2018 5th International Workshop on Compressed Sensing applied to Radar, Multimodal Sensing, and Imaging (CoSeRa)

Iterative Hard Thresholding with Optimal Measurement Matrices

Miguel Heredia Conde, Otmar Loffeld
Center for Sensorsystems (ZESS)
University of Siegen, Paul-Bonatz-Straße 9-11
57076 Siegen, Germany
E-mail: {heredia, loffeld}@zess.uni-siegen.de

***Abstract*—In compressive sensing (CS), signal recoverability is subject to satisfaction of the Restricted Isometry Property (RIP). Satisfying the RIP inequality with the tightest possible constants is related to finding the measurement matrix with the lowest possible coherence. Furthermore, methods exist that are able to generate sets of vectors tightly approaching the corresponding theoretical lower bounds on the coherence. Matrices constructed this way are optimal measurement matrices from a CS perspective. Nevertheless, the effective advantages of using these optimal matrices over randomly-generated ones have not been sufficiently studied in the literature and random matrices continue being the option of choice in most works on CS.**

In this work we compare different types of measurement matrices, in terms of preservation of l_2 distance between sparse vectors. As expected, minimal-coherence matrices outperform random matrices in this regard. Furthermore, since the distortion of l_2 distances is minimal and can be tightly bounded, its upper bound is used to calculate optimal values of the step size in the Iterative Hard Thresholding (IHT) algorithm. We show that IHT with properly adapted step size cannot converge any faster than using an optimal sensing matrix. We back this observation with simulations, which also show that optimal sensing matrices yield superior results in terms of l_2 reconstruction error. The faster convergence, together with the absence of an online adjustment of the IHT step size, yields reductions of the recovery time over 80%.

I. Introduction

Compressive sensing (CS) theory [1], [2] shows that real-world signals can often be recovered from a number of measurements that is well below that prescribed by the Shannon sampling theorem. CS models the sensing process as a linear system, where the measurements are obtained as linear projections of the signal, and exploits the *sparsity* or *compressibility* of the latter. Whether the signal can be reconstructed from the measurements or not directly depends on the properties of the measurement matrix used as linear projector. More specifically, it is critical that the linear transform preserves distances between the sparse signals we deal with up to some extent. This has been enunciated by means of the well-known Uniform Uncertainty Principle (UUP) [3] and the Restricted Isometry Property (RIP) [4].

Finding a measurement matrix satisfying the UUP or the RIP with the tightest possible constants can be shown to be related to finding a matrix with the lowest possible inter-column coherence. Furthermore, approximate methods can be found in the literature that can generate sets of vectors tightly approaching the corresponding theoretical lower bounds on the coherence, eventually with equality [5]. Matrices constructed this way are optimal measurement matrices from a CS perspective. In the real case, sets of vectors approaching the lower bound on the coherence as much as possible are known as *Best Antipodal Spherical Codes* (BASCs) and have been suggested as suitable CS matrices in [6]. In the complex case, all possible complex rotations of each vector are to be taken into account to ensure the construction of an antipodal code. This increases the computational cost required for constructing the codes. Sets of complex vectors achieving minimal coherence are known as *Best Complex Antipodal Spherical Codes* (BCASCs). An approximate method for constructing BCASCs has been presented in [5] and an accelerated version of the latter in [7]. Thanks to a dramatic reduction on the computational complexity of the algorithm, the method in [7] can construct large close-to-optimal measurement matrices in a reasonable amount of time. Despite the availability of construction methods, the effective advantages of using these optimal matrices over randomly-generated ones have not been sufficiently studied in the literature and random matrices continue being the option of choice in most works on CS.

In this work we aim to provide solid evidence of the actual advantages of using optimal measurement matrices over random ones. For generality, we work with complex signals throughout the paper. First we provide a comparison between the capabilities of different types of measurement matrices for bounding the distortion of the l_2 distance between sparse vectors. As expected, minimal-coherence matrices widely outperform random matrices in this regard. Furthermore, since the distortion of l_2 distances is minimal and can be tightly bounded, its upper bound is used to calculate theoretically optimal values of the step size in the Iterative Hard Thresholding (IHT) algorithm. We show that an IHT algorithm with properly adapted step size cannot converge any faster than using a BCASC. We conduct thorough simulations that show both the fastest convergence of IHT when using BCASCs as sensing matrices and the best global performance in terms of reconstruction error. We provide Donoho-Tanner graphs of reconstruction error for the two best-performing matrices, namely, random Gaussian matrices and BCASCs.

II. From Isometry Requirements to Coherence Minimization and Back

A. A Linear Sensing Model

Differently from the Shannon sampling theorem, which requires the signal to be bandlimited, CS theory [1], [2], [8], [9] imposes the more general requirement of the signal being *sparse* in some basis or tight frame. Measurements are no longer equidistant samples, but linear projections of the signal. In the discrete case, the signal is a vector whose coefficients may be real or complex, depending on the specific problem formulation. For generality, let $\vec{x} \in \mathbb{C}^n$ be the discrete signal we want to recover, in its sparse representation. Then CS theory states that $\vec{x}$ can be exactly recovered from few non-adaptive linear measurements of the shape

$$\vec{y} = \boldsymbol{A}\vec{x} \tag{1}$$

if the sparsity requirement is satisfied and the so-called *measurement matrix*, $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ is *good enough* for the sparsity level of $\vec{x} \in \mathbb{C}^n$. The great potential of CS lies on the fact that if such requirements are satisfied, $\vec{x} \in \mathbb{C}^n$ can be exactly recovered from $\vec{y} \in \mathbb{C}^m$ for $m \ll n$, using the sparsity constraint as a further restriction to the underdetermined system in Eq. 1. If the sparsity of the signal, s, is known beforehand, one may try to approach a solution to the system in Eq. 1 as follows:

$$\hat{\vec{x}} = \underset{\vec{x}\in\mathbb{C}^n}{\arg\min} \|\vec{y} - \boldsymbol{A}\vec{x}\|_2^2 \text{ subject to } \|\vec{x}\|_0 \leq s, \tag{2}$$

where the l_0 norm is defined as the cardinality of the support of $\vec{x}$:

$$\|\vec{x}\|_0 \coloneqq \lim_{p\to 0} \|\vec{x}\|_p^p = |\operatorname{supp}(\vec{x})| \tag{3}$$

B. From Isometry Requirements to Coherence Minimization

The sparsity of $\vec{x} \in \mathbb{C}^n$ depends on the application and cannot be arbitrarily reduced. The requirement of $\boldsymbol{A}$ being *good enough* for an expected s is now the critical point. Probably the most widely-spread characterization of the *goodness* of $\boldsymbol{A}$ from a CS perspective is by means of the so-called *Restricted Isometry Property* (RIP) [4]. A matrix $\boldsymbol{A}$ is said to satisfy the RIP of order k if there exists a constant $\delta_k \in (0, 1)$ such that

$$(1-\delta_k)\|\vec{x}\|_2^2 \leq \|\boldsymbol{A}\vec{x}\|_2^2 \leq (1+\delta_k)\|\vec{x}\|_2^2, \ \forall \vec{x} \in \Sigma_k \tag{4}$$

being δ_k known as the *k-restricted isometry constant* and Σ_k the subset of $\mathbb{C}^n$ containing all k-sparse vectors. Note that the RIP ensures that $\boldsymbol{A}$ is close to an isometry for k-sparse vectors, i. e., that the transformation preserves their l_2 norm to some extent. Clearly, uniqueness of the solution to Eq. 3 is only guaranteed if the l_2 distances between s-sparse vectors do not vanish, which requires that $\boldsymbol{A}$ satisfies the RIP of order $k = 2s$, since differences between s-sparse vectors are, at maximum, $2s$-sparse. Furthermore, if $\boldsymbol{A}$ satisfies the RIP of order $2s$ with δ_{2s} low enough, e. g., $\delta_{2s} < \sqrt{2} - 1$ [10], then successful signal recovery is guaranteed. The RIP imposes lower and upper bounds on the ratio $\|\boldsymbol{A}\vec{x}\|_2^2 / \|\vec{x}\|_2^2$ for $\vec{x} \in \Sigma_k$. For any specific realization of $\vec{x}$, let Ω_k denote its support set and $\boldsymbol{A}_{\Omega_k}$ the matrix obtained by selecting the k columns of $\boldsymbol{A}$ indexed by Ω_k. It can be shown that such ratio admits the following analytical bounds:

$$\lambda_{\min}\left(\boldsymbol{A}_{\Omega_k}^*\boldsymbol{A}_{\Omega_k}\right) \leq \frac{\|\boldsymbol{A}\vec{x}\|_2^2}{\|\vec{x}\|_2^2} \leq \lambda_{\max}\left(\boldsymbol{A}_{\Omega_k}^*\boldsymbol{A}_{\Omega_k}\right). \tag{5}$$

Consequently, a lower bound on $\lambda_{\min}\left(\boldsymbol{A}_{\Omega_k}^*\boldsymbol{A}_{\Omega_k}\right)$ and an upper bound on $\lambda_{\max}\left(\boldsymbol{A}_{\Omega_k}^*\boldsymbol{A}_{\Omega_k}\right)$, $|\Omega_k| = k$, are also lower and upper bounds, respectively, on the ratio $\|\boldsymbol{A}\vec{x}\|_2^2 / \|\vec{x}\|_2^2$ for $\vec{x} \in \Sigma_k$. This means that the Uniform Uncertainty Principle (UUP) [3], which requires actually such bounding of the eigenvalues of $\boldsymbol{A}_{\Omega_k}^*\boldsymbol{A}_{\Omega_k}$, implies the RIP up to appropriate choice of constants.

The challenge is then constructing $\boldsymbol{A}$, such that the spread of the eigenvalues of the Gram matrix of the Ω_k-restricted $\boldsymbol{A}$, $G_{\Omega_k} = \boldsymbol{A}_{\Omega_k}^*\boldsymbol{A}_{\Omega_k}$, is minimal over all possible support sets Ω_k of size k. It has been shown that matrices whose elements are drawn from Gaussian and Bernoulli distributions satisfy the RIP [11], but do not yield the desired minimal eigenvalue spread of G_{Ω_k}. Note that the aim is that $G_{\Omega_k} \to I_k$, elementwise:

$$\begin{aligned} g_{i,i} &\to 1, \forall i \leq k \\ |g_{i,j}| &\to 0, \forall i \leq k, j < i \end{aligned} \tag{6}$$

The requirement $g_{i,i} \to 1$ can be attained with equality, since it is a simple normalization matter, meaning that all columns of $\boldsymbol{A}$ should be of unit norm. The second requirement is related to the *coherence* between the columns of $\boldsymbol{A}$. Provided that it is not known beforehand which k indices are contained in Ω_k, it becomes necessary to minimize the *worst case coherence*, that is, the largest value of $|g_{i,j}|$. Formally, this often called *matrix coherence* [12] of $\boldsymbol{A}$ is defined as

$$\mu(\boldsymbol{A}) = \max_{u<v\leq n} \frac{|\langle \vec{a}_u, \vec{a}_v \rangle|}{\|\vec{a}_u\|_2 \|\vec{a}_v\|_2} \tag{7}$$

where $\langle \cdot, \cdot \rangle$ denotes (complex) scalar product. Clearly, the denominator does not play a role if the columns of $\boldsymbol{A}$ have unit norm. The problem is now reduced to finding a matrix $\boldsymbol{A}$ with minimal $\mu(\boldsymbol{A})$, formally:

$$\hat{\boldsymbol{A}} = \underset{\boldsymbol{A}\in\mathbb{C}^{m\times n}}{\arg\min}\, \mu(\boldsymbol{A}) = \underset{\boldsymbol{A}\in\mathbb{C}^{m\times n}}{\arg\min} \max_{u<v\leq n} \frac{|\langle \vec{a}_u, \vec{a}_v \rangle|}{\|\vec{a}_u\|_2 \|\vec{a}_v\|_2}. \tag{8}$$

A set of n unit vectors in $\mathbb{C}^m$ attaining minimal coherence is a Best Complex Antipodal Spherical Code (BCASC) [5]. That is, a BCASC is a solution to Eq. 8 and thus an optimal CS measurement matrix.

C. From Coherence Minimization to Isometry Guarantees

Theorem 2 of [13] states that the eigenvalues of a matrix $\boldsymbol{A} \in \mathbb{C}^{k\times k}$ lie in the union of n discs, $\bigcup_{i=1}^{k} d_i\left(c_i, r_i\right)$, centered at $c_i = a_{i,i}$ and with radius $r_i = \sum_{j\neq i}|a_{i,j}|$. Applying the theorem to $G_{\Omega_k} \in \mathbb{C}^{k\times k}$ we have the following bounds on its eigenvalues:

$$\begin{aligned}\lambda_{\min}\left(G_{\Omega_k}\right) &= \lambda_{\min}\left(\boldsymbol{A}_{\Omega_k}^{*}\boldsymbol{A}_{\Omega_k}\right) \geq 1-(k-1)\mu\left(\boldsymbol{A}\right)\\ \lambda_{\max}\left(G_{\Omega_k}\right) &= \lambda_{\max}\left(\boldsymbol{A}_{\Omega_k}^{*}\boldsymbol{A}_{\Omega_k}\right) \leq 1+(k-1)\mu\left(\boldsymbol{A}\right),\end{aligned} \tag{9}$$

which may be loose in the general case, but become tighter as $\mu\left(\boldsymbol{A}\right) \to \mu_{\min}\left(\boldsymbol{A}\right)$. Theoretical lower bounds on the $\mu_{\min}\left(\boldsymbol{A}\right)$ that can be attained when solving Eq. 8 can be found for any size of $\boldsymbol{A}$ (Eq. (9) of [5]) and both the BCASC construction method in [5] and its accelerated version in [7] have been shown to tightly approach them. Note that combining the inequalities in Eq. 9 with the double inequality in Eq. 5 we can bound the l_2 norm distortion undergone when projecting k-sparse vectors through $\boldsymbol{A}$, both from below and above, and exclusively in terms of $\mu\left(\boldsymbol{A}\right)$. More specifically, if the columns of $\boldsymbol{A}$ are of unit norm, one can show that $\boldsymbol{A}$ satisfies the RIP of order k with δ_k given by:

$$\delta_k = (k-1)\mu\left(\boldsymbol{A}\right),\ \forall k < 1/\mu\left(\boldsymbol{A}\right). \tag{10}$$

III. Fast Signal Recovery via Iterative Hard Thresholding

As critical as sensing matrix design is how signals are to be reconstructed from the measurements. One of the most attractive classes of recovery algorithms is that of *thresholding* algorithms, due to their simplicity and speed. Thresholding algorithms work iteratively, in a two-step structure in which first the current estimate of $\vec{x}$ is corrected by residual projection to signal space and then the coefficients of the updated $\vec{x}$ are thresholded to obtain an exactly-sparse estimate. The simplest algorithm implementing this scheme is the Iterative Hard Thresholding (IHT) [14], [15], and whose basic structure is given in Algorithm 1, where the stopping criterion is a threshold on the residual norm, ε_{tol}.

Algorithm 1 Iterative Hard Thresholding (IHT)

Initialize: $\vec{r}^{(0)} = \vec{y}, \vec{x}^{(0)} = \vec{0}$
1: **while** $\left(\left\|\vec{r}^{(i)}\right\|_2 > \varepsilon_{\text{tol}}\right)$ **do**
2: $\quad i := i+1$
3: $\quad$ Update estimate: $\vec{x}^{(i)} = H_k\left(\vec{x}^{(i-1)} + \alpha\boldsymbol{A}^{*}\vec{r}^{(i-1)}\right)$
4: $\quad$ Update residual: $\vec{r}^{(i)} = \vec{y} - \boldsymbol{A}\vec{x}^{(i)}$
5: **end while**

The *hard thresholding operator* in line 3, H_k, preserves the largest k entries of the input vector and sets the rest to zero. The input vector is the updated estimate and, if an estimate of s is known *a priori*, $k = s$. The update is controlled by the step size α. This parameter has a critical effect on the algorithm, since execution may become too slow if α is too low, eventually producing stagnation, while divergence may occur if it is too large. The need for a careful adjustment of α is one of the main weaknesses of IHT and similar alternatives, often requiring a manual adjustment for each specific problem. Alternatively, one can compute the optimal value of α at each iteration, at the price of increasing the computational cost per iteration. Supposing that the temporal support of $\vec{x}$, $\Omega^i = \text{supp}\left(\vec{x}^{(i)}\right)$, does not change from one iteration to the next, the optimal step size α can be calculated as follows:

$$\begin{aligned}\alpha_{\text{opt}}^{(i)} &= \frac{\left\|\vec{g}_{\Omega^{k-1}}^{(i-1)}\right\|_2^2}{\left\|\boldsymbol{A}_{\Omega^{k-1}}\vec{g}_{\Omega^{k-1}}^{(i-1)}\right\|_2^2}\\ \vec{g}^{(i-1)} &= \boldsymbol{A}^{*}\vec{r}^{(i-1)}\end{aligned} \tag{11}$$

where $(\cdot)_{\Omega^i}$ denotes restriction to the vector entries or matrix columns indexed by Ω^i. If $\text{supp}\left(\vec{x}^{(i)}\right) = \text{supp}\left(\vec{x}^{(i-1)}\right)$, then using the step size $\alpha_{\text{opt}}^{(i)}$ yields maximal error reduction without divergence risk.

The update step is not more than a gradient descent (GD) step aiming to minimize the square reprojection error. The smallest Lipschitz constant of the gradient of this cost function is $L = 2\lambda_{\max}\left(\boldsymbol{A}^{*}\boldsymbol{A}\right)$. Provided that the update step happens after a previous k-thresholding step and supposing support invariance, an estimate of $\lambda_{\max}\left(\boldsymbol{A}_{\Omega_k}^{*}\boldsymbol{A}_{\Omega_k}\right)$ can be used to calculate an analytically optimal IHT step size as follows:

$$\alpha_{\lambda_{\max}} = \frac{2}{L} = \frac{1}{\lambda_{\max}\left(\boldsymbol{A}_{\Omega_k}^{*}\boldsymbol{A}_{\Omega_k}\right)}. \tag{12}$$

If a unique value of $\alpha_{\lambda_{\max}}$ can be precomputed, there is no need for computing $\alpha_{\text{opt}}^{(i)}$ at each iteration i. The suitability of the constant step size given by Eq. 12 depends on two matters. On the one hand, the actual variability of $\lambda_{\max}\left(\boldsymbol{A}_{\Omega_k}^{*}\boldsymbol{A}_{\Omega_k}\right)$ over all possible realizations of $\boldsymbol{A}$ and Ω_k subject to $|\Omega_k| = k$ for a given k. If this variability is too large, the step size given by Eq. 12 may become uninformative. On the other hand, the magnitude of the estimate of $\lambda_{\max}\left(\boldsymbol{A}_{\Omega_k}^{*}\boldsymbol{A}_{\Omega_k}\right)$ itself plays a major role, since large eigenvalue spreads might translate into overpessimistic step sizes. Fortunately, due to their optimal coherence, the (eventually Ω_k-restricted) Gram matrix of a BCASC exhibits minimal eigenvalue spread for a given size, thus leading to the sharpest Lipschitz constant, thus yielding the largest possible step size in Eq. 12.

IV. Simulation Results

In this section we explore the feasibility of using an analytically-obtained $\alpha_{\lambda_{\max}}$ as fixed step size in IHT (Algorithm 1) by means of simulations. We consider the following types of complex sensing matrices:

- Complex Gaussian: both the real and imaginary parts of each matrix coefficient are drawn from *i.i.d.* normal distributions of zero mean and unit variance.
- Complex Bernoulli 0/1: both the real and imaginary parts of each matrix coefficient are drawn from *i.i.d.* Bernoulli distributions with 0.5 success probability.

- Complex Bernoulli -1/1: similar to the previous, but using $\{-1, 1\}$ instead of $\{0, 1\}$ as binary realization domain.
- BCASC: an approximate BCASC constructed using an accelerated version of the method in [5].

Matrices of the first three types undergo a subsequent columnwise normalization, which is not necessary for BCASCs, as the codewords are of unit norm. Note that the coefficients of the two matrices arising from Bernoulli distributions are effectively subject to specific phase restrictions in the complex plane. Nonzero coefficients of the Bernoulli 0/1 matrices are phasors with angles in $\{0, 45°, 90°\}$, while the coefficients of the Bernoulli -1/1 matrices can only take phases in $\{45°, 135°, 225°, 315°\}$. In all simulations the signal size is set to $n = 128$.

A. Eigenvalues of the Restricted Gram Matrices

In this section we present the results of an empirical statistical evaluation of the eigenvalues of the Gram matrices $G_{\Omega_k} = \boldsymbol{A}^*_{\Omega_k}\boldsymbol{A}_{\Omega_k}$ over all feasible values of k for the four matrix types considered. The mean and standard deviation of $\lambda_{\min}(G_{\Omega_k})$ and $\lambda_{\max}(G_{\Omega_k})$ are computed over 10^3 realizations of each matrix type. The support sets Ω_k are randomly generated for each matrix realization. Since BCASCs are unique, all their realizations are identical. Fig. 1 shows the obtained statistics for four values of $\delta = m/n \in \{0.25, 0.50, 0.75, 1\}$. The abscissa is $\rho = s/m$ and it is supposed that s is known and one can set $k = s$.

Gaussian matrices are not a bad option in terms of eigenvalue spread for low δ (Fig. 1a), but its inferiority w.r.t. BCASCs becomes evident as δ increases. For $\delta = 1$ the BCASC is an orthonormal basis of $\mathbb{C}^n$ and $\lambda_{\min}(G_{\Omega_k}) = \lambda_{\max}(G_{\Omega_k}) = 1$, $\forall k$ (Fig. 1d). In contrast, the curves for the Gaussian matrices do not experience any improvement with increasing δ, but a slight degradation. Both Bernoulli alternatives exhibit a large eigenvalue spread, with linear growth of $\lambda_{\max}(G_{\Omega_k})$ with k and a large slope that also grows with δ.

B. Sparse Signal Recovery

The very low standard deviation from the mean values of $\lambda_{\min}(G_{\Omega_k})$ and $\lambda_{\max}(G_{\Omega_k})$ showed in Fig. 1 suggests that the mean values of $\lambda_{\max}(G_{\Omega_k})$ could be directly used in Eq. 12 for the offline calculus of $\alpha_{\lambda_{\max}}$, provided that $k = s$ is known beforehand. We do so and compare the performance of IHT when the step size is computed online via Eq. 11 and offline via Eq. 12. Experiments are carried out for different values of $\delta = m/n$ and $\rho = s/m$, as suggested in [16]. More specifically, we consider a complete evaluation of the entire $\delta - \rho$ plane, i. e., $0 < \delta \leq 1$, $0 < \rho \leq 1$. For each parameter sweep, 32 equally-spaced discrete steps are considered, yielding 1024 different experimental cases. Each experimental case is repeated 256 times with different l_2-normalized s-sparse signals and mean and standard deviation values of the l_2 recovery error are recorded. The algorithm is forced to run 10^3 iterations in all cases. The resulting Donoho-Tanner graphs of mean (normalized) l_2 recovery error are given in Fig. 2, both for Gaussian matrices and BCASCs.

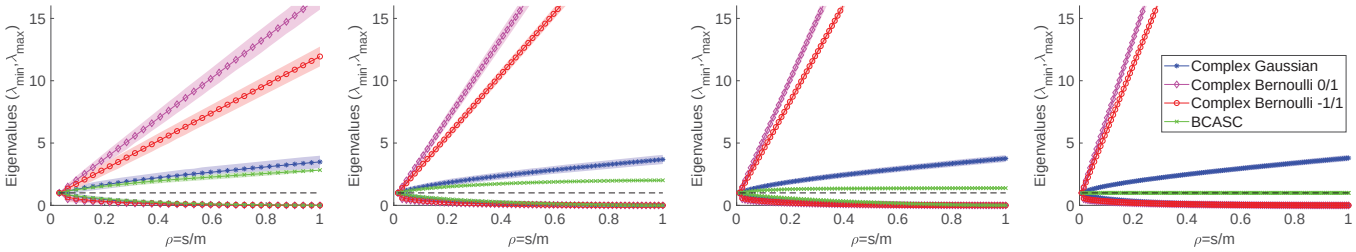
The interpretation of the graphs in Fig. 2 is twofold: on the one hand, IHT performs better with a BCASC than with a Gaussian matrix as measurement matrix, regardless of how α is adjusted. On the other hand, when using BCASCs one can precompute an optimal value of α as prescribed in Eq. 12 without degrading the phase transition in the Donoho-Tanner graphs (cf. Fig. 2d to Fig. 2c). This is not the case if Gaussian matrices are used instead (cf. Fig. 2b to Fig. 2a).

Obviously, using a constant step size cannot be any better than calculating the optimal one at each iteration. For this reason, the advantage of using BCASCs together with the step size given by Eq. 12 is not an improvement of the phase transition in the Donoho-Tanner graphs, but a massive speedup without degrading it. This speedup is twofold: on the one hand, the fact that the maximum eigenvalue of G_{Ω_k} provides a rather tight upper bound on the l_2 norm distortion allows using it to compute a tight Lipschitz constant, which can be used to calculate the maximum allowable step size in the correction step, yielding the fastest convergence. On the other hand, using a constant precomputed step avoids the computational cost of adjusting it at each IHT iteration. In order to give specific speedup figures, let's compare the performance of IHT both using BCASCs and the constant step size from Eq. 12 (our approach) and using Gaussian matrices with the optimal step size from Eq. 11 (standard approach). Suppose we require a (normalized) residual norm of $< 10^{-4}$, which means an SNR of $80\,\mathrm{dB}$ in measurement space. For the case of, e. g., $\delta = 0.75$, $\rho = 0.25$, the standard approach requires 31 iterations to meet the residual norm requirement, while ours meets the requirement after only 17 iterations, i. e., 45% less. Furthermore, the iterations of our approach are faster than those of the standard one. For the selected case the time cost per iteration is 38% lower. Composing both figures we have that the proposed approach can reach the same results as the standard one with a recovery time reduction of 83%.

V. Conclusion

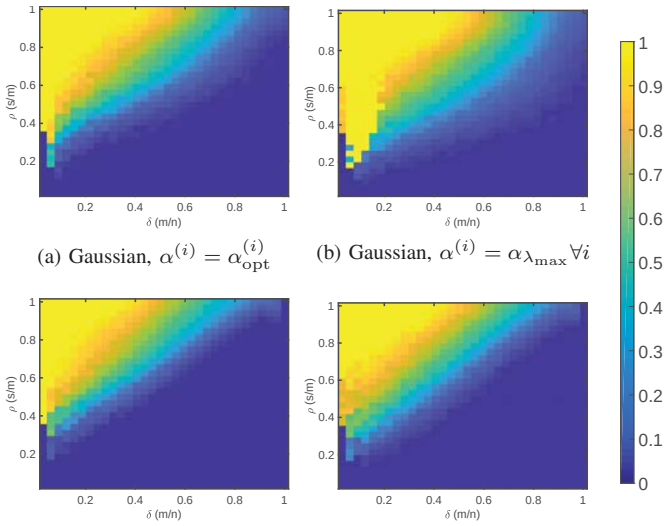
The RIP plays a central role in CS. Fulfilling the RIP with tighter restricted isometry constants translates into the ability of reconstructing less sparse signals from the same number of measurements or, complementary, reducing the minimum number of measurements that are necessary to reconstruct a signal of given sparsity. We have shown how designing the *best* measurement matrices in terms of RIP translates into the problem of finding a set of vectors with minimal coherence. We have also recalled how coherence reduction directly translates into lower restricted isometry constants. We have pointed out the existence of methods for constructing such optimal CS measurement matrices both in the real and complex case and proposed using them to maximally speed up the IHT algorithm.

For different types of measurement matrices, including complex Gaussian matrices and BCASCs, we have proposed using expected values of the maximum eigenvalue of the Ω_k-restricted Gram matrix $G_{\Omega_k} = \boldsymbol{A}^*_{\Omega_k}\boldsymbol{A}_{\Omega_k}$ to analytically compute a constant step size for IHT. Due to the minimal



(a) $\delta = 0.25$ (b) $\delta = 0.50$ (c) $\delta = 0.75$ (d) $\delta = 1$

Fig. 1: Solid lines: mean values of the minimum and maximum eigenvalues of the Gram matrix of the Ω_k-restricted measurement matrices, for four different classes of matrices (see legends in (d)) and $k = s$, against $\rho = s/m$, for all feasible values of s. Shaded areas: 6σ-regions, i. e., only 3.4 cases per million fall outside them. The statistics were computed over 10^3 realizations of each random matrix and randomly-generated Ω_k for each k. Recall that $\delta = m/n$ simply regulates the matrix size.



(a) Gaussian, $\alpha^{(i)} = \alpha^{(i)}_{\text{opt}}$ (b) Gaussian, $\alpha^{(i)} = \alpha_{\lambda_{\max}} \forall i$

(c) BCASC, $\alpha^{(i)} = \alpha^{(i)}_{\text{opt}}$ (d) BCASC, $\alpha^{(i)} = \alpha_{\lambda_{\max}} \forall i$

Fig. 2: Evaluation of the (normalized) l_2 recovery error with respect to $\delta = m/n$ and $\rho = s/m$ obtained after 10^3 IHT iterations. Each pixel in the plots encodes the mean value over 256 independent signal realizations. Plots in the first row are for complex Gaussian matrices, both using optimal (online adapted) and fixed (based on the expected maximum eigenvalues of the partial Gram matrix) IHT step sizes, while plots in the second row are for BCASCs, also in both cases.

eigenvalue spread of the Gram matrices of BCASCs the precomputed step size is approximately the largest possible. Simulations have shown that using the constant steps calculated this way did not yield a visible degradation of the phase transition in the Donoho-Tanner graphs of l_2 recovery error for the case of BCASCs, while allowing for a massive speedup of the algorithm. This speed up arises both from eliminating the online adjustment of the step size and from the fact that using BCASCs yields the fastest convergence of IHT. Typical recovery time reductions by a factor of 6 have been registered w.r.t. using Gaussian matrix with online step size adjustment.

REFERENCES

[1] E. J. Candès, "Compressive sampling," in *Proceedings oh the International Congress of Mathematicians*, Aug. 2006, pp. 1433–1452.

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

[3] 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, Dec. 2006.

[4] ——, "Decoding by linear programming," *IEEE Trans. Inf. Theor.*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.

[5] H. Zörlein and M. Bossert, "Coherence optimization and best complex antipodal spherical codes," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6606–6615, Dec. 2015.

[6] D. E. Lazić, H. Zörlein, and M. Bossert, "Low coherence sensing matrices based on best spherical codes," in *Systems, Communication and Coding (SCC), Proceedings of 2013 9th International ITG Conference on*, Jan. 2013, pp. 1–6.

[7] M. Heredia Conde and O. Loffeld, "Fast approximate construction of best complex antipodal spherical codes," *CoRR*, vol. abs/1705.03280, May 2017.

[8] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[9] R. Baraniuk, "Compressive sensing [lecture notes]," *Signal Processing Magazine, IEEE*, vol. 24, no. 4, pp. 118–121, Jul. 2007.

[10] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comptes Rendus Mathematique*, vol. 346, no. 9-10, pp. 589 – 592, 2008.

[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] E. J. Candès, Y. C. Eldar, D. Needell, and P. Randall, "Compressed sensing with coherent and redundant dictionaries," *Applied and Computational Harmonic Analysis*, vol. 31, no. 1, pp. 59 – 73, 2011.

[13] S. A. Geršgorin, "Über die abgrenzung der eigenwerte einer matrix," *Bulletin de l'Académie des Sciences de l'URSS. Classe des Sciences Mathématiques et na*, vol. 6, pp. 749–754, 1931.

[14] T. Blumensath and M. E. Davies, "Iterative thresholding for sparse approximations," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 629–654, 2008.

[15] J. Portilla, "Image restoration through l0 analysis-based sparse optimization in tight frames," in *2009 16th IEEE International Conference on Image Processing (ICIP)*, Nov. 2009, pp. 3909–3912.

[16] D. Donoho and J. Tanner, "Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing," *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 367, no. 1906, pp. 4273–4293, 2009.