi}_\Omega\right)^{-1}\Big], \tag{7}$$

which can be evaluated in closed-form when $\mathbf{\Sigma}_z$ has all equal eigenvalues, *i.e.*, in the case where the noise distribution is spherically symmetric. In this case, there exists an orthogonal matrix $\mathbf{Q}$ such that $\mathbf{Q}\mathbf{\Sigma}_z\mathbf{Q}^\mathsf{T} = \sigma_z^2\mathbf{I}_M$ and it follows that

$$\begin{aligned} &\mathbb{E}\Big[\left(\mathbf{\Phi}_\Omega^\mathsf{T}\mathbf{\Phi}_\Omega\right)^{-1} \mathbf{\Phi}_\Omega^\mathsf{T} \underbrace{\mathbf{z}M\mathbf{z}^\mathsf{T}}_{M\mathbf{\Sigma}_z} \mathbf{\Phi}_\Omega \left(\mathbf{\Phi}_\Omega^\mathsf{T}\mathbf{\Phi}_\Omega\right)^{-1}\Big] \\ &= M\sigma_z^2 \mathbb{E}\left[\left(\mathbf{\Phi}_\Omega^\mathsf{T}\mathbf{\Phi}_\Omega\right)^{-1} \mathbf{\Phi}_\Omega^\mathsf{T}\mathbf{Q}^\mathsf{T}\mathbf{Q}\mathbf{\Phi}_\Omega \left(\mathbf{\Phi}_\Omega^\mathsf{T}\mathbf{\Phi}_\Omega\right)^{-1}\right] \\ &= \frac{M}{M-L-1}\frac{\sigma_z^2}{\sigma_\phi^2}\mathbf{I}_L . \end{aligned} \tag{8}$$

Summing the terms of (5), (6) and (8) the final covariance matrix can be obtained. It has to be remarked that (6) and (8) hold for $M > L+1$.

The error norm can be evaluated as follows

$$\begin{aligned} \mathbb{E}\left[\|\mathbf{e}\|^2\right] &= \mathbb{E}\left[(\widehat{\mathbf{x}}-\mathbf{x})^\mathsf{T}(\widehat{\mathbf{x}}-\mathbf{x})\right] \\ &= \mathbb{E}\left[(\widehat{\mathbf{x}}_\Omega-\mathbf{x}_\Omega)^\mathsf{T}(\widehat{\mathbf{x}}_\Omega-\mathbf{x}_\Omega)\right] + \mathbb{E}\left[\mathbf{x}_{\Lambda\cap\Omega^c}^\mathsf{T}\mathbf{x}_{\Lambda\cap\Omega^c}\right] \end{aligned} \tag{9}$$

and it can be derived for a generic $\mathbf{\Sigma}_z$ since the trace of (7) can be expressed in closed form. It follows that[1]

$$\begin{aligned} &\mathbb{E}\left[\|\mathbf{e}\|^2\right] = (K-k)\sigma_x^2 + \\ &+ \mathrm{Tr}\left(\sigma_x^2\frac{K-k}{M-L-1}\mathbf{I}_L\right) + \mathrm{Tr}\left(\frac{1}{\sigma_\phi^2}\frac{L}{M-L-1}\mathbf{\Sigma}_z\right) \\ &= \sigma_x^2(K-k)\frac{M-1}{M-L-1} + \frac{L}{M-L-1}\frac{\mathrm{Tr}\left(\mathbf{\Sigma}_z\right)}{\sigma_\phi^2} \end{aligned} \tag{10}$$

from which the contributions of the mismatched components and of the noise can be easily distinguished. Of course, when there is no mismatch ($k = L = K$) the first term disappears and (10) reduces to (45) in [4, Appendix B] . Moreover, in absence of noise ($\sigma_\phi^2 = 0$), (10) specializes to (28) in [17]. It can be noticed that the error norm only depends on the variance of the elements of $\mathbf{z}$ and not on its covariance matrix. Therefore, our result holds even if the noise is correlated (for instance if vector quantization is used). As a consequence, we can apply our result to any quantization algorithm or to noise not resulting from quantization. Note that, if the elements of $\mathbf{z}$ have the same variance, (10) reduces to

$$\mathbb{E}\left[\|\mathbf{e}\|^2\right] = \sigma_x^2(K-k)\frac{M-1}{M-L-1} + \frac{\sigma_z^2}{\sigma_\phi^2}\frac{LM}{M-L-1} . \tag{11}$$

□

Corollary 2. *In the presence of observation noise, the mismatched sparse denoiser performs better in case of overestimation of the support length.*

Proof. (11) shows that by increasing the size of the estimated support L will increase the impact of noise in the reconstruction error. However, for the usual range of signal-to-noise ratios, the signal variance σ_x^2 is much greater than the observation noise variance. Since in the first term of (11) σ_x^2 is multiplied by $K-k$, *i.e.,* the number of missed non-zero coefficients in the estimated support Ω, this term dominates the second term. Thus, to avoid penalizing too much the estimate noise variance, one should avoid missing some significant terms in the support. This can be easier achieved if one is allowed to estimate a bigger support than the true one Λ. □

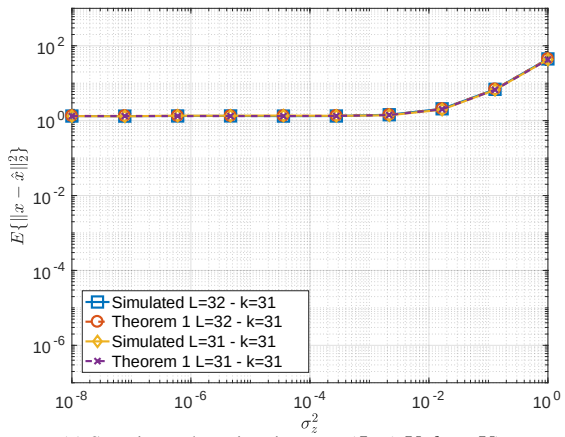
Corollary 3. *If the estimated support length goes to M, the noise variance of the estimate goes to infinity. Moreover, to avoid amplification of the noise variance by a factor $\alpha > 1$, the length of the support should be smaller than $M\frac{\alpha-1}{\alpha}$.*

Proof. The noise variance (11) on the signal estimate contains a factor $\approx M/(M-L-1)$ that goes to ∞ as $L \to M$. Moreover, $M/(M-L-1) < \alpha \Leftrightarrow L+1 < M(\alpha-1)/\alpha$. □

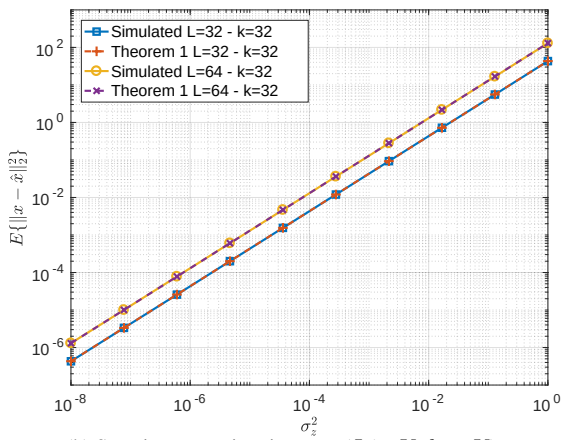
4. NUMERICAL RESULTS

In this section, we show the validity of the results of Section 3 by comparing the equations to the results of simulations. Here and in the following sections, signal length is $N = 512$ with

[1] Errata from the published version. The modification is written in red.



(a) Sparsity underestimation case ($L \leq K, k < K$).



(b) Sparsity overestimation case ($L \geq K, k = K$).

Fig. 2. Mismatched reconstruction error. Simulations vs. (1). $N = 512$, $K = 32$, $M = 128$. White noise: $\mathbf{\Sigma}_z = \sigma_z^2 \mathbf{I}_M$.

sparsity $K = 32$. $M = 128$ measurements are taken. The nonzero elements of the signal are distributed as $\mathcal{N}(0, 1)$. The sensing matrix is composed by i.i.d. elements distributed as zero–mean Gaussian with variance $\sigma_\phi^2 = 1$. The noise vector is Gaussian with zero mean, while the covariance matrix depends on the specific test and will be discussed later. The reconstructed signal $\widehat{\mathbf{x}}$ is obtained using the mismatched estimator of (3). A different realization of the signal, noise and sensing matrix is drawn for each trial, and the reconstruction error, evaluated as $\mathbb{E}\left[\|\mathbf{e}\|_2^2\right]$, is averaged over 10,000 trials.

4.1. White noise

In this first experiment, the measurement vector $\mathbf{y}$ is corrupted with white Gaussian noise, *i.e.,* $\mathbf{z} \sim \mathcal{N}_p(\mathbf{0}, \sigma_z^2 \mathbf{I}_M)$. The first set of experiments shows the comparison between the simulated reconstruction error and (11) in two cases. Fig. 2(a) shows the case in which the decoder underestimates the sparsity of the signal missing only 1 component of the true nonzeros ($k = 31$ and $L = 31, 32$). On the other hand, Fig. 2(b) shows the case in which the decoder overestimates the sparsity support of the signal ($\Lambda \subseteq \Omega$ with $L = 32, 64$). In both cases, it can be easily noticed that the match between simulated and theoretical curve is perfect. In the former case, it can be noticed that missing only one component of the true signal from the estimated support strongly affects the performance, making the noise contribution to the error almost negligible for high SNRs. Instead, in the latter case the first term of (11) goes to 0 with the performance affected only by L . This means that even when the measurements are noisy it is better to overestimate the sparsity support of the signal to increase the likelihood to include all the true signal components in the estimated support.

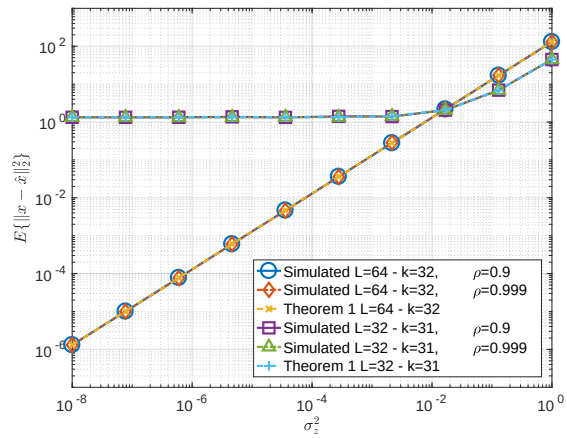


Fig. 3. Same as Figure 2 but with correlated noise: $(\mathbf{\Sigma}_z)_{i,j} = \sigma_z^2 \rho^{|i-j|}$ and $\rho = 0.9,\ 0.999$.

4.2. Correlated noise

We also report in Fig. 3 the results obtained reconstructing the signal from the measurements corrupted by correlated noise. In particular, the i, j-th element of the noise covariance matrix will be given by $(\mathbf{\Sigma}_z)_{i,j} = \sigma_z^2 \rho^{|i-j|}$. The correlation coefficient takes the values of $\rho = 0.9$ and 0.999. We compare the simulations with (1). It can be noticed from Fig. 3 that simulations confirm the result that the performance of the oracle does not depend on noise covariance but only on its variance. This is shown by the fact that simulations for $\rho = 0.9$ overlap the ones for $\rho = 0.999$, and both match (1) even in the correlated noise scenario.

5. CONCLUSIONS

In this paper, we analytically derived the closed form expressions of the covariance matrix of the estimation of a sparse signal from noisy linear measurements with mismatch in the estimated sparsity support, as well as the closed form of the estimation error norm. The results hold also for correlated noise. They confirm that for these kind of decoders it is a better choice to increase the probability of picking all the signal nonzeros by overestimating the sparsity support, because the penalty due to the miss of a signal nonzero in the estimation is far greater than the additional noise caused by the overestimation of the signal support.

6. REFERENCES

[1] Simon Foucart and Holger Rauhut, *A Mathematical Introduction to Compressive Sensing*, Birkhäuser Basel, 2013.

[2] Michael Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010.

[3] Giulio Coluccia, Aline Roumy, and Enrico Magli, "Exact performance analysis of the oracle receiver for compressed sensing reconstruction," in *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2014, pp. 1005–1009.

[4] Giulio Coluccia, Aline Roumy, and Enrico Magli, "Operational rate-distortion performance of single-source and distributed compressed sensing," *IEEE Transactions on Communications*, vol. 62, no. 6, pp. 2022–2033, 2014.

[5] A. K. Fletcher and S. Rangan, "Orthogonal matching pursuit: A Brownian motion analysis," *IEEE Trans. on Signal Processing*, vol. 60, no. 3, pp. 1010–1021, 2012.

[6] G. Reeves and M. Gastpar, "The Sampling Rate-Distortion Tradeoff for Sparsity Pattern Recovery in Compressed Sensing," *IEEE Trans. on Information Theory*, vol. 58, no. 10, pp. 3065–3092, May 2012.

[7] G. Reeves and M. Gastpar, "Approximate Sparsity Pattern Recovery: Information-Theoretic Lower Bounds," *IEEE Trans. on Information Theory*, vol. 59, no. 6, pp. 3451–3465, June 2013.

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

[9] E.J. Candès and T. Tao, "Near-Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?," *IEEE Transactions on Information Theory*, vol. 52, no. 12, pp. 5406–5425, 2006.

[10] Y.C. Eldar and G. Kutyniok, Eds., *Compressed Sensing: Theory and Applications*, Cambridge University Press, 2012.

[11] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, "A simple proof of the restricted isometry property for random matrices," *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[12] Joel A Tropp and Anna C Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4655–4666, 2007.

[13] Deanna Needell and Joel A Tropp, "Cosamp: Iterative signal recovery from incomplete and inaccurate samples," *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301–321, 2009.

[14] Wei Dai and Olgica Milenkovic, "Subspace pursuit for compressive sensing signal reconstruction," *IEEE Transactions on Information Theory*, vol. 55, no. 5, pp. 2230–2249, 2009.

[15] Robb John Muirhead, *Aspects of multivariate statistical theory*, Wiley series in probability and mathematical statistics. Probability and mathematical statistics. John Wiley & Sons, New York, 1982.

[16] D. von Rosen, "Moments for the inverted wishart distribution," *Scandinavian Journal of Statistics*, pp. 97–109, 1988.

[17] Rémi Gribonval, Volkan Cevher, and Mike E Davies, "Compressible distributions for high-dimensional statistics," *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 5016–5034, 2012.