Recovery guarantees for compressed sensing with unknown errors

Simone Brugiapaglia*, Ben Adcock†
Department of Mathematics
Simon Fraser University
Burnaby, BC V5A 1S6, Canada
Email: *simone_brugiapaglia@sfu.ca, †ben_adcock@sfu.ca

Richard K. Archibald
Computer Science and Mathematics Division
Oak Ridge National Laboratory
Oak Ridge, TN 37831, USA
Email: archibaldrk@ornl.gov

Abstract—From a numerical analysis perspective, assessing the robustness of $\ell^1$-minimization is a fundamental issue in compressed sensing and sparse regularization. Yet, the recovery guarantees available in the literature usually depend on a priori estimates of the noise, which can be very hard to obtain in practice, especially when the noise term also includes unknown discrepancies between the finite model and data. In this work, we study the performance of $\ell^1$-minimization when these estimates are not available, providing robust recovery guarantees for quadratically constrained basis pursuit and random sampling in bounded orthonormal systems. Several applications of this work are approximation of high-dimensional functions, infinite-dimensional sparse regularization for inverse problems, and fast algorithms for non-Cartesian Magnetic Resonance Imaging.

I. INTRODUCTION

In Compressed Sensing (CS) and sparse representations we deal with underdetermined linear systems of equations

$$y = Ax + n,$$

where $A \in \mathbb{C}^{m \times N}$, with $m \ll N$, is the sensing matrix, $x \in \mathbb{C}^N$ is an unknown signal, and $y \in \mathbb{C}^m$ is the vector of measurements perturbed by noise $n \in \mathbb{C}^m$. This corruption could be due to physical noise produced by the measuring device, to approximation errors in the model, or to numerical factors. Some examples are model error in inverse problems such as MRI, the expansion error in infinite-dimensional CS when truncating the signal to its finite dimensional representation, or the quadrature error involved in the evaluation of the bilinear form associated with a PDE.

A standard tool to regularize the inverse problem and recover a good approximation $\hat{x}(\eta)$ to the solution $x$ (assumed to be sparse or compressible) is the Quadratically Constrained Basis Pursuit (QCBP) optimization program

$$\hat{x}(\eta) \in \arg \min_{z \in \mathbb{C}^N} \|z\|_1, \quad s.t. \quad \|Az - y\|_2 \leq \eta,$$

also called Basis Pursuit (BP) when $\eta = 0$. Usually, in order to study the recovery guarantees of (2), the parameter $\eta$ is assumed to control the noise magnitude, i.e.,

$$\|n\|_2 \leq \eta.$$

Indeed, under the regime and with suitable hypotheses on the sensing matrix $A$ (e.g., based on the restricted isometry property), the following type of recovery error estimate holds with high probability

$$\|x - \hat{x}(\eta)\|_2 \lesssim \frac{\sigma_s(x)_1}{\sqrt{s}} + \eta,$$

where $\sigma_s(x)_1$ is the best $s$-term approximation error of $x$ with respect to the $\ell^1$-norm. Unfortunately, a priori estimates of the noise of the form may not be available in real applications of CS. Moreover, since the recovery error estimate is sensitive to $\eta$, the choice of this parameter is crucial (see Figure 1). In practice, one could resort to cross-validation in order to tune this parameter, but this technique could be time-consuming or inaccurate and it is not properly understood from a theoretical perspective.

The goal of this work is to establish robust recovery guarantees for QCBP (and BP) under the regime

$$\|n\|_2 \geq \eta.$$
Moreover, we recall the well-known restricted isometry property (also known as “RIP”), where the sensing matrix is required to behave similarly to an isometry when its action is restricted to the set of sparse vectors \([12]\).

**Definition 2 (Restricted isometry property):** The \(s^{th}\) restricted isometry constant \(\delta_s\) of a matrix \(A \in \mathbb{C}^{m \times N}\) is the smallest constant \(\delta \geq 0\) such that

\[
(1 - \delta)\|z\|_2^2 \leq \|Az\|_2^2 \leq (1 + \delta)\|z\|_2^2, \quad \forall z \in \Sigma^N.
\]

The matrix \(A\) has the Restricted Isometry Property (RIP) of order \(s\) if \(0 < \delta_s < 1\).

It is well-known that \(\delta_{2s} < 4/\sqrt{m}\) is a sufficient condition for the \(\ell^2\)-robust NSP to hold \([17]\) Theorem 6.13 (we decide to use this condition and not that presented in \([11]\) for ease of exposition).

### B. Bounded orthonormal systems

Our analysis focuses on the case of measurement matrices \(A\) arising from random sampling from a Bounded Orthonormal System (BOS) \([20]\). Some significant examples of random sampling from a BOS are the subsampled Fourier transform, nonharmonic Fourier measurements, and random evaluation of orthogonal polynomials. We will discuss these case studies in more detail in Section III-C.

**Definition 3 (Random sampling from a BOS):** Let \(D \subseteq \mathbb{R}^d\) be endowed with probability measure \(\nu\). Then, a set \(\Phi = \{\phi_1, \ldots, \phi_N\}\) of complex-valued functions on \(D\) is called a Bounded Orthonormal System (BOS) with constant \(K\) if, for every \(j, k \in [N]\), it holds \(\int_D \phi_j(\tau)\overline{\phi_k(\tau)} d\nu(\tau) = \delta_{jk}\) and \(\|\phi_j\|_{\infty} := \sup_{\tau \in D} |\phi_j(\tau)| \leq K\). Moreover, given \(m\) independent random variables \(\tau_1, \ldots, \tau_m\), distributed according to \(\nu\), we define the random sampling matrix \(A \in \mathbb{C}^{m \times N}\) associated with a the BOS \(\Phi\) as

\[
A_{kj} := m^{-1/2} \phi_j(\tau_k), \quad \forall k \in [m], j \in [N].
\]

A crucial property of this kind of matrices is the following. For every \(\delta \in (0, 1)\), assuming

\[
m \geq s \mathcal{L}(N, s, \delta, \varepsilon, K),
\]

the \(s^{th}\) restricted isometry constant of \(A\) satisfies \(\delta_s \leq \delta\) with probability at least \(1 - \varepsilon\), where the factor \(\mathcal{L}(N, s, \delta, \varepsilon, K)\) is polylogarithmic in \(N\) and \(s\) and can be chosen in different ways (see \([20], [17], [11]\)). For example, according to \([17]\) Theorem 12.32, it can be defined as follows

\[
\mathcal{L} = \frac{K^2}{\delta^2} \max \left\{ \ln^2(s) \left( \ln \frac{K^2}{\delta^2} s \ln(N) \right) \ln(N), \ln \left( \frac{1}{\varepsilon} \right) \right\}.
\]

### III. Robustness analysis

In this section, we illustrate our robustness analysis. For proofs of the results stated here, we refer to \([8]\).
\[ s_1(M) \geq s_2(M) \geq \cdots \geq s_m(M) = s_{\min}(M). \]  

First, we provide a robust error estimate for QCBP as defined in [2], assuming the $\ell^2$-robust NSP. This estimate depends on the minimum singular value of the Hermitian conjugate of the sensing matrix $A$.

**Proposition 4 (Robust error estimate based on the NSP):** Let $A \in \mathbb{C}^{m \times N}$ be a matrix of rank $m$ that satisfies the $\ell^2$-robust NSP of order $s$ with constants $0 < \rho < 1$ and $\tau > 0$ with respect to $\| \cdot \|_2$ (see Definition 1). Then, the following error estimate holds for problem (2)

\[ \| \tilde{x}(\eta) - x \|_2 \lesssim \frac{\sigma_s(x)}{\sqrt{s}} + \eta + \sqrt{m} \max\left\{ \| \mathbf{u} \|_2 - \eta, 0 \right\}, \]

where the hidden constant depends on $\rho$ and $\tau$.

Proposition 4 shows that, in order to get robust error estimates for QCBP, we need to understand the asymptotic behavior of $s_{\min}(\sqrt{\frac{N}{m}}A^*)$. In particular, our goal is to show that $s_{\min}(\sqrt{\frac{N}{m}}A^*) \approx 1$. With this aim, we employ principles and ideas from the theory of random matrices with isotropic heavy-tailed columns [25].

Recall that a random vector $z \in \mathbb{C}^N$ is said to be *isotropic* if $E[zz^*] = I$. Then, consider random matrices of the form

\[ M = [m_1] \cdots [m_n] \in \mathbb{C}^{N \times m}, \]

where the columns $m_j$ are independent random isotropic vectors. We introduce the *cross-coherence parameter*, defined as

\[ \mu := \frac{1}{N^2} E \max_{k \in [m]} \sum_{\ell \in [m] \setminus \{k\}} |\langle m_k, m_\ell \rangle|^2. \]

This parameter controls the off-diagonal part of the Gram matrix of $M$. We also define the *distortion parameter* as

\[ \xi := E \max_{k \in [m]} \left\| \frac{m_k}{\sqrt{N}} \right\|^2 - 1. \]

It measures how far the columns of $M$ are from being normalized. We notice that if $\| m_k \|_2 = \sqrt{N}$ almost surely for every $k \in [m]$, then $\xi$ vanishes. Using these two parameters, we give a generalization of [25] Theorem 5.62.

**Theorem 5 (Singular values of heavy-tailed matrices):** Let $M$ be an $N \times m$ matrix ($N \geq m$) whose columns are independent isotropic random vectors in $\mathbb{C}^N$. Then, its singular values satisfy the following asymptotic estimate

\[ E \max_{j \in [m]} s_j \left( \frac{1}{\sqrt{N}} M \right) - 1 \lesssim \xi + \sqrt{(1 + \xi)\mu \ln m}, \]

where $\mu$ and $\xi$ are defined as in [13] and [14], respectively.

We observe that a necessary condition for estimate (15) to be nontrivial is that the quantity $\xi$ should be bounded uniformly in $N$ (but not necessarily in $m$). Regarding the incoherence parameter, exploiting the isotropy of the columns of $M$, and assuming $\| m_j \|_2 \leq K \sqrt{N}$ for a suitable constant $K > 0$ (this is always the case for $M = \sqrt{m}A^*$, where $A$ is the random sampling matrix associated with a BOS), we can prove that

\[ N \mu \leq (Km)^2. \]

We check the sharpness of this upper bound for moderate values of $m$ numerically in the case of the subsampled Fourier transform, where $K = 1$ (see Figure 2).

**B. Recovery error estimate**

The following theorem is the main result of the paper. It provides a robust error estimate in probability for QCBP, when the solution $x$ is very sparse and $A$ is a random sampling matrix associated with a BOS.

**Theorem 6 (Robust recovery error estimate for QCBP):** Consider a BOS $\Phi$ with constant $K \geq 1$ and with a distortion parameter that satisfies

\[ \xi \leq \min \left\{ D_1 \sqrt{\frac{m^2 \ln(m)}{N}}, D_2 \right\}, \]

for suitable constants $D_1, D_2$ independent of $m$ and $N$. Then, there exist constants $c, d, C, D, E > 0$ and a function $L = L(N, s, \varepsilon, K)$ depending polylogarithmically on $N$ and $s$ such that the following holds. For every $N \in \mathbb{N}$ and $\varepsilon \in (0, 1)$, assume that the sparsity level satisfies

\[ s \leq c \varepsilon \sqrt{N}, \]

and let $A \in \mathbb{C}^{m \times N}$ be the random sampling matrix associated with $\Phi$ and with a number of measurements

\[ m = [d s L(N, s, \varepsilon, K)]. \]

Then, the following robust error estimate holds for QCBP

\[ \| \tilde{x}(\eta) - x \|_2 \leq \frac{C}{\sqrt{s}} \sigma_s(x) + D\eta + E L^2 \max\{\| u \|_2 - \eta, 0\}, \]

for suitable constants $D_1, D_2$ independent of $m$ and $N$. Then, there exist constants $c, d, C, D, E > 0$ and a function $L = L(N, s, \varepsilon, K)$ depending polylogarithmically on $N$ and $s$ such that the following holds. For every $N \in \mathbb{N}$ and $\varepsilon \in (0, 1)$, assume that the sparsity level satisfies

\[ s \leq c \varepsilon \sqrt{N}, \]

and let $A \in \mathbb{C}^{m \times N}$ be the random sampling matrix associated with $\Phi$ and with a number of measurements

\[ m = [d s L(N, s, \varepsilon, K)]. \]

Then, the following robust error estimate holds for QCBP

\[ \| \tilde{x}(\eta) - x \|_2 \leq \frac{C}{\sqrt{s}} \sigma_s(x) + D\eta + E L^2 \max\{\| u \|_2 - \eta, 0\}, \]
with probability at least $1 - \varepsilon$. The constant $c$ depends on $D_1$, $D_2$, and $K$, whereas the constants $d$, $C$, $D$, and $E$ are universal. The function $\mathcal{L}$ can be defined as in \[ \text{(10)}, \] with $\delta = 1/2$.

The required relation between $s$ and $m$ is linear up to logarithmic factors, in accordance with the usual recovery error estimate in CS. However, there are three main limitations of Theorem \[ \text{(6)} \] that are worth underlining. First, the result holds for a particular sparsity regime \[ \text{(18)} \]. Essentially, fixed the failure probability $\varepsilon$, we require $s \lesssim \sqrt{N}$ (up to logarithmic factors). Second, the error estimate \[ \text{(20)} \] actually depends on $N$, but this dependence is only polylogarithmic (the factor $\mathcal{L}^{\frac{1}{2}}$ is due to the term $\sqrt{m}/s$ in the error estimate of Proposition \[ \text{(4)} \]). Third, there is a linear dependence between $s$ and $\varepsilon$ in \[ \text{(18)} \]. Thus, the failure probability of the estimate is not “overwhelmingly low”. These three issues are open problems currently under investigation.

### C. Applications

To conclude, we discuss some applications of Theorem \[ \text{(6)} \] to concrete examples from signal processing and high-dimensional polynomial approximation.

#### 1) Fourier and Chebyshev BOSs: We discuss two examples of BOSs very popular in CS. First, we consider the subsampled Fourier transform. We have $D = [N]_0$, the system is defined as

$$\phi_j(\tau) = \exp \left(2\pi i j \tau/N\right), \quad \forall j, k \in [N]_0,$$

and the sampling measure $\nu$ is the uniform discrete distribution on $D$. It turns out that $K = 1$ and, consequently, the distortion parameter is $\xi = 0$. As a result, condition \[ \text{(17)} \] of Theorem \[ \text{(6)} \] is satisfied.

In the case of the Chebyshev system, we consider the Chebyshev orthogonal polynomials on $D = [-1, 1]$, defined as

$$\phi_0(\tau) \equiv 1,$$

$$\phi_j(\tau) = \sqrt{2} \cos(j \arccos(\tau)), \quad \forall j \in [N - 1].$$

They form a BOS with respect to the Chebyshev measure $d\nu(\tau) = \pi^{-1}(1 - \tau^2)^{-1/2}$ on $D$ and with constant $K = \sqrt{2}$. By studying the normalized Christoffel function associated with this system, we estimate that the distortion parameter $\xi$ decays proportionally to $\sqrt{m/N}$. Therefore, hypothesis \[ \text{(17)} \] holds true and we can apply Theorem \[ \text{(6)} \] to Chebyshev polynomials as a sparse basis. We set maximum degree 10 on each variable and we restrict the multi-index space to the *hyperbolic cross* shape \[ \text{(5)} \]. These choices leads to a total of $N = 581$ degrees of freedom. We evaluate $f$ at $m = 50$ random independent sampling points $\tau_1, \ldots, \tau_m$ identically distributed according to the tensorized Chebyshev measure over $D = [-1, 1]^{10}$. In order to assess the robustness to unknown error, we artificially add centered gaussian noise with standard deviation $\zeta$ to the measurements, namely

$$y_i = f(\tau_i) + N(0, \zeta^2), \quad \forall i \in [m].$$

#### 2) High-dimensional polynomial approximation: We assess the robustness of QCBB for polynomial approximation in high dimension \[ \text{(1)} \]. Consider the multivariate function

$$f(x) = \ln \left(d + 1 + \sum_{i=1}^{d} x_i\right), \quad x \in [-1, 1]^d.$$  

We fix $d = 10$ and we employ the tensorized version of the Chebyshev polynomials defined in \[ \text{(22)} \] over $[-1, 1]^{10}$ as a sparsity basis. We set maximum degree 10 on each variable and we restrict the multi-index space to the *hyperbolic cross* shape \[ \text{(5)} \].

### References

[1] B. Adcock, “Infinite-dimensional $\ell^p$ minimization and function approximation from pointwise data,” *Constr. Approx.*, (to appear), 2016.

[2] B. Adcock, “Infinite-dimensional compressed sensing and function interpolation,” *Found. Comput. Math.*, (in revision), 2016.

[3] B. Adcock, R. Archibald, A. Gelb, G. Song, R. B. Platte and E. G. Walsh, “Parameter assessment from time-dependent MR signals using sequential interpolation”, preprint, 2016.

[4] B. Adcock and A.C. Hansen, “Generalized sampling and infinite-dimensional compressed sensing,” *Found. Comput. Math.*, vol. 16, no. 5, pp. 1263–1323, 2016.

[5] K.I. Babenko, “Approximation by trigonometric polynomials in a certain class of periodic functions of several variables,” *Dokl. Akad. Nauk SSSR.*, Vol. 132, no. 5, 1960.
Fig. 3. We plot the absolute error $|\hat{x}(\eta) - x|_2$ as a function of the parameter $\eta$ when QCBP is applied to the sparse polynomial approximation of the function $f$ with $d = 10$. We consider 50 values of $\eta$ varying between $10^{-5}$ and $10^3$ with constant ratio and plot the results corresponding to 2 values of standard deviation $\zeta$ in $\mathcal{N}_p$: $\zeta = 0.1$ (top), $\zeta = 1$ (bottom). We estimate the optimal value $\eta_{opt}$ as in (20). Moreover, we assess the performance of cross-validation, where $\eta$ is chosen from a grid on the interval $[10^{-2}\eta_{opt}, 10^2\eta_{opt}]$ discretized by 21 points with constant ratio. The ratio between the reconstruction samples and the validation samples is set to 3/4 (see [10] for more details).

Fig. 4. The effect of gridding error on non-Cartesian MRI. A $256 \times 256$ phantom (the McGill phantom [6]) is sampled using a rosette sampling trajectory (top left) giving $m = 15626 \approx 0.238N$ non-Cartesian Fourier measurements. Upfront gridding of the data is performed, and then the images are recovered using total variation minimization. Top right: recovered image using standard nearest neighbor gridding. The signal-to-error ratio is SER = 8.38dB. Bottom row: recovered image using the novel fractional integer nearest neighbor gridding introduced in [6] with parameter $n_{up} = 2$ (left) and $n_{up} = 4$ (right). The signal-to-error ratios are SER = 14.88dB and SER = 17.84dB, respectively. Standard gridding leads to $O(1)$ model error. The fractional nearest neighbor gridding has $O(1/n_{up})$ model error.

[6] BrainWeb: Simulated Brain Database. [http://brainweb.bic.mni.mcgill.ca/brainweb/]

[7] S. Brugiapaglia, “Compressed Solving: sparse approximation of PDEs based on compressed sensing”, Ph.D. Thesis, Politecnico di Milano, 2016.

[8] S. Brugiapaglia, B. Adcock, R.K. Archibald, “Robustness to unknown error in sparse regularization”, in preparation, 2017.

[9] S. Brugiapaglia, S. Micheletti, and S. Perotto, “Compressed solving: a numerical approximation technique for elliptic PDEs based on Compressed Sensing”, Comput. Math. Appl., vol. 70, no. 6, pp. 1306–1335, 2015.

[10] S. Brugiapaglia, F. Nobile, S. Micheletti, and S. Perotto, “A theoretical study of Compressed Solving for advection-diffusion-reaction problems”, Math. Comp., (to appear), 2016.

[11] T. T. Cai and A. Zhang, “Sparse representation of a polytope and recovery of sparse signals and low-rank matrices.” IEEE Trans. Inf. Theory, vol. 60, no. 1, pp. 122–132, 2014.

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

[13] A. Chikifa, N. Dexter, H. Tran, and C. G. Webster, “Polynomial approximation via compressed sensing of high-dimensional functions on lower sets,” arXiv preprint arXiv:1602.05825, 2016.

[14] A. Cohen, W. Dahmen, and R.A. DeVore, “Compressed sensing and best k-term approximation,” J. Amer. Math. Soc., vol. 22, no. 1, pp. 211–231, 2009.

[15] D. L. Donoho, “Compressed Sensing,” IEEE Trans. Inf. Theory, vol. 52, no. 4, pp. 1289–1306, 2006.

[16] A. Doostan and H. Owhadi, “A non-adapted sparse approximation of PDEs with stochastic inputs,” J. Comput. Phys., vol. 230, no. 8, pp. 3015–3034, 2011.