Compressive Deconvolution in Random Mask Imaging
Sohail Bahmani and Justin Romberg, Senior Member

Abstract

We investigate the problem of reconstructing signals and images from a subsampled convolution of masked snapshots and a known filter. The problem is studied in the context of coded imaging systems, where the diversity provided by the random masks makes the deconvolution problem significantly better conditioned than it would be from a set of direct measurements. We start by studying the conditioning of the forward linear measurement operator that describes the system in terms of the number of masks $K$, the dimension of image $L$, the number of sensors $N$, and certain characteristics of the blur kernel. We show that stable deconvolution is possible when $KN \geq L \log L$, meaning that the total number of sensor measurements is within a logarithmic factor of the image size. Next, we consider the scenario where the target image is known to be sparse. We show that under mild conditions on the blurring kernel, the linear system is a restricted isometry when the number of masks is within a logarithmic factor of the number of active components, making the image recoverable using any one of a number of sparse recovery techniques.

I. INTRODUCTION

In this paper, we study deconvolution methods in the context of imaging systems that utilize masks of random codes as depicted in Figure 1. The target image, on the right side of the diagram, is first focused on a Digital Micromirror Device (DMD). The pattern generated by the DMD creates a masked reflection of the image. The reflection from the DMD is then blurred by a secondary lens, and the result is spatially subsampled by a small number of sensors. The diversity introduced by the DMD masks makes reconstruction from these measurements possible; our goal in this paper is to find conditions the number of masks $K$ and sensors $N$ sufficient to stably reconstruct an $L$ pixel image.

In Section II-A, we study the stability of the standard least squares reconstruction by analyzing the behavior of the extreme singular values of the linear measurement operator in terms of properties of the known blurring kernel. Our results suggest that the conditioning of the measurement operator, as can be expected, is related to the degree to which the blurring kernel distributes the energy in the image evenly across the sensors; the architecture performs its best when the net sensitivity across all of the sensors is roughly the same for each pixel in the image.

The number of masks required to attain the mentioned conditioning is quantified by the product of a term that characterizes the conditioning of the subsampled blurring operator and the logarithm of the image dimension.

The authors are with the School of Electrical and Computer Engineering, Georgia Institute of Technology in Atlanta, GA. E-mail: {sohail.bahmani,jrom}@ece.gatech.edu. This work was supported by ONR grant N00014-11-1-0459, and NSF grants CCF-1415498 and CCF-1422540.
Next, in Section [III-B] we analyze the performance $\ell_1$-minimization for reconstruction of sparse images in terms of the Restricted Isometry Property (RIP) of the measurement matrix. Theoretically, we show that the required number of masks to guarantee a sufficiently small RIP constant is proportional to the sparsity of the image with a logarithmic factor of the ambient dimension. As in the case of unstructured deconvolution mentioned above, the relative sensitivity of measurements to each pixel of the image also affects the required number of measurements and the stability of the $\ell_1$-minimization.

Classical deconvolution techniques can be broadly categorized in two frameworks based on their approaches to regularization of the inverse problem. Methods of the first category, including Wiener filtering and a variety of Bayesian methods, assume some stochastic model for the image or the blurring kernel that is often application specific. Methods of the second category, that are essentially some variants of the least squares, only use the deterministic spatial or spectral structures of the image such as smoothness for regularization. For a comprehensive survey of classic deconvolution methods for image restoration and reconstruction we refer the interested readers to [1] and [2].

In recent years there has been an increasing interest in the application of Compressive Sensing (CS) [3]–[5] in various imaging modalities. The study of practical advantages and challenges of various CS imaging systems can be found in [6]. In particular, in [7] an imaging system called the “single-pixel camera” is proposed where multiple snapshots of a randomly masked image is captured using a single sensor. Using the fact that natural images are often (nearly) sparse in some basis, it is shown in [7] that compressive sensing allows accurate image reconstruction in this single-pixel architecture. The use of random masks in the single-pixel camera is very similar to that in our considered imaging system. The single-pixel camera basically measures the inner product of inner product of the image and any random mask using a single sensor.

In the imaging system considered in this paper, the randomly masked reflections of the image are blurred before being spatially subsampled by a small number of sensors. Our analysis ties the number of masks sufficient for reconstruction of the image to certain characteristics of the blurring kernel and the subsampling operator. Because the integration that occurs at each sensor involves the convolution with the Point Spread Function (PSF) of the lens, our analysis has similarity with the analyses used for CS with random convolution [8]–[10]. These analyses address the problems where convolution with a known random signal/filter is of interest. In our problem, however, the convolution is deterministic and randomness is used for modulation only. Our theoretical results are based on recent theoretical developments regarding random matrices and chaos processes.

Section [II] elaborates on the considered imaging architecture and the formulation of the deconvolution problems. The main theoretical guarantees are stated in Section [III] The proofs are subsumed to the appendices. To further validate the theoretical guarantees, the result of some numerical simulations are also provided in Section [IV].

**Notation:** We use the following notation convention throughout this paper. Matrices and vectors are denoted by bold capital and small letters, respectively. The vectors $e_i$ for $i = 1, 2, \ldots$ denote the canonical basis vectors that are zero except at their $i$-th entry which is one. The diagonal matrix whose diagonal entries form a vector $x$ is denoted by $D_x$. The matrix of diagonal entries of a matrix $X$ is denoted by $\text{diag}(X)$. The vector norms $\|\cdot\|_p$
Fig. 1: Schematic of the masked imaging system. The reflection of the target image from a DMD with a random pattern is blurred by a lens and then subsampled by a few sensors. The result is one set of measurements that correspond to the chosen DMD pattern. Our main results are conditions the number of masks $K$ and sensors $N$ (the length of $y_k$ and the number of rows in $G$) that guarantee the stable recovery of an $L$ pixel image (length of $x$ and number of rows/columns in $D_{e_k}$ and $H$).

for $p \geq 1$ are the standard $\ell_p$-norms. The so called $\ell_0$-norm, which counts the nonzero entries of its argument, is denoted by $\|\cdot\|_0$. The matrix norms $\|\cdot\|$ and $\|\cdot\|_F$ are used to denote the operator norm and the Frobenius norm, respectively. The largest (or the smallest) eigenvalue of symmetric matrices are denoted by $\lambda_{\text{max}}(\cdot)$ (or $\lambda_{\text{min}}(\cdot)$). Also, we use $\text{cond}(\cdot)$ to denote the condition number of matrices. Occasionally, expressions of the form $f \gtrless g$ (or $f \lesssim g$) are used that should be interpreted as $f \geq cg$ (or $cf \leq g$), where $c > 0$ is some absolute constant.

II. PROBLEM SETUP

While real images and blurring kernels have 2D representations, in this paper we focus on deconvolution of 1D signals for the sake of an uncluttered exposition. We emphasize, however, that 2D models can be treated in the same manner and the theoretical analysis remains fundamentally the same. As a proof of concept, the results of sparse deconvolution on a test 2D image are also reported in Section IV.

Figure 1 provides a schematic illustration of the considered imaging system. We model the image of interest by a vector $x^* \in \mathbb{R}^L$. We will model the DMD patterns as a vector of independent Rademacher random variables ($\phi_k \in \{\pm 1\}^L$) — we can move from this type of measurement to the more physical model of $\{0,1\}$ random

\footnote{The target image is an image of the Milky Way and the depicted sensor is a CMOS image sensor, both adapted from NASA images.}
variables by simply including a mask of all ones. The effect of the blurring lens and the subsequent subsampling at the sensor is modeled as multiplication by matrices $H \in \mathbb{R}^{L \times L}$ and $G \in \mathbb{R}^{N \times L}$, respectively. Therefore, each of the $K$ measurement $y_k \in \mathbb{R}^N$ for $k = 1, 2, \ldots, K$ can be expressed as

$$y_k = GHD\phi_k x^*.$$  

Approximating the blur by a circular (or linear) convolution, imposes a circulant (or Toeplitz) structure on $H$. Throughout this paper $H \in \mathbb{R}^{L \times L}$ is assumed to be a circulant matrix generated from the blurring kernel $h \in \mathbb{R}^L$. With a slight modification of our treatment, the results can be extended to linear convolution (i.e., Toeplitz $H$).

There are many options for the subsampling matrix $G$. For example, a uniform instantaneous sampling can be model with a uniformly subsampled rows of the identity matrix as $G$. A uniform subsampling with a windowed integration can be modeled by a $G$ whose rows are different shifts of the considered window. In fact, the latter subsampling scheme can be reduced to the former by absorbing the window into the blurring operator $H$. Throughout the paper, we do not impose any specific model on the structure of $G$. However, our results depends on certain characteristics of the subsampled blurring operator (i.e., $GH$) which describes the imaging system.

In this paper, we address the problem of deconvolution in the described system where the blurring kernel is known. In particular, we address the following problems.

1) **Deconvolution of generic images**: In this scenario, the aim is to estimate the image $x$ that has no specific structure. For this deconvolution problem we analyze the performance of the least squares estimator

$$\hat{x} := \arg \min_x \sum_{k=1}^K \|GHD\phi_k x - y_k\|_2^2. \tag{1}$$

The performance of the above least squares depends solely on the conditioning of the measurement matrix. We show that this matrix is well-conditioned if the number of masks is sufficiently larger than a quantity that is directly proportional to the sensitivity to each pixel of the image and the redundancy in the measurements, times the logarithm of the dimension.

2) **Deconvolution of sparse images**: Our major contribution is with regard to recovery of sparse images in the described imaging system. We consider the estimator

$$\hat{x} := \arg \min_x \|x\|_1 \tag{2}$$

subject to $GHD\phi_k x = y_k$ for $k = 1, 2, \ldots, K$,

which is inspired by the compressed sensing framework. A common sufficient condition used in compressed sensing to guarantee accuracy of not only the $\ell_1$-minimization algorithm, but also a variety of greedy algorithms, is the Restricted Isometry Property (RIP) (see [11] and references therein). A matrix $A$ is said to satisfy the RIP with constant $\delta_s \in (0, 1)$ if

$$(1 - \delta_s) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_s) \|x\|_2^2,$$

holds for all $s$-sparse vectors $x$. Exact recovery of $s$-sparse signals via $\ell_1$-minimization is shown in [12] under the condition $\delta_{2s} < \sqrt{2} - 1$, which was later improved to $\delta_{2s} < 3/(4 + \sqrt{6})$ in [13].
We provide a sufficient condition on the number of masks that can guarantee RIP for the measurement matrix associated with (2). Our result shows that the image can be recovered through (2), if the number of masks grows linearly in the sparsity of the target image and logarithmically in its dimension with a constant factor that quantifies the relative sensitivity of the measurements to different pixels.

III. MAIN RESULTS

To state the main results we use the shorthand

$$\rho := \|GH\|, \quad \theta_{\text{max}} := \max_{i=1,2,...,L} \|GHe_i\|_2, \quad \theta_{\text{min}} := \min_{i=1,2,...,L} \|GHe_i\|_2,$$

for the spectral norm, the largest column $\ell_2$-norm, and the smallest column $\ell_2$-norm of $GH$, respectively.

The accuracy and robustness of the solution obtained by the least squares estimator in (1), depends solely on the condition number of the matrix $M_K^T M_K$ where

$$M_K = \left[ D_{\phi_1} H^T G^T \quad D_{\phi_2} H^T G^T \quad \cdots \quad D_{\phi_K} H^T G^T \right]^T.$$

Our Theorem 1 in the next section establishes a bound on the mentioned condition number.

To ensure that the imaging system described in Section II produces identifiable measurements for every image, it is necessary for the system to retain the relative intensity of the image pixels in the measurements to a great extent. For instance, no single-pixel image should be mapped to an all-zero measurement (i.e., $\theta_{\text{min}} > 0$). Similarly, there should be no drastic amplification of any pixel with respect to other pixels. These requirements translate into desirability of matrices $GH$ whose columns $\ell_2$-norms are almost equal (i.e. $\frac{\theta_{\text{max}}}{\theta_{\text{min}}} \approx 1$). In fact, straightforward calculations reveal that the expectation of the matrix $M_K^T M_K$, which describes the system, has a condition number equal to $(\frac{\theta_{\text{max}}}{\theta_{\text{min}}})^2$.

A. Deconvolution by Least Squares

As discussed above, identifiability of the target image in the considered imaging system depends on certain characteristics of the matrix $M_K^T M_K$. The conditioning of the same matrix determines the stability of the estimate obtained by (1). To verify this fact, it suffices to observe that the least squares solution can be written as

$$\hat{x} = (M_K^T M_K)^{-1} M_K^T (y+z) = x^* + (M_K^T M_K)^{-1} M_K^T z,$$

where $y = M_K x^*$ is the vertical concatenation the vectors $y_k$ for $k = 1, 2, \ldots, K$ and $z$ is a measurement perturbation. It is clear from the equation above that the conditioning of $(M_K^T M_K)^{-1} M_K^T$, or equivalently that of $M_K^T M_K$, determines the stability against the perturbation. The following theorem establishes a relation between the number of applied masks (i.e., $K$) and the condition number of $M_K^T M_K$.

**Theorem 1** (Conditioning of the least squares for deconvolution). Suppose that

$$K \geq \frac{\beta + 1}{\log 4 - 1} \delta^{-2} \frac{\rho^2}{\theta_{\text{min}}^2} \log L,$$

(3)
for absolute constants $\delta \in [0, 1]$ and $\beta > 0$. Then with probability at least $1 - 2L^{-\beta}$ we have

$$\text{cond} \left( M_K^T M_K \right) \leq \frac{1 + \delta}{1 - \delta} \frac{\theta_{\max}^2}{\theta_{\min}^2}.$$  

**Remark 1.** As the proof provided in the appendix shows, the bound in (3) can be improved slightly to

$$K \geq (\beta + 1) \max \left\{ \frac{\rho^2}{\psi(-\delta) \theta_{\min}^2}, \frac{\rho^2}{\psi(\delta) \theta_{\max}^2} \right\},$$

where $\psi(t) := (1 + t) \log (1 + t) - t$. The bound (3) is preferred merely because of its simpler expression.

**Remark 2.** It is worthwhile to inspect the tightness of the bound imposed by (3) qualitatively. We can express the term $\rho^2/\theta_{\min}^2$ in (3) as the product of $\|GH\|_F^2/\theta_{\min}^2$ and $\rho^2/\|GH\|_F^2$. As discussed above, for a fixed number of masks the more uneven the column norms of $GH$ are, the more unbalanced the pixel amplification/attenuation and thereby the more unstable the least squares become. The first term (i.e., $\|GH\|_F^2/\theta_{\min}^2$) that varies in the interval $[L, \infty)$, measures how equally the energy of $GH$ is distributed among its columns. Another crucial factor that determines the required number of masks is the redundancy of the measurements that can be captured by the rank of the matrix $GH$. A severely rank-deficient $GH$ provides less information per mask than a full-rank $GH$.

The second term mentioned above (i.e., $\rho^2/\|GH\|_F^2$) that varies in the interval $[1/rank(GH), 1]$ quantifies the dependence on the rank of $GH$. In the ideal case of a full-rank $GH$ with equally normed columns (3) imposes $K \gtrsim L \log L$, which is suboptimal merely by a factor of $\log L$. In the case that $GH$ has columns of equal norm, but $\text{rank}(GH) = 1$, (3) requires $K \gtrsim L \log L$ that is also suboptimal by a mere factor of $\log L$.

**B. Sparse Deconvolution by $\ell_1$-Minimization**

It is also desirable to ensure that the $\ell_1$-minimization (2) is strongly stable and its performance degrades gracefully with perturbations. Significant shrinkage or expansion of the distances between different sparse signals under the measurement matrix can lead to ambiguity or noise amplification, respectively. The RIP provides a formal characterization of the measurement matrices that have the desired properties. The next theorem establishes a RIP for a properly scaled version of the matrix $M_K$.

Below, the set of $s$-sparse vectors inside the $L$-dimensional unit ball is denoted by

$$\mathcal{D}_{s,L} := \left\{ x \in \mathbb{R}^L \mid \|x\|_2 = 1 \text{ and } \|x\|_0 \leq s \right\}.$$

**Theorem 2** (RIP for sparse deconvolution). Let $\mu := \theta_{\max}^2/\theta_{\min}^2 \geq 1$, $\theta_{\text{avg}}^2 := \theta_{\max}^2 + \theta_{\min}^2/2$, and suppose that

$$K \gtrsim \frac{\delta^2}{\mu^2} \mu^2 s \log L,$$

for some parameter $\delta \in (0, 1)$. Then

$$1 - \frac{\mu - 1 + 2\delta}{\mu + 1} \leq \frac{1}{\theta_{\text{avg}}^2} \frac{\|M_K x\|_2^2}{K} \leq 1 + \frac{\mu - 1 + 2\delta}{\mu + 1}$$

(4)

holds for all $x \in \mathcal{D}_{s,L}$ with probability at least $1 - 2e^{-CK \min\{\delta/(2\mu), \delta^2/(2\mu^2)\}}$, where $C > 0$ is an absolute constants.
Corollary 3. If $\mu$ is sufficiently close to one and $K \gtrsim \mu^2 \|x^*\|_0 \log L$, then there exist $\delta \in (0, 1)$ for which \eqref{eq:ell1} recovers the true $s$-sparse signal $x^*$ with probability at least $1 - 2e^{-CK \min\{\delta/(2\mu), \delta^2/(2\mu)^2\}}$, where $C > 0$ is an absolute constant.

Proof: It is known from the standard compressed sensing literature that if a measurement matrix satisfies the RIP of order $2s$ with a sufficiently small constant, then the associated $\ell_1$-minimization exactly recovers any $s$-sparse signal [11, and references therein]. Note that \eqref{eq:K} guarantees that if $K \gtrsim \delta^2 \|x^*\|_0 \log L$ with probability at least $1 - 2e^{-CK \min\{\delta/(2\mu), \delta^2/(2\mu)^2\}}$ the matrix $\frac{1}{\sqrt{K\theta_{\text{avg}}}} M_K$ obeys the RIP of order $2 \|x^*\|_0$ with a constant no more than $(\mu + 2\delta - 1) / (\mu + 1) = 1 - 2(1 - \delta) / (\mu + 1)$. For a $\mu$ that is sufficiently close to one we can choose a $\delta \in (0, 1)$ for which the desired RIP constant can be achieved and thus the $\ell_1$-minimization \eqref{eq:ell1} successfully recovers $x^*$.

Remark 3. Robustness of the proposed sparse deconvolution to noise can also be established using the standard RIP-based bounds known for $\ell_1$-minimization [11, and references therein], but we do not reproduce them here. Note that, as can be expected, for large values of $\mu$ RIP and thereby robustness cannot be guaranteed.

IV. Numerical Experiments

We conducted a series of Monte Carlo simulations to evaluate the deconvolution methods studied in this paper, numerically. In all of the reported simulation results, $N = 4$ sensors with uniform spacing are considered. Also, as mentioned in Section II the masks are drawn independently each with independent Rademacher entries.

A. Deconvolution by Least Squares

We considered signals of dimension $L = 2048$ and varied the number of masks from $K = 512$ to $K = 2048$ with the steps of size 32. The entries of the signal (i.e., $x^*$) are iid standard normal random variables. For the (blurring) filter (i.e., $h$) we consider an all-pass model and a low-pass model. For the all-pass model the filter is drawn from a $\mathcal{N}(0, I)$ distribution. For the low-pass model, however, the filter is obtained by suppressing the DFT coefficients of a standard normal random vector whose indices are between 128 and $1920 = 2047 - 127$ leading to a random filter of “bandwidth” 127. Given that the stability of the least squares approach can be characterized by the conditioning of the associated matrix, we merely measured the ratio of the smallest and largest singular values of the matrix of interest. For each value of $K$ we ran the simulation 100 times and computed the 10\%, the 50\%, and 90\% quantiles of the condition number.

Figures 2a and 2b show the condition number of the matrix associated with the least squares problem \eqref{eq:LS} in logarithmic scale versus the number of masks $K$ for the considered all-pass and low-pass models, respectively. The first few values of $K$ are cropped out because the matrix was severely ill-conditioned and the condition number reached values in the order of $10^4$. As expected, increasing the number of masks improves the considered condition number. Furthermore, for any given number of masks the condition number for the all-pass model is better that the
low-pass model by a factor of about two. Comparison between the quantiles of the condition number at any given value of $K$ also suggests that the condition number has more fluctuations in the case of the low-pass model.

\subsection*{B. Sparse Deconvolution by $\ell_1$-Minimization}

For this set of simulations we considered signals of dimension $L = 2048$ whose nonzero entries are drawn independently from a standard normal distribution and are located on a support set drawn uniformly at random. However, we varied the number of masks from $K = 8$ to $K = 256$ with the steps of size eight. The (blurring) filter is generated according to the random all-pass and low-pass models described above for the case of unstructured deconvolution. For each value of $K$ the sparsity of the signal is increased as a multiple of eight until the successful recovery rate drops below 1%. Among each of the 100 trials ran for each pair of $K$ and $S$ (i.e., the $\ell_0$-norm of the signal), those that yield a relative error no more than 5\% are counted as successful reconstructions.

Figures 3a and 3b show the phase transition diagram of the estimate produced by (2) in terms of the number of masks $K$ and $S$ for the considered all-pass and low-pass models, respectively. As can be seen from the figures phase transition boundary is almost linear in both models, which is in agreement with the relation between $K$ and $S$ suggested by the Theorem 2. While the phase transition boundary for the low-pass model is slightly lower (worse) than that of the all-pass model, the difference does not appear to be significant.

Furthermore, we applied the $\ell_1$-minimization technique to recover a 2D image from (synthetic) measurements in the considered masked imaging system and the result is depicted in Figure 4. Instead of reconstruction in the canonical basis, in this experiment we considered reconstruction of the image with respect to the 2D-DCT basis.
Fig. 3: The phase transition diagram of $\ell_1$-minimization in terms of $K$, the number of masks, and $S$, the $\ell_0$-norm of the signals of dimension $L = 2048$.

The target image at the top left corner is a fluorescent microscopy image that has $L = 188 \times 256$ pixels with 8-bit per color channel. Although the image appears to be sparse, only about 80% of its pixels are zero-valued. In the 2D-DCT basis, however, the image is more compressible as, for instance, its squared $\ell_1$-norm to $\ell_2$-norm ratio is less than $4428 \approx 0.09L$. The blurring kernel, a $128 \times 128$ PSF generated using the PSFGenerator package [14], is depicted in the bottom left corner of Figure 4. For comparison, the blurred unmasked image is shown at the top center. We applied $K = 50$ random Rademacher masks and subsampled each of the blurred outputs by a rate of $1/11$ in each direction which yields $N = 18 \times 24$ scalar measurements per mask. These 50 measurements are shown at the center of the second row. The overall undersampling rate is $K \times N/L \approx 45\%$. We applied the $\ell_1$-minimization with the same blur kernel for each color channel, independently. The relative error of estimated image obtained by $\ell_1$-minimization, shown at the top right corner of Figure 4, is less than 11%.

**APPENDIX A**

**PROOF OF THEOREM**

To prove Theorem 1, we use a variant of matrix Chernoff tail bounds for sums of random positive-semidefinite matrices due to [15, Corollary 5.2].

\[ \text{http://commons.wikimedia.org/wiki/File:S.cerevisiae_septins.jpg} \]
Fig. 4: Sparse deconvolution applied to fluorescent micrograph of cell outlines (red) and septins (green) in Saccharomyces. The first column shows the original $188 \times 256$ image (top) and the $128 \times 128$ blur kernel (bottom). The second column shows the blurred image (top) for comparison and the 50 masked, blurred, and subsampled measurements (bottom). The last column shows the estimated image (top) and the absolute error (bottom).

Proof of Theorem 1: Let $Z_k = D_{\phi_k} H^T G^T G H D_{\phi_k}$. Our goal is to show that for sufficiently large $K$, the random matrix $\sum_{k=1}^K Z_k = M_K^T M_K$ is well-conditioned with high probability. It is straightforward to verify that $Z_k \succeq 0$ and $\|Z_k\| = \rho^2$. Furthermore, we have

$$E \left[ \sum_{k=1}^K Z_k \right] = K \text{diag} \left( H^T G^T G H \right),$$

from which we can deduce that

$$\lambda_{\min} \left( E \left[ \sum_{k=1}^K Z_k \right] \right) = K \theta_{\min}^2$$

and

$$\lambda_{\max} \left( E \left[ \sum_{k=1}^K Z_k \right] \right) = K \theta_{\max}^2.$$ 

Therefore, with $\psi(\delta) := (1 + \delta) \log (1 + \delta) - \delta$, we can apply the matrix Chernoff bound [15, Corollary 5.2] and obtain

$$P \left( \lambda_{\min} \left( \sum_{k=1}^K Z_k \right) \leq (1 - \delta) K \theta_{\min}^2 \right) \leq L e^{-\frac{K \theta_{\min}^2 \psi(\delta)}{\rho^2}}.$$
for any $\delta \in [0, 1]$ and
\[
P \left( \lambda_{\max} \left( \sum_{k=1}^{K} Z_k \right) \geq (1 + \delta) K \theta_{\max}^2 \right) \leq L e^{-K \theta_{\max}^2 \psi(\delta)},
\]
for any $\delta \geq 0$. In particular, using (3) and the fact that $\psi(\pm \delta) \geq (\log 4 - 1) \delta^2$ for all $\delta \in (0, 1)$, it is straightforward to show that
\[
P \left( \lambda_{\min} \left( \sum_{k=1}^{K} Z_k \right) \leq (1 - \delta) K \theta_{\min}^2 \right) \leq L^{-\beta}
\]
and
\[
P \left( \lambda_{\max} \left( \sum_{k=1}^{K} Z_k \right) \geq (1 + \delta) K \theta_{\max}^2 \right) \leq L^{-\beta}.
\]
Therefore, for values of $K$ that obey (3) we have
\[
\frac{\lambda_{\max} \left( \sum_{k=1}^{K} Z_k \right)}{\lambda_{\min} \left( \sum_{k=1}^{K} Z_k \right)} \leq \frac{1 + \delta}{1 - \delta} \cdot \frac{\theta_{\max}^2}{\theta_{\min}^2},
\]
with probability at least $1 - 2L^{-\beta}$.

**APPENDIX B**

**PROOF OF THEOREM 2**

Let $A_x := GHD_x$ and
\[
\tilde{A}_x = I_{K \times K} \otimes A_x = \begin{bmatrix} A_x & 0 \\ & \ddots \\ & & A_x \end{bmatrix}.
\]
Furthermore, define
\[
A := \left\{ \tilde{A}_x \mid x \in D_{s,L} \right\}.
\]
To state the proof, we also need to define certain quantities as follows. Let
\[
d_F(A) := \sup_{A \in A} \| A \|_F = \sup_{x \in D_{s,L}} \| \tilde{A}_x \|_F
\]
\[
= \sqrt{K} \sup_{x \in D_{s,L}} \| A_x \|_F = \sqrt{K} \theta_{\max}
\]
and
\[
d(A) := \sup_{A \in A} \| A \| = \sup_{x \in D_{s,L}} \| \tilde{A}_x \|
\]
\[
= \sup_{x \in D_{s,L}} \| A_x \| = \theta_{\max}.
\]
Let $\gamma_2(A, \| \cdot \|)$ be the Talagrand’s functional [16] for the set $A$ with respect to the operator norm. It is known that this functional can be bounded from above by the Dudley’s integral as
\[
\gamma_2(A, \| \cdot \|) \leq c_1 \int_0^{d(A)} \sqrt{\log N(A, \| \cdot \|, u)} du,
\]
where \( N(\mathcal{X}, \| \cdot \|, \varepsilon) \) denotes the covering number of a set \( \mathcal{X} \) with respect to \( \varepsilon \)-balls of the norm \( \| \cdot \| \). Proof of Theorem 2 follows from the results of [17]. In particular, we use the following theorem.

**Theorem 4** (Krahmer et al [17, Theorem 1.4]). Let \( A \) be a symmetric set of matrices (i.e., \( A = -A \)) and \( \phi \) be a vector of iid Rademacher random variables. Then for some positive absolute constants \( c_1 \) and \( c_2 \) we have

\[
P \left( \sup_{A \in \mathcal{A}} \left\| A \phi \right\|^2_2 - E \left| A \phi \right|^2 \geq c_1 E + t \right) \leq 2e^{-c_2 \min \left\{ \frac{t^2}{U^2}, \frac{2}{t^2} \right\}},
\]

where \( E = \gamma_2(A, \| \cdot \|) (\gamma_2(A, \| \cdot \|) + d_F(A)) + d_F(A) d(A), U = d^2(A), \) and \( V = d(A) (\gamma_2(A, \| \cdot \|) + d_F(A)) \).

**Proof of Theorem 2:** It can be easily verified that

\[
\tilde{A}_x \begin{bmatrix} \phi_1^T \\ \phi_2^T \\ \cdots \\ \phi_K^T \end{bmatrix}^T = M_K x.
\]

Therefore, the conditioning of \( M_K \) on sparse vectors can be obtained by means of appropriate tail bounds for \( \left\| \tilde{A}_x \phi \right\|^2_2 \) with \( x \in D_{s,L} \) and \( \phi \) being a vector of \( KL \) iid Rademacher random variables. Our goal is to apply Theorem 4 for the set \( A \) defined by (5), to obtain the desired tail bounds. It follows from Theorem 4 that for some positive absolute constants \( c_1 \) and \( c_2 \) we have

\[
\sup_{x \in D_{s,L}} \left\| M_K x \right\|^2_2 \leq \sup_{x \in D_{s,L}} E \left\| \tilde{A}_x \phi \right\|^2_2 + c_1 E + t
\]

\[
= \sup_{x \in D_{s,L}} K \left\| A_x \right\|_F^2 + c_1 E + t
\]

\[
\leq K \theta_{\max}^2 + c_1 E + t,
\]

(9)

and

\[
\inf_{x \in D_{s,L}} \left\| M_K x \right\|^2_2 \geq \inf_{x \in D_{s,L}} E \left\| \tilde{A}_x \phi \right\|^2_2 - c_1 E - t
\]

\[
= \inf_{x \in D_{s,L}} K \left\| A_x \right\|_F^2 - c_1 E - t
\]

\[
\geq K \theta_{\min}^2 - c_1 E - t
\]

(10)

with probability at least \( 1 - 2e^{-c_2 \min \left\{ \frac{t^2}{U^2}, \frac{2}{t^2} \right\}} \). It only remains to properly bound the quantities \( U, V, \) and \( E \) and choose a reasonable value for \( t \).

First we need to bound \( \gamma_2(A, \| \cdot \|) \). Using the special structure of \( \tilde{A}_x \) we deduce that

\[
\left\| \tilde{A}_x - \tilde{A}_x' \right\| = \left\| A_x - A_{x'} \right\|
\]

\[
\leq \|GH\| \|x - x'\|_\infty
\]

\[
= \rho \|x - x'\|_\infty.
\]

Therefore, \( N(A, \| \cdot \|, u) \leq N(D_{s,L}, \rho \| \cdot \|_\infty, u) \). Then a simple volumetric argument yields

\[
N(D_{s,L}, \rho \| \cdot \|_\infty, u) \leq \left( \frac{L}{s} \right)^{\left( 1 + \frac{2\rho}{u} \right)^s}
\]

\[
\leq \left( \frac{Le}{s} \left( 1 + \frac{2\rho}{u} \right) \right)^s.
\]
Thus, using $\delta$ and for sufficiently large absolute constant $\beta_0$, we can write

\[
\gamma_2 (\mathcal{A}, \|\cdot\|) \leq c_0 \int_0^{\theta_{\text{max}}} \sqrt{\log N (\mathcal{A}, \|\cdot\|, u)} du \\
\leq c_0 \int_0^{\theta_{\text{max}}} \sqrt{\log N (\mathcal{D}, \|\cdot\|, u)} du \\
\leq c_0 \int_0^{\theta_{\text{max}}} \sqrt{s \log \frac{Le}{s} + s \log \left(1 + \frac{2\rho}{u}\right)} du \\
\leq c_0 \theta_{\text{max}} \sqrt{s \left(\sqrt{\frac{Le}{s}} + 2 \sqrt{\log \left(1 + \frac{2\rho}{\theta_{\text{max}}}\right)}\right)} \\
\leq \theta_{\text{max}} \sqrt{\beta_0 s \log L},
\]

where the last two inequalities follow from the bounds

\[
\int_0^{\alpha} \sqrt{\log \left(1 + \frac{1}{u}\right)} du \leq 2\alpha \sqrt{\log \left(1 + \frac{1}{\alpha}\right)}
\]

and

\[
\rho = \|G\mathcal{H}\| \leq \|G\mathcal{H}\|_F \leq \theta_{\text{max}} \sqrt{L},
\]

respectively. Then we can write

\[
U = d^2 (\mathcal{A}) = \theta_{\text{max}}^2, \quad V = d (\mathcal{A}) (\gamma_2 (\mathcal{A}, \|\cdot\|) + d_F (\mathcal{A})) \\
\quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \leq \theta_{\text{max}}^2 \left(\sqrt{\beta_0 s \log L + \sqrt{K}}\right),
\]

and

\[
E = \gamma_2 (\mathcal{A}, \|\cdot\|) (\gamma_2 (\mathcal{A}, \|\cdot\|) + d_F (\mathcal{A})) + d_F (\mathcal{A}) d (\mathcal{A}) \\
\quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \quad \leq \theta_{\text{max}}^2 \left(\beta_0 s \log L + \sqrt{\beta_0 K s \log L + \sqrt{K}}\right).
\]

Setting $t = \frac{\delta \theta_{\text{min}}^2}{2} \left(\sqrt{\beta_0 s \log L + \sqrt{K}}\right) \sqrt{K}$ we have

\[
c_1 E + t \leq c_1 \theta_{\text{max}}^2 \left(\beta_0 s \log L + \sqrt{\beta_0 K s \log L + \sqrt{K}}\right) \\
+ \frac{\delta \theta_{\text{min}}^2}{2} \left(\sqrt{\beta_0 s \log L + \sqrt{K}}\right) \sqrt{K}
\]

(11)

Recall from the statement of the theorem that $\mu = \frac{\theta_{\text{max}}^2}{\theta_{\text{min}}^2}$. If $K$ obeys

\[
K \geq \left(\sqrt{\beta_0 + \frac{2c_1}{\delta} \cdot \mu \left(\sqrt{\beta_0 + 2}\right)}^2 + \frac{4c_1}{\delta} \cdot \mu \beta_0\right) s \log L \\
= O (\delta^{-2} \mu^2 s \log L),
\]
then with straightforward algebra we can upperbound the right-hand side of (11) as
\[ c_1 E + t \leq \delta K \theta^2_{\min}. \] (12)
Furthermore, we have
\[ \frac{t}{U} \geq \frac{\delta}{2\mu} K \] and
\[ \frac{t}{V} \geq \frac{\delta}{2\mu} \sqrt{K} \]
which imply
\[ e^{-c_2 \min\left\{ \frac{t^2}{V^2}, \frac{t^2}{U^2} \right\}} \leq e^{-c_2 K \min\left\{ \frac{\delta^2}{4\mu}, \frac{\delta^2}{4\mu^2} \right\}}. \] (13)
Applying (12) to (9) and (10) and bounding the tail probability using (13) shows that with probability at least
\[ 1 - 2e^{-c_2 K \min\left\{ \frac{\delta}{2(\mu)}, \frac{\delta}{2(\mu)^2} \right\}} \] the inequalities
\[ (1 - \delta) K \theta^2_{\min} \leq \| M_K x \|_2^2 \leq K \left( \theta^2_{\max} + \delta \theta^2_{\min} \right), \]
which are equivalent to (4), hold for all \( x \in D_{s,L} \). 

REFERENCES

[1] M. Banham and A. Katsaggelos, “Digital image restoration,” IEEE Signal Processing Magazine, vol. 14, no. 2, pp. 24–41, Mar. 1997.
[2] R. Puetter, T. Gosnell, and A. Yahil, “Digital image reconstruction: deblurring and denoising,” Annual Review of Astronomy and Astrophysics, vol. 43, no. 1, pp. 139–194, 2005.
[3] 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.
[4] 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.
[5] D. L. Donoho, “Compressed sensing,” IEEE Transactions on Information Theory, vol. 52, no. 4, pp. 1289–1306, 2006.
[6] R. M. Willett, R. F. Marcia, and J. M. Nichols, “Compressed sensing for practical optical imaging systems: a tutorial,” Optical Engineering, vol. 50, no. 7, pp. 072601–1–072601–13, 2011. [Online]. Available: http://dx.doi.org/10.1117/1.3596602
[7] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, “Single-pixel imaging via compressive sampling,” IEEE Signal Processing Magazine, vol. 25, no. 2, pp. 83–91, Mar. 2008.
[8] J. Romberg, “Compressive sensing by random convolution,” SIAM Journal on Imaging Sciences, vol. 2, no. 4, pp. 1098–1128, 2009.
[9] H. Rauhut, Theoretical foundations and numerical methods for sparse recovery, ser. Radon series on computational and applied mathematics. deGruyter, 2010, vol. 9, ch. Compressive sensing and structured random matrices, pp. 1–92.
[10] H. Rauhut, J. Romberg, and J. A. Tropp, “Restricted isometries for partial random circulant matrices,” Applied and Computational Harmonic Analysis, vol. 32, no. 2, pp. 242–254, 2012.
[11] S. Foucart, “Sparse recovery algorithms: sufficient conditions in terms of restricted isometry constants,” in Approximation Theory XIII: San Antonio 2010, ser. Springer Proceedings in Mathematics, vol. 13. San Antonio, TX: Springer New York, 2012, pp. 65–77.
[12] E. J. Candès, “The restricted isometry property and its implications for compressed sensing,” Comptes Rendus Mathématique, vol. 346, no. 910, pp. 589 – 592, 2008.
[13] S. Foucart, “A note on guaranteed sparse recovery via \( \ell_1 \)-minimization,” Applied and Computational Harmonic Analysis, vol. 29, no. 1, pp. 97 – 103, 2010.
[14] H. Kirshner, F. Aguet, D. Sage, and M. Unser, “3-D PSF fitting for fluorescence microscopy: Implementation and localization application,” *Journal of Microscopy*, vol. 249, no. 1, pp. 13–25, January 2013, software available online at: http://bigwww.epfl.ch/algorithms/psfgenerator/.

[15] J. A. Tropp, “User-friendly tail bounds for sums of random matrices,” *Foundations of Computational Mathematics*, Aug. 2011.

[16] M. Talagrand, *The Generic Chaining*, ser. Springer Monographs in Mathematics. Springer, 2005.

[17] F. Krahmer, S. Mendelson, and H. Rauhut, “Suprema of chaos processes and the restricted isometry property,” *Communications on Pure and Applied Mathematics*, Jan. 2014.