Sensing Matrix Design via Capacity Maximization for Block Compressive Sensing Applications

Richard Obermeier and Jose Angel Martinez-Lorenzo, *Senior Member, IEEE*

***Abstract*—It is well-established in the compressive sensing (CS) literature that sensing matrices whose elements are drawn from independent random distributions exhibit enhanced reconstruction capabilities. In many CS applications, such as electromagnetic imaging, practical limitations on the measurement system prevent one from generating sensing matrices in this fashion. Although one can usually randomize the measurements to some degree, these sensing matrices do not achieve the same reconstruction performance as the aforementioned truly random sensing matrices. This paper presents a novel method, based upon capacity maximization, for designing sensing matrices with enhanced block-sparse signal reconstruction capabilities. Additionally, several numerical examples are also included to show how the proposed method enhances reconstruction performance.**

***Index Terms*—Compressive sensing (CS), block compressive sensing, sensing matrix design, nonconvex optimization.**

I. Introduction

A CLASSICAL problem in science and engineering is reconstructing an unknown vector $\mathbf{x} \in \mathbb{C}^N$ from a set of linear measurements $\mathbf{y} = \mathbf{A}\mathbf{x} \in \mathbb{C}^M$. When $M < N$, there exist an infinite number of solutions satisfying $\mathbf{y} = \mathbf{A}\mathbf{x}$ and so regularization techniques need to be employed in order to induce a unique solution. In practice, the regularization term is selected from prior knowledge of the unknown vector. When the vector is known to be sparse, then Compressive Sensing (CS) theory [1]–[3] states that it can be recovered exactly as the solution to a convex and computationally tractable ℓ_1-norm minimization problem, provided that the sensing matrix is "well-behaved" according to a performance metric such as the mutual coherence [4] or the Restricted Isometry Property (RIP) [5].

CS theory also considers the case where the unknown vector is block sparse. When a signal is block sparse, the non-zero values are distributed over $K = N/L$ disjoint blocks of size L. Although block sparse signals can be reconstructed using the standard techniques, such as ℓ_1-norm minimization, that are applied to general sparse signals, specialized techniques based on joint ℓ_2/ℓ_1 minimization have been shown to provide better reconstruction performance [6]–[11]. Unsurprisingly, extensions to the coherence [11] and RIP [9], [10], [12], [13] determine whether or not the block sparse recovery techniques will be successful for a given sensing matrix.

In general, one cannot verify if a given sensing matrix satisfies the RIP or the block-sparse variant. Frequently, researchers resort to random matrix theory in order to generate sensing matrices that satisfy the RIP condition with high probability. Unfortunately, this approach cannot be used in many applications, such as electromagnetic imaging, where the elements of the sensing matrix are constrained by practical limitations. In this paper, we introduce a method based upon capacity maximization for designing sensing matrices with enhanced block sparse signal reconstruction capabilities.

The remainder of this paper is organized as follows. In Section II, we introduce the capacity-based design method and analyze it from two perspectives, one that is deterministic in nature and one that is probabilistic. We also discuss how our method relates to existing work in the literature. In Section III, we describe how the capacity-based design problem can be solved using the method of multipliers [14]. In Section IV, we present results for several design scenarios to demonstrate the effectiveness of the algorithm. Finally, in Section V we conclude the paper.

II. Capacity-based Design Method

A. Deterministic Perspective

Consider the noise-corrupted linear system $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$, where $\mathbf{x} \in \mathbb{C}^N, \mathbf{y}, \mathbf{w} \in \mathbb{C}^M, \mathbf{A} \in \mathbb{C}^{M\times N}$, and $M < N$. It is assumed here that $\mathbf{A}$ has normalized columns. Suppose that the unknown vector is known to be block sparse with block size L and let us denote $\mathbf{P}_k \in \{0,1\}^{L\times N}, k = 1, \ldots, K$ as the binary projection matrix that extracts the elements of $\mathbf{x}$ in the $k-$th block. Note that because $\mathbf{P}_k$ is a projection matrix, $\mathbf{P}_k\mathbf{P}_k^T = \mathbf{I}_{L,L}$, the identity matrix, and $\mathbf{P}_k\mathbf{P}_j^T = \mathbf{0}_{L,L}$, the zero matrix, for $k \neq j$. In order to induce sparsity in the solution vector, one would ideally use a mixed ℓ_2/ℓ_0 objective function, where the ℓ_0-"norm" counts the number of non-zeros. Unfortunately, this problem is NP-hard, and so it cannot be solved even for moderately sized

Manuscript received May 9, 2018; revised August 22, 2018 and October 12, 2018; accepted October 22, 2018. Date of publication November 30, 2018; date of current version February 7, 2019. This work was supported in part by the Department of Energy under Award DE-SC0017614, and in part by the National Science Foundation CAREER program under Award 1653671. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Ali Bilgin. *(Corresponding author: Jose Angel Martinez-Lorenzo.)*

R. Obermeier is with the Electrical and Computer Engineering, Northeastern University, Winthrop, MA 02152 USA (e-mail: obermeier.ri@husky.neu.edu).

J. A. Martinez-Lorenzo is with the Electrical & Computer Engineering Department and MIE, Northeastern University, Boston, MA 02115 USA (e-mail: jmartinez@coe.neu.edu).

Digital Object Identifier 10.1109/TCI.2018.2884291

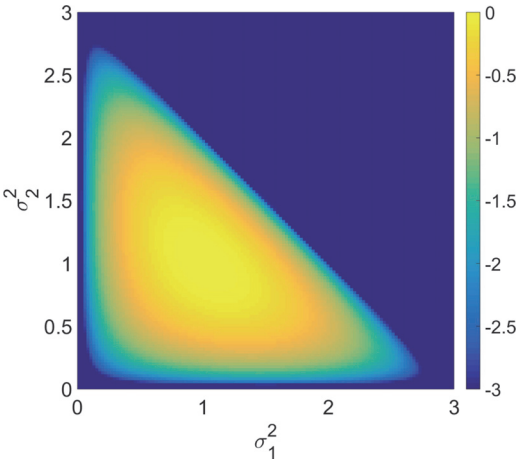

Fig. 1. Capacity of a $M \times 3$ matrix with normalized columns.

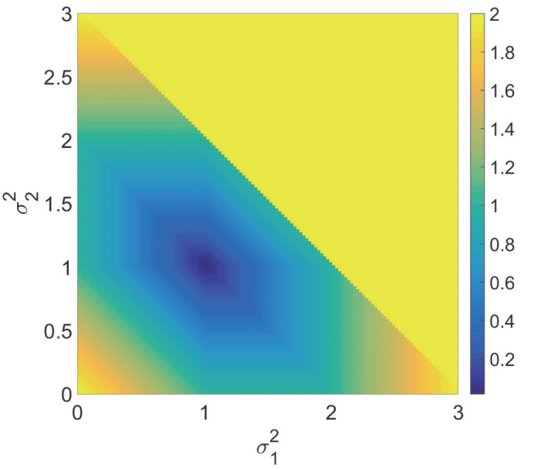

Fig. 2. Restricted isometry constant δ_i of a $M \times 3$ matrix with normalized columns.

problems. However, block CS theory states that the vector can be stably recovered using the following joint ℓ_2/ℓ_1 technique [6]–[11]

$$\begin{aligned} \underset{\boldsymbol{x}}{\text{minimize}} \quad & \sum_{k=1}^{K} \|\mathbf{P}_k \boldsymbol{x}\|_{\ell_2} \\ \text{subject to} \quad & \|\mathbf{A}\boldsymbol{x} - \mathbf{y}\|_{\ell_2} \leq \eta \end{aligned} \tag{1}$$

provided that the sensing matrix $\mathbf{A}$ is "well-behaved" according to some design metric. The most powerful design metric is the block RIP [9], [10], [12], [13], which can be defined as follows. For a fixed block sparsity level T, the block restricted isometry constant $\delta_{L,T}$ is the smallest positive constant such that

$$(1-\delta_{L,T})\|\mathbf{x}_i\|_{\ell_2}^2 \leq \|\mathbf{A}_i\mathbf{x}_i\|_{\ell_2}^2 \leq (1+\delta_{L,T})\|\mathbf{x}_i\|_{\ell_2}^2 \tag{2}$$

where $\mathbf{x}_i = \mathbf{\Phi}_i\mathbf{x}$ and $\mathbf{A}_i = \mathbf{A}\mathbf{\Phi}_i^T$, for all projection matrices $\mathbf{\Phi}_i \in \{0,1\}^{LT\times N}\, i = 1,\ldots,\binom{K}{T}$ obtained by concatenating T of the K projection matrices. Generally speaking, the block RIP requires $\delta_{L,T}$ to be small. When the block RIP is satisfied, it guarantees exact reconstruction in the noiseless measurement scenario and stable reconstruction in the noisy scenario. Note that if $L = 1$, then Eq. (1) reduces to ℓ_1-norm minimization and Eq. (2) reduces to the standard RIP.

Suppose that the sensing matrix can be expressed as a function of $\mathbf{p} \in \mathbb{C}^P$ design variables, i.e., $\mathbf{A} = \mathbf{F}(\mathbf{p})$, where $\mathbf{F} : \mathbb{C}^P \rightarrow \mathbb{C}^{M\times N}$. From the RIP perspective, the optimal design method should determine the design variables $\mathbf{p}$ such that the restricted isometry constant $\delta_{L,T}$ is minimized. There are two issues with this approach. First, it is impractical to compute the block restricted isometry constant $\delta_{L,T}$ for general matrices, as the number of sub-matrices $\binom{K}{T}$ that must be assessed grows exponentially with K. This issue can be addressed by optimizing $\delta_{L,2}$ instead, as it only requires one to assess $\binom{K}{2} = K(K-1)/2$ sub-matrices. In practice, one hopes that this will lead to a decrease in $\delta_{L,T}$, although one can only guarantee that the lower bound for $\delta_{L,T}$ is decreased due to the inequality $\delta_{L,T} \leq \delta_{L,T+1}$. This relaxation is similar to that taken by coherence minimization in standard CS problems [4].

The second issue is that it is impractical (if even possible) to directly optimize the maximum deviation $\delta_i = \max_{m=1,\ldots,LT} |\sigma_{m,i}^2 - 1|$ of a single sub-matrix $\mathbf{A}_i \in \mathbb{C}^{M\times LT}$. Instead of directly optimizing that quantity, we desire a more practical cost function that indirectly decreases the restricted isometry constant. One design metric can be realized through an analysis of the singular values $\sigma_{m,i}$ of the sub-matrix $\mathbf{A}_i$, which has normalized columns. Assuming that $LT \leq M$, the following two relationships are easy to prove:

$$\sum_{m=1}^{LT} \sigma_{m,i}^2 = LT \tag{3}$$

$$\sum_{m=1}^{LT} \log(\sigma_{m,i}^2) \leq 0 \tag{4}$$

where equality in Eq. (4) holds only when all of the singular values equal one. Intuitively, one expects that increasing $\sum_{m=1}^{LT} \log(\sigma_{m,i}^2)$ will decrease δ_i. This result is not guaranteed, but is likely, as can be seen in Figs. 1 and 2, which display the capacity and restricted isometry constant for a $M \times 3$ matrix ($M \geq 3$) with normalized columns. While the level curves of the capacity are relatively smooth, and those of the restricted isometry constant are diamond shaped, the optimal values for each are achieved at the same location. As a result, an appropriately sized step in the direction of the gradient of the capacity is guaranteed to bring the result closer to an optimal solution (i.e., $\sum_{m=1}^{LT} |\sigma_{m,i}^2 - 1|$ gets closer to zero).

Taking this analysis into consideration, we propose solving the following optimization problem to design sensing matrices with enhanced block CS reconstruction capabilities:

$$\begin{aligned} \underset{\boldsymbol{p}}{\text{minimize}} \quad & \max_{r=1,\ldots,R} -\log\det\left(\mathbf{\Phi}_r^T\mathbf{F}^H(\boldsymbol{p})\mathbf{F}(\boldsymbol{p})\mathbf{\Phi}_r + \beta\mathbf{I}_{M_r,M_r}\right) \\ \text{subject to} \quad & \boldsymbol{p} \in Q_p \end{aligned} \tag{5}$$

where $\mathbf{\Phi}_r \in \{0,1\}^{N\times M_r}, r = 1,\ldots,R,\ R = K(K-1)/2,\ \beta$ is a small positive constant that ensures that the arguments to det

are positive-definite, and Q_p is the feasible set for the design variables. For appropriately conditioned problems, β can be made arbitrarily small (we use $\beta = 10^{-6}$), such that it has a minuscule affect on the final solution.

B. Probabilistic Perspective

Now consider the noise-corrupted linear system $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{w}$ from a probabilistic viewpoint. Suppose that $\mathbf{w} \sim \mathcal{N}(0, \mathbf{\Sigma}_w)$ and $\mathbf{x}$ is distributed according to a Gaussian Mixture Model (GMM) such that

$$f_{\mathbf{x}}(\mathbf{x}) = \sum_{r=1}^{R} \pi_r f_{\mathbf{x}}(\mathbf{x}|\theta = r) = \sum_{r=1}^{R} \pi_r \mathcal{N}(0, \mathbf{\Sigma}_r) \tag{6}$$

where θ is a hidden random variable that determines which Gaussian mixture is active. We assume that $\mathbf{x}$ and $\mathbf{w}$ are independent. The success of any reconstruction algorithm in this case is limited by the amount of information that is conveyed through the linear projection $\mathbf{y} = \mathbf{A}\mathbf{x}$:

$$\begin{aligned} I(\mathbf{x};\mathbf{y}) &= I(\mathbf{x};\mathbf{A}\mathbf{x}+\mathbf{w}) \\ &= h(\mathbf{A}\mathbf{x}+\mathbf{w}) - h(\mathbf{A}\mathbf{x}+\mathbf{w}|\mathbf{x}) \end{aligned} \tag{7}$$

where $I(\cdot)$ is the mutual information and $h(\cdot)$ is the entropy rate. It is difficult to evaluate $I(\mathbf{x};\mathbf{y})$ because $\mathbf{x}$ follows a GMM, and so we instead turn our attention to the conditional mutual information, $I(\mathbf{x};\mathbf{A}\mathbf{x}+\mathbf{w}|\theta = r)$, which has the following closed-form solution:

$$\begin{aligned} &I(\mathbf{x};\mathbf{A}\mathbf{x}+\mathbf{w}|\theta = r) \\ &\quad = h(\mathbf{A}\mathbf{x}+\mathbf{w}|\theta = r) - h(\mathbf{A}\mathbf{x}+\mathbf{w}|\mathbf{x},\theta = r) \\ &\quad = \frac{1}{2}\left(\log\det\left(\mathbf{A}\mathbf{\Sigma}_r\mathbf{A}^H + \mathbf{\Sigma}_w\right) - \log\det\left(\mathbf{\Sigma}_w\right)\right) \\ &\quad = \frac{1}{2}\log\det\left(\mathbf{A}\mathbf{\Sigma}_r\mathbf{A}^H\mathbf{\Sigma}_w^{-1} + \mathbf{I}_{M,M}\right) \end{aligned} \tag{8}$$

Eq. (8) can easily be used to compute $I(\mathbf{x};\mathbf{A}\mathbf{x}+\mathbf{w}|\theta) = \sum_{r=1}^{R}\pi_r I(\mathbf{x};\mathbf{A}\mathbf{x}+\mathbf{w}|\theta = r)$, however we are more interested in the worst-case scenario. Therefore, we propose designing the sensing matrix $\mathbf{A}$ such that the worst-case conditional mutual information is maximized. Once again assuming that $\mathbf{A} = \mathbf{F}(\mathbf{p})$, this can be achieved by solving the following optimization problem:

$$\begin{aligned} &\underset{\boldsymbol{p}}{\text{minimize}} \ \max_{r=1,\ldots,R} -\log\det\left(\mathbf{F}(\mathbf{p})\mathbf{\Sigma}_r\mathbf{F}^H(\mathbf{p})\mathbf{\Sigma}_w^{-1} + \mathbf{I}_{M,M}\right) \\ &\text{subject to} \quad \boldsymbol{p} \in Q_p \end{aligned} \tag{9}$$

Block CS problems can be described as a specific version of the GMM problem. In this version, the hidden variable θ defines the support of the variable $\mathbf{x}$, so that the diagonal of $\mathbf{\Sigma}_r$ has LT non-zero values and $N - LT$ values equal to zero, depending on whether or not the element is part of the support for $\theta = r$. The values in the support set can have non-zero off-diagonal elements, or they can be independent of each other. Assuming that $\mathbf{\Sigma}_w = \beta\mathbf{I}_{M,M}$ and $\mathbf{\Sigma}_r = \mathbf{\Phi}_r\mathbf{\Phi}_r^T$, where $\mathbf{\Phi}_r \in \{0,1\}^{N\times M}, r = 1,\ldots,R$, the objective functions of Eq. (5) and 9 differ only by a constant.

C. Comparison With Previous Work

Measurement matrix design is a widely researched area in the CS literature. Many methods have been developed for standard CS problems [15]–[20] and block CS problems [21]–[25] using coherence-based design metrics. These techniques (with the exception of [20]) are limited in that they can only be applied to sensing matrices that are linear projections of a dictionary, i.e., $\mathbf{A} = \mathbf{\Phi}\mathbf{D}$, where the elements of $\mathbf{\Phi}$ are the design variables. Our method, like the one described in [20] for standard CS problems, can be used to design sensing matrices that are nonlinear functions of the design variables, provided that the relationship is differentiable over the feasible set. This property makes our capacity-based method available to a wider range of applications.

There are also many design methods in the literature that are motivated by information theory. Several papers [26]–[29] consider the mutual information between the unknown solution vector and the measurements, $I(\mathbf{x};\mathbf{y})$. As we discussed in the previous sub-section, it is not possible to optimize $I(\mathbf{x};\mathbf{y})$ for block CS problems, at least without making significant approximations to the objective function or its derivative, like is done in [28]. This was the motivation, in the probabilistic viewpoint, for optimizing the conditional mutual information $I(\mathbf{x};\mathbf{y}|\theta = r)$. However, the two quantities are undeniably related. Indeed, the following inequalities hold for the block CS problem:

$$\begin{aligned} \min_{r=1,\ldots,R} I(\mathbf{x};\mathbf{y}|\theta = r) &\le I(\mathbf{x};\mathbf{y}|\theta) \\ &= I(\mathbf{x};\mathbf{y}) + H(\theta|\mathbf{x}) + H(\theta|\mathbf{y}) \\ &\quad - H(\theta|\mathbf{x},\mathbf{y}) - H(\theta) \\ &= I(\mathbf{x};\mathbf{y}) + H(\theta|\mathbf{y}) - H(\theta) \\ &= I(\mathbf{x};\mathbf{y}) - I(\mathbf{y};\theta) \\ &\le I(\mathbf{x};\mathbf{y}) \end{aligned} \tag{10}$$

where we have used the fact that $H(\theta|\mathbf{x}) = H(\theta|\mathbf{x},\mathbf{y}) = 0$ for the block CS problem (the non-zero elements of $\mathbf{x}$ defines its support set, which maps to a unique value of θ). There are also more specialized techniques that adaptively construct the sensing matrix using an greedy information theoretical approach [30] and jointly reconstruct the sensing matrix and the unknowns [31]. Once again, these methods can only be applied to problems where the sensing matrix is a linear function of the design variables.

Finally, we would like to discuss the method presented in [32], which maximizes the mutual information of a nonlinear sensing system $\mathbf{y} = \mathbf{A}\boldsymbol{\psi}(\mathbf{x}) + \mathbf{w}$ by directly optimizing the elements of the matrix $\mathbf{A}$. While this method considers a nonlinear sensing system in which the measurement functions are a linear function of the design variables $\mathbf{A}$, our method considers a linear sensing problem in which the measurement functions are a nonlinear function of the design variables $\mathbf{A} = \mathbf{F}(\mathbf{p})$. So, although both

methods consider nonlinearity, they do so in different ways, and so it is not possible to directly compare them.

D. The Argument for Normalizing the Sensing Matrix

In the most general block CS problems, the only prior information we have about the unknown vector is that it is block sparse, so it may not be appropriate to consider the GMM of Eq. (6). One can instead consider the deterministic viewpoint and the design problem of Eq. (5), which requires the sensing matrix to behave like an approximate isometry. In most problems, one cannot guarantee that this condition is satisfied by the "raw" sensing matrix, as the norms of each column can vary widely. This issue can be overcome by instead considering the normalized sensing matrix $\hat{\mathbf{A}} = \mathbf{A}\,\mathrm{diag}(\mathbf{e})^{-1}$, where the elements of $\mathbf{e} \in \mathbb{R}^N$ are the ℓ_2-norms of each column of $\mathbf{A}$. If $\hat{\mathbf{A}}$ is used, one must also replace $\mathbf{x}$ with the scaled design variables $\hat{\mathbf{x}} = \mathrm{diag}(\mathbf{e})\mathbf{x}$. This is equivalent to minimizing $\sum_{k=1}^{K} \|\mathbf{P}_k\,\mathrm{diag}(\mathbf{e})\mathbf{x}\|_{\ell_2}$ in Eq. (1) instead of $\sum_{k=1}^{K} \|\mathbf{P}_k\mathbf{x}\|_{\ell_2}$. While this scaling does not affect the ℓ_2/ℓ_0 optimization problem that we would ideally solve, it can have a significant effect on the solution generated by the ℓ_2/ℓ_1 problem of Eq. (1). As a result, when the deterministic perspective of Sec. II-A is the primary motivation for the design technique, one should optimize over the normalized sensing matrix $\hat{\mathbf{A}}$. However, if the probabilistic perspective of Sec. II-B is the primary motivation for the design technique, one should optimize over $\mathbf{A}$ instead of $\hat{\mathbf{A}}$, as the solutions to Eq. (9) can greatly vary with the structure of, and especially the trace of the noise covariance $\boldsymbol{\Sigma}_w$.

III. Solving the Design Problem

As we discussed in the previous section, the deterministic design problem of Eq. (5) is a special case of the general design problem of Eq. (9). As a result, we only consider the latter in this section. The general design problem is a nonlinear, nonconvex optimization problem, even when $\mathbf{F}(\mathbf{p})$ is linear. This property makes it difficult to find globally optimal solutions. Nevertheless, in many applications it is sufficient to simply find solutions that are "good enough". This section describes how Eq. (9) can be solved using the method of multipliers [14]. Applying this technique requires a minor modification to the problem. To start, we introduce the auxiliary variable $\mathbf{c} = (c_1, \ldots, c_R)^T \in \mathbb{R}^R$ to represent the capacities of the submatrices, so that the problem be expressed in the equivalent form:

$$\begin{aligned} \underset{\boldsymbol{p}, c_1, \ldots, c_R}{\text{minimize}} \quad & \max_{r=1,\ldots,R} - c_r \\ \text{subject to} \quad & \boldsymbol{p} \in Q_p \\ & c_r = \log\det\left(\mathbf{F}(\boldsymbol{p})\boldsymbol{\Sigma}_r\mathbf{F}^H(\boldsymbol{p})\boldsymbol{\Sigma}_w^{-1} + \mathbf{I}_{M,M}\right) \end{aligned} \tag{11}$$

Eq. (11) has a very similar form to the coherence minimization algorithm displayed in Eq. 11 of [20]; it simply replaces the coherence equality constraints with the capacity equality constraints. As a result, the method of multipliers [14] approach described in [20] can also be used to solve Eq. (11), provided that the feasible set Q_p has an easy to compute proximal operator. Formally, the scaled Augmented Lagrangian can be written as follows:

$$\begin{aligned} \mathcal{L}_{\mathcal{A}}(\boldsymbol{p}, \boldsymbol{c}, \boldsymbol{\gamma}; \rho) = & \left(\max_{r=1,\ldots,R} - c_r\right) + I_{Q_p}(\boldsymbol{p}) + \sum_{r=1}^{R} \frac{\rho}{2} \\ & \left|c_r - \log\det\left(\mathbf{F}(\boldsymbol{p})\boldsymbol{\Sigma}_r\mathbf{F}^H(\boldsymbol{p})\boldsymbol{\Sigma}_w^{-1} + \mathbf{I}_{M,M}\right) + \gamma_r/\rho\right|^2 \end{aligned} \tag{12}$$

where $\boldsymbol{\gamma} \in \mathbb{R}^R$ are the Lagrange multipliers and I_{Q_p} is the indicator function for the feasible set. The method of multipliers solves Eq. (11) by solving a series of unconstrained problems of the form of Eq. (12), where $\boldsymbol{\gamma}$ is held fixed. The unconstrained sub-problems can be solved using an alternating minimization procedure, in which $\mathbf{c}$ is updated by evaluating the proximal operator for $\max$ and $\mathbf{p}$ is updated using a proximal gradient update. This procedure is described in the Appendix. When a given instance of Eq. (12) is solved, the Lagrange multipliers are updated as follows:

$$\begin{aligned} \gamma_r^{(k+1)} = \gamma_r^{(k)} + \rho^{(k)}\Bigg(c_r^{(k)} - \log\det \\ \left(\mathbf{F}(\mathbf{p})\boldsymbol{\Sigma}_r\mathbf{F}^H(\mathbf{p})\boldsymbol{\Sigma}_w^{-1} + \mathbf{I}_{M,M}\right)\Bigg) \end{aligned} \tag{13}$$

where the superscripts indicate the iteration number, i.e., the Lagrange multiplier $\gamma_r^{(k)}$ is used on the $k-$th instance of Eq. (12). To improve the convergence rate of the algorithm, it is often necessary to increase the penalty parameter ρ based upon the convergence level. Our design method utilizes an update approach based upon the one described in [14]. The optimization procedure is summarized in Algorithm 1.

IV. Numerical Results

In this section, we present design results for several sensing problems. All results were generated by solving Eq. (5) using the normalized sensing matrix $\hat{\mathbf{A}}$.

A. Pulse Reconstruction Problem

In the first example, the design algorithm is applied to a pulse reconstruction problem. Consider the scenario in which a discrete time-series signal x_n, where $n = 0, \ldots, N-1$, needs to be reconstructed from a set of incomplete Fourier measurements y_m, where $m = 0, \ldots, M-1$. Formally, the $m-$th measurement can be expressed as follows:

$$y_m = \sum_{n=0}^{N-1} x_n\, e^{-\jmath\hat{\omega}_m\, n} \tag{14}$$

where $\hat{\omega}_m$ is the normalized digital frequency of the $m-$th measurement. The pulses were known to be distributed on $K = 16$ non-overlapping segments of a fixed width $L = 32$ samples. This is a simplified example of a communication network that uses the Time Division Multiple Access (TDMA) channel access method. The objective, then, is to select the normalized digital frequencies $\hat{\omega}_m$ such that the minimum capacities over

Algorithm 1 : Summary of the Augmented Lagrangian Update Procedure for the Capacity Maximization Problem of Eq. (11).

Choose the initial values for $\mathbf{p}^{(0)}$, $\rho^{(1)}$;
Choose convergence rate tolerances η_* and ω_* ;
Choose parameter $\tau > 1$;
Set $\omega^{(1)} = 1/\rho^{(1)}$ and $\eta^{(1)} = 1/(\rho^{(1)})^{0.1}$;
Set $c_r^{(0)} = -\log\det\left(\mathbf{F}(\mathbf{p}^{(0)})\mathbf{\Sigma}_r\mathbf{F}^H(\mathbf{p}^{(0)})\mathbf{\Sigma}_w^{-1} + \mathbf{I}_{M,M}\right)\big)$, $\gamma_r^{(1)} = 0$;
for k = *1,2,3...* **do**
  Approximately solve the unconstrained subproblem

$$\left(\mathbf{p}^{(k)}, \mathbf{c}^{(k)}\right) = \underset{\boldsymbol{p},\boldsymbol{c}}{\operatorname{argmin}}\ \mathcal{L}_{\mathcal{A}}\left(\boldsymbol{p}, \boldsymbol{c}, \boldsymbol{\gamma}^{(k)}; \rho^{(k)}\right)$$

  so that after m iterations

$$\max(\|\mathbf{p}_m^{(k)} - \mathbf{p}_{m-1}^{(k)}\|_{\ell_\infty}, \|\mathbf{c}_m^{(k)} - \mathbf{c}_{m-1}^{(k)}\|_{\ell_\infty}) \le \omega^{(k)}$$

  Test for convergence ;

$$e^{(k)} = \left(\sum_{r=1}^{R}\left[c_r^{(k)} - \log\det\left(\mathbf{F}(\mathbf{p}^{(k)})\mathbf{\Sigma}_r\mathbf{F}^H(\mathbf{p}^{(k)})\mathbf{\Sigma}_w^{-1} + \mathbf{I}_{M,M}\right)\right]^2\right)^{0.5}$$

  if $e^{(k)} \le \eta^{(k)}$ **then**
    if $e^{(k)} \le \eta_*$ and $\max(\|\mathbf{p}^{(k)} - \mathbf{p}^{(k-1)}\|_{\ell_\infty}, \|\mathbf{c}^{(k)} - \mathbf{c}^{(k-1)}\|_{\ell_\infty}) \le \omega_*$ **then**
      stop with solution $\mathbf{p}^{(k)}$

$$\gamma_r^{(k+1)} = \gamma_r^{(k)} + \rho^{(k)}\left(c_r^{(k)} - \log\det\left(\mathbf{F}(\mathbf{p}^{(k)})\mathbf{\Sigma}_r\mathbf{F}^H(\mathbf{p}^{(k)})\mathbf{\Sigma}_w^{-1} + \mathbf{I}_{M,M}\right)\right)$$

    $\rho^{(k+1)} = \rho^{(k)}$;
    $\omega^{(k+1)} = \omega^{(k)}/\rho^{(k+1)}$;
    $\eta^{(k+1)} = \eta^{(k)}/(\rho^{(k+1)})^{0.9}$;
  else
    $\gamma_r^{(k+1)} = \gamma_r^{(k)}$;
    $\rho^{(k+1)} = \tau\rho^{(k)}$;
    $\omega^{(k+1)} = 1/\rho^{(k+1)}$;
    $\eta^{(k+1)} = 1/(\rho^{(k+1)})^{0.1}$;

TABLE I
SUMMARY OF DESIGN PARAMETERS AND CONSTRAINTS FOR THE PULSE RECONSTRUCTION SENSING MATRIX DESIGN PROBLEM

Design parameters and constraints		
Parameter	baseline value	Constraint
M	256	–
N	512	–
K	16	–
L	32	–
$\hat{\omega}_m$	Randomly distributed between $-\pi$ and π	Unbounded

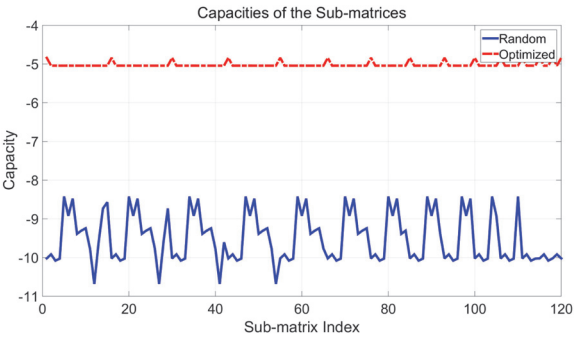

Fig. 3. Capacities of the sub-matrices evaluated during the optimization procedure for the pulse reconstruction problem.

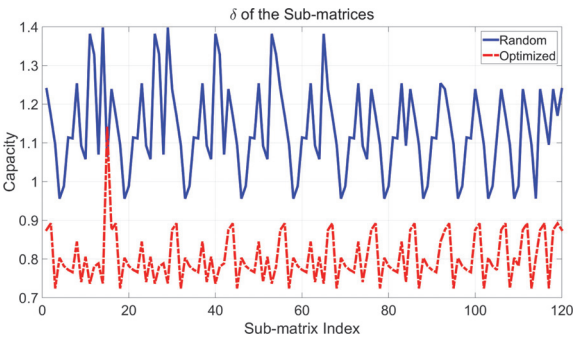

Fig. 4. Bounds on the singular values of the sub-matrices evaluated during the optimization procedure for the pulse reconstruction problem.

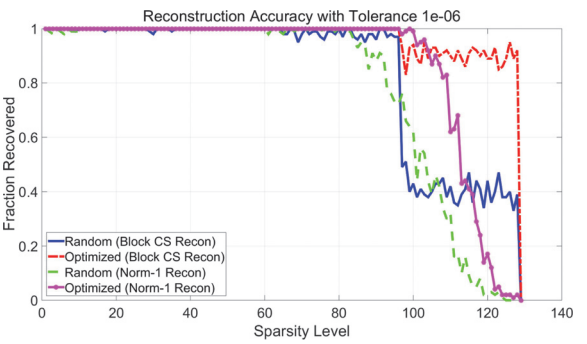

Fig. 5. Numerical comparison of the reconstruction accuracies of joint ℓ_2/ℓ_1 reconstruction and standard ℓ_1 reconstruction, using the baseline random and optimized designs for the pulse reconstruction problem.

all $\binom{16}{2} = 120$ pairs of blocks are maximized according to the design parameters and constraints displayed in Table I. Note that, due to the 2π modulo nature of $\hat{\omega}_m$, it was considered unbounded.

Fig. 3 displays the capacities of all of the sub-matrices evaluated in the optimization procedure, and Fig. 4 displays the bounds on the singular values of the sub-matrices, i.e., $\delta = \max_{m=1,\ldots,LT}|\sigma_{m,i}^2 - 1|$. The optimized sensing matrix outperforms the random matrix according to both metrics, which suggests that it will provide better block CS reconstruction performance. This result is confirmed by Fig. 5 which displays the CS reconstruction accuracies achieved by the baseline and optimized sensing matrices when joint ℓ_2/ℓ_1 and standard ℓ_1 reconstruction techniques are applied to noiseless data. These results were generated by reconstructing 100 vectors at each sparsity level $S = 1, \ldots, M/2$ (block sparsity $S_B = \frac{S}{L}$) and comparing the solutions to the ground truth vectors. The ℓ_1-norm minimization results were included to provide a comparison with the joint ℓ_2/ℓ_1 results. Unsurprisingly, joint ℓ_2/ℓ_1 minimization outperformed ℓ_1 minimization for each of the sensing matrices. Remarkably, the optimized sensing matrix was able to reconstruct $>90\%$ of block-sparse vectors up to a block sparsity $S_B = 4$ (total sparsity $S = 128$) using joint ℓ_2/ℓ_1 minimization, whereas the baseline random sensing matrix reconstructed $< 50\%$. It is important to note that exact reconstruction cannot be guaranteed

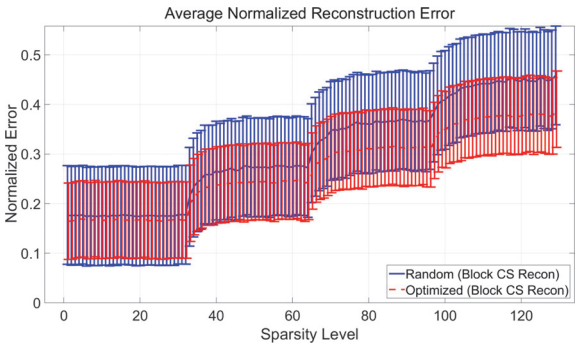

Fig. 6. Numerical comparison of the reconstruction accuracies of joint ℓ_2/ℓ_1 reconstruction and standard ℓ_1 reconstruction, using the baseline random and optimized designs for the pulse reconstruction problem using noisy measurements (SNR = 20 dB). The error bars denote the first standard deviation.

TABLE II
SUMMARY OF DESIGN PARAMETERS AND CONSTRAINTS FOR THE ELECTROMAGNETIC IMAGING SENSING MATRIX DESIGN PROBLEM

Design Parameters and Constraints		
Parameter	Baseline Value	Constraint
M	64	–
N	144	–
K	9	–
L	16	–
$\mathbf{r}_n$	5λ by 5λ grid centered at origin	–
$\mathbf{r}_m$	Uniformly spaced over 5λ by 5λ grid at $z = 5\lambda$	$\lvert x_m\rvert \leq 2.5\lambda$ $\lvert y_m\rvert \leq 2.5\lambda$ $z_m = 5\lambda$

for total sparsity levels greater than $M/2$ (128 for this problem). Although it does not achieve the theoretical limit, the optimized sensing matrix achieves a level of performance that is significantly better than that of the randomized sensing matrix. The improvement in block CS reconstruction performance is also confirmed by Fig. 6, which displays the averaged normalized reconstruction errors ($\|\hat{\mathbf{x}} - \mathbf{x}_{\text{true}}\|_{\ell_2}/\|\mathbf{x}_{\text{true}}\|_{\ell_2}$, with error bars) achieved by the baseline and optimized sensing matrices when joint ℓ_2/ℓ_1 minimization is applied to noisy measurement data ($\text{SNR} = 20\log(\frac{\|\mathbf{A}\mathbf{x}\|_{\ell_2}}{\|\mathbf{w}\|_{\ell_2}}) = 20$ dB).

B. Electromagnetic Imaging Problem

In this example, the design algorithm is applied to a linearized inverse electromagnetic problem, in which multiple monostatic antennas are used to image a region of interest using a single frequency. The simplified discrete measurement model for this system is given as follows [20]:

$$y_m = \sum_{n=0}^{N-1} x_n e^{-j2k\|\mathbf{r}_m - \mathbf{r}_n\|_{\ell_2}} = \sum_{n=0}^{N-1} A_{mn} x_n \qquad (15)$$

where y_m is the $m-$th scattered field measurement, $\mathbf{r}_m$ is the position of the $m-$th antenna, $\mathbf{r}_n$ is the $n-$th position in the imaging region, k is the wavenumber, and x_n is the reflectivity at the $n-$th position in the imaging region. Keeping the wavenumber fixed, the objective in this case is to select the antenna positions $\mathbf{r}_m$.

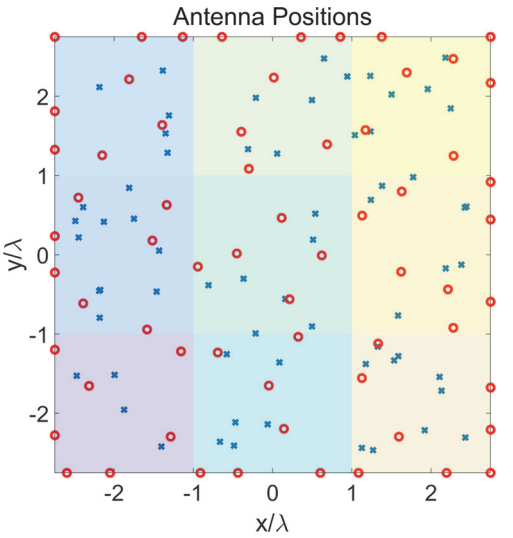

Fig. 7. Antenna positions of the baseline (blue) and optimized (red) designs. The shaded boxes in the background represent the squares on which the capacity was evaluated.

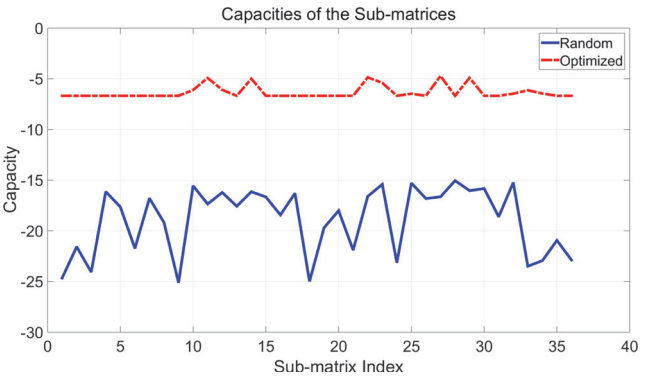

Fig. 8. Capacities of the sub-matrices evaluated during the optimization procedure for the electromagnetic imaging problem.

Table II displays the design parameters and constraints for the optimization problem, and Fig. 7 displays the positions of the baseline random antenna configuration, which was used as the starting point to the optimization procedure, and the positions of the optimized antenna configuration. The colored regions in the background of Fig. 7 represent the nine blocks on which the unknown signal was known to be block-sparse. The optimization procedure was configured to maximize the minimum capacity of all 36 pairs of blocks. Figs. 8 and 9 display the capacities and singular value bounds for the evaluated sub-matrices. Once again, the optimized sensing matrix shows clear improvement over the randomized starting point. This directly led to an improvement in CS reconstruction accuracy, as shown in Figs. 10 and 11, which display noiseless and noisy reconstruction results. In the noiseless case, ℓ_2/ℓ_1 minimization reconstructed $>90\%$ of block-sparse vectorsup to a block sparsity $S_B = 2$ (total sparsity $S = 32$) using the optimized sensing matrix, whereas it only reconstructed $<20\%$ using the random sensing matrix. Once again, the fact that the optimized design performs so well up to the theoretical maximum sparsity level, $M/2 = 32$, truly demonstrates the capabilities of the design method. A specific

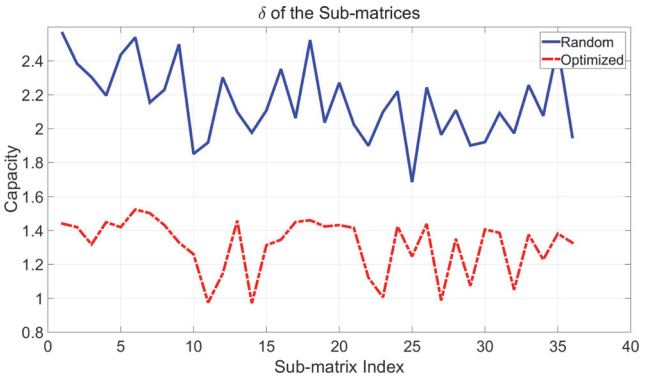

Fig. 9. Bounds on the singular values of the sub-matrices evaluated during the optimization procedure for the electromagnetic imaging problem.

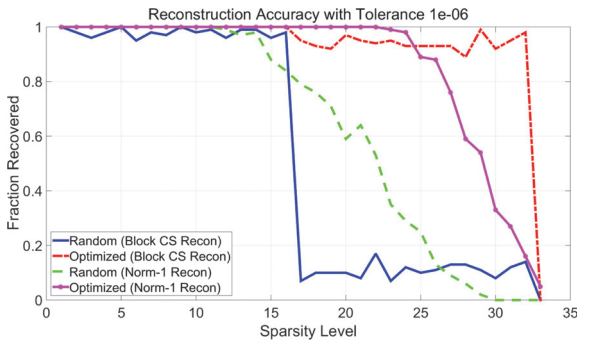

Fig. 10. Numerical comparison of the reconstruction accuracies of joint ℓ_2/ℓ_1 reconstruction and standard ℓ_1 reconstruction using the baseline random and optimized designs for the electromagnetic imaging problem.

Fig. 11. Numerical comparison of the reconstruction accuracies of joint ℓ_2/ℓ_1 reconstruction and standard ℓ_1 reconstruction using the baseline random and optimized designs for the electromagnetic imaging problem using noisy measurements (SNR $=$ 20 dB). The error bars denote the first standard deviation.

instance of the noiseless planar reconstruction problem is displayed in Figs. 12–14, which display the ground-truth reflectivity, the reflectivity reconstructed by the baseline random sensing matrix, and the reflectivity reconstructed by the optimized sensing matrix.

C. General Linear System

In the final example, the design algorithm is tested in a general linear system, $\mathbf{y} = \mathbf{A}\mathbf{x}$. The objective for this problem is to optimize the $M \cdot N$ coefficients $a_{mn} \in \mathbb{C}$ of the sensing matrix.

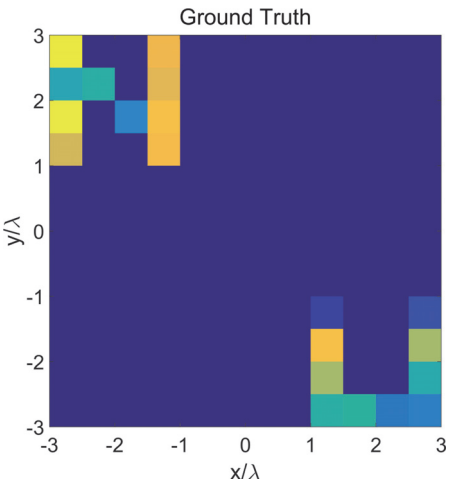

Fig. 12. Magnitude of the ground-truth reflectivity.

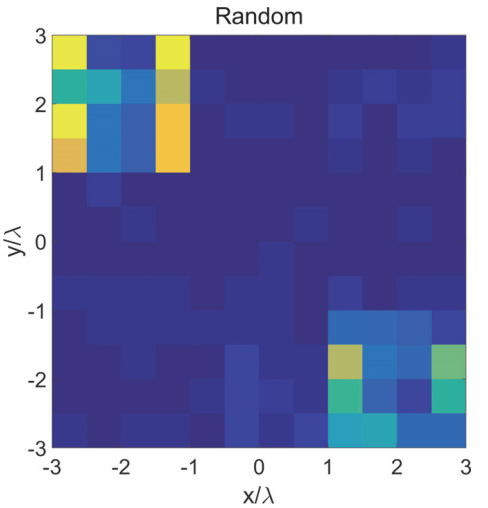

Fig. 13. Magnitude of the reflectivity reconstructed by the baseline random sensing matrix using joint ℓ_2/ℓ_1 minimization. Normalized error $=$ 0.2863.

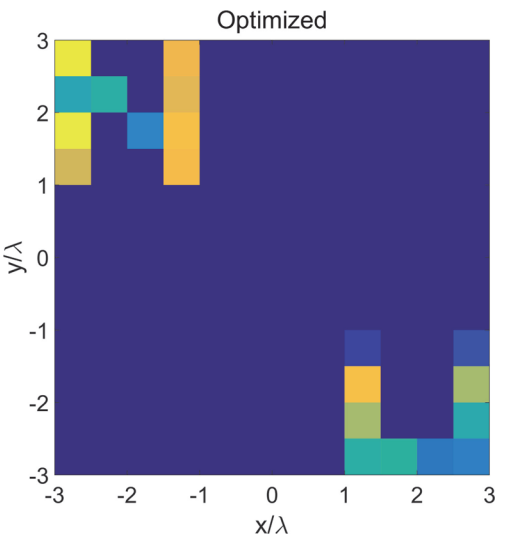

Fig. 14. Magnitude of the reflectivity reconstructed by the baseline random sensing matrix using joint ℓ_2/ℓ_1 minimization. Normalized error $=$ 0.0.

TABLE III
SUMMARY OF DESIGN PARAMETERS AND CONSTRAINTS FOR THE GENERAL LINEAR SYSTEM SENSING MATRIX DESIGN PROBLEM

Design Parameters and Constraints		
Parameter	Baseline Value	Constraint
M	64	–
N	192	–
K	24	–
L	8	–
a_{mn}	Randomly distributed according to i.i.d. complex Normal distribution	Unbounded

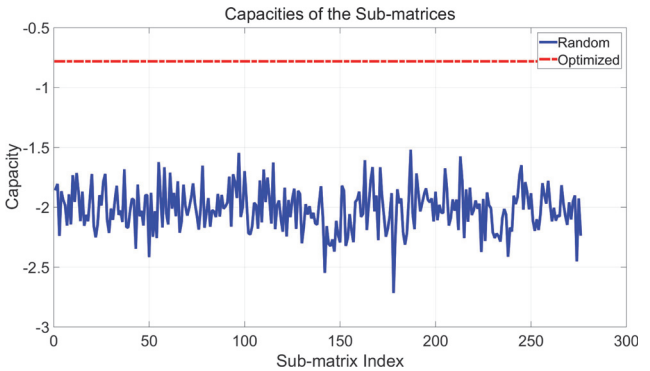

Fig. 15. Capacities of the sub-matrices evaluated during the optimization procedure for the general linear system problem.

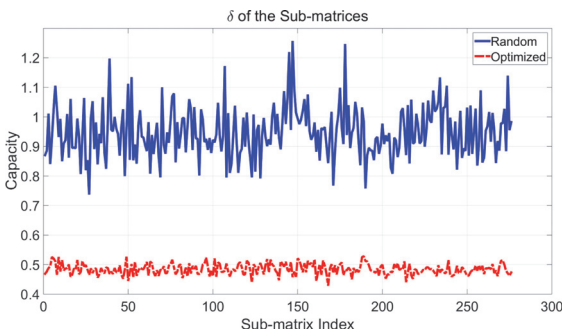

Fig. 16. Bounds on the singular values of the sub-matrices evaluated during the optimization procedure for the general linear system problem.

This is the ideal design scenario, since complete control over the sensing matrix is given. The initial values of the sensing matrix coefficients are drawn from i.i.d. complex Normal distributions. The full set of design parameters and constraints are displayed in Table III.

Figs. 15 and 16 display the capacities and singular value bounds for the evaluated sub-matrices in the linear design problem. While the increase in capacity is smaller than in the previous examples, the singular value bounds improved significantly. In fact, the optimized sensing matrix satisfies the block restricted isometry constant with value $\delta_{8,2} \approx 0.53$. As one would expect, this leads to an improvement in reconstruction accuracy, as shown in Figs. 17 and 18. In the noiseless case, ℓ_2/ℓ_1 minimization reconstructed $>90\%$ of block-sparse vectors up to a block sparsity $S_B = 4$ (total sparsity $S = 32$) using the optimized sensing matrix, whereas it only reconstructed $75-80\%$ using the random sensing matrix. Of the three examples presented in this paper, the reconstruction accuracy is (unsurprisingly) increased the least in the last one. This is due to the significant control that the designer has over the sensing matrix. Therefore, running the optimization procedure in such scenarios may not be worth the effort when compared to the random solution.

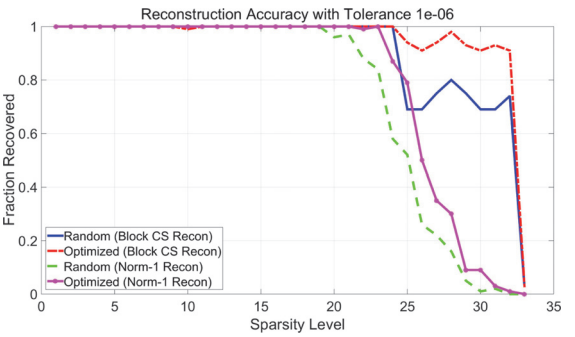

Fig. 17. Numerical comparison of the reconstruction accuracies of joint ℓ_2/ℓ_1 reconstruction and standard ℓ_1 reconstruction using the baseline random and optimized designs for the general linear system problem.

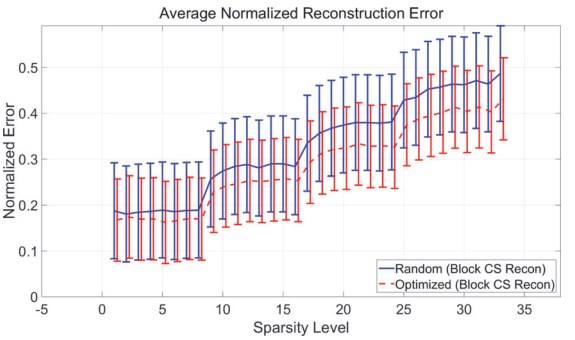

Fig. 18. Numerical comparison of the reconstruction accuracies of joint ℓ_2/ℓ_1 reconstruction and standard ℓ_1 reconstruction using the baseline random and optimized designs for the general linear system problem using noisy measurements (SNR $=$ 20 dB). The error bars denote the first standard deviation.

V. CONCLUSION

This paper describes a novel method for designing sensing matrices with enhanced block-sparse signal recovery capabilities. By maximizing the minimum capacity over a set of sub-matrices selected from columns of the full sensing matrix, the design method is capable of significantly improving the reconstruction results obtained using joint ℓ_2/ℓ_1 minimization. Our technique was motivated from two perspectives, a deterministic one seeking to satisfy the block RIP, and an information theoretical one. The proposed design method is intimately related to many existing techniques, but differentiates itself in that it optimizes worst-case performance and can be applied to problems where the sensing matrix is a nonlinear function of the design variables.

The design method's ability to improve the joint ℓ_2/ℓ_1 minimization reconstruction performance was demonstrated in three applications: a sparse pulse reconstruction problem, an electromagnetic imaging problem, and a general linear system. These results showed that the design method can be extremely beneficial in applications where the measurement system is constrained by practical limitations, but less beneficial when one has greater control over the sensing matrix.

Algorithm 2: Summary of the Alternating Minimization Procedure for Solving the Augmented Lagrangian Subproblem of Eq. (12).

Given $\mathbf{p}_{(0)}^{(k)}$, $\mathbf{c}_{(0)}^{(k)}$, $\boldsymbol{\gamma}^{(k)}$, $\rho^{(k)}$;
for *m = 0,1,2,...* **do**
  Update $\mathbf{c}$ while holding $\mathbf{p}$ fixed

$$\mathbf{c}_{(m+1)}^{(k)} = \underset{\boldsymbol{c}}{\operatorname{argmin}}\ \mathcal{L}_{\mathcal{A}}\left(\mathbf{p}_{(m)}^{(k)}, \boldsymbol{c}, \boldsymbol{\gamma}^{(k)}; \rho^{(k)}\right)$$

  Update $\mathbf{p}$ while holding $\mathbf{u}$ fixed

$$\mathbf{p}_{(m+1)}^{(k)} = \underset{\boldsymbol{p}}{\operatorname{argmin}}\ \mathcal{L}_{\mathcal{A}}\left(\boldsymbol{p}, \mathbf{c}_{(m+1)}^{(k)}, \boldsymbol{\gamma}^{(k)}; \rho^{(k)}\right)$$

Appendix A
Augmented Lagrangian Subproblem

In Section III, we described an Augmented Lagrangian method for solving the sensing matrix design problem. One of the key steps in this procedure solves the unconstrained subproblem of Eq. (12), which is repeated here for convenience:

$$\mathcal{L}_{\mathcal{A}}(\boldsymbol{p}, \boldsymbol{c}, \boldsymbol{\gamma}; \rho) = \left(\max_{r=1,\ldots,R} - c_r\right) + I_{Q_p}(\boldsymbol{p}) + \sum_{r=1}^{R} \frac{\rho}{2} \left|c_r - \log\det\left(\mathbf{F}(\boldsymbol{p})\boldsymbol{\Sigma}_r\mathbf{F}^H(\boldsymbol{p})\boldsymbol{\Sigma}_w^{-1} + \mathbf{I}_{M,M}\right) + \gamma_r/\rho\right|^2 \tag{12}$$

This subproblem can be solved using an alternating minimization procedure, in which $\mathbf{c}$ is updated while $\mathbf{p}$ is held fixed and vice versa. The alternating minimization procedure is summarized in Algorithm 2. Note that the subscript (m) denotes the value of the variable at the m-th iteration of the unconstrained subproblem, while the superscript (k) denotes the value of the variable at the k-th iteration of the outer loop defined in Algorithm 1. Consequently, $\mathbf{c}_{(m)}^{(k)}$ denotes the value of $\mathbf{c}$ at the m-th inner iteration of the k-th outer iteration, and by convention, $\mathbf{c}^{(k)} = \mathbf{c}_{(m_k)}^{(k)}$, where m_k is the last iteration of the k-th subproblem. The $\mathbf{c}$ and $\mathbf{p}$ update steps displayed in Algorithm 2 are described in detail in Appendix B and Appendix C respectively.

Appendix B
$\mathbf{c}$ Update Step

In the $\mathbf{c}$ update step, we seek the minimizer of Eq. (12) with respect to $\mathbf{c}$ while keeping $\mathbf{p}$, ρ, and $\boldsymbol{\gamma}$ constant. By introducing the auxiliary variables $z_r = -\log\det(\mathbf{F}(\mathbf{p})\boldsymbol{\Sigma}_r\mathbf{F}^H(\mathbf{p})\boldsymbol{\Sigma}_w^{-1} + \mathbf{I}_{M,M}) + \gamma_r/\rho$ and $\mathbf{d} = -\mathbf{c}$, this subproblem can be reduced to the proximal operator for $\max$, i.e.:

$$\underset{\boldsymbol{d}}{\text{minimize}} \quad \left(\max_{r=1,\ldots,R} d_r\right) + \frac{\rho}{2}\|\boldsymbol{d} - \mathbf{z}\|_{\ell_2}^2 \tag{16}$$

By introducing the auxiliary variable t, this problem can be recast as follows:

$$\begin{aligned} \underset{\boldsymbol{d},t}{\text{minimize}} \quad & t + \frac{\rho}{2}\|\boldsymbol{d} - \mathbf{z}\|_{\ell_2}^2 \\ \text{subject to} \quad & \boldsymbol{d} \preceq t \end{aligned} \tag{17}$$

Algorithm 3: Summary of the Procedure Used to Solve Eq. (16).

Compute $z_r = \gamma_r/\rho - \log\det\left(\mathbf{F}(\mathbf{p})\boldsymbol{\Sigma}_r\mathbf{F}^H(\mathbf{p})\boldsymbol{\Sigma}_w^{-1} + \mathbf{I}_{M,M}\right)$;
Compute $\mathbf{d} = -\mathbf{c}$;
Sort $\mathbf{z}$ in descending order ;
Find t^* by finding the smallest value of m such that Eq. (24) is satisfied ;
Compute $\mathbf{d}^* = \min(\mathbf{z}, t^*)$;
Unsort $\mathbf{d}^*$ so that the elements are arranged in their original order ;
Compute $\mathbf{c}^* = -\mathbf{d}^*$;

The Lagrangian for this problem can be written in terms of a single dual variable $\boldsymbol{\alpha}$:

$$\mathcal{L}(\boldsymbol{d}, t, \boldsymbol{\alpha}) = t + \frac{\rho}{2}\|\boldsymbol{d} - \mathbf{z}\|_{\ell_2}^2 + \sum_r \alpha_r(d_r - t) \tag{18}$$

and the Karush Kuhn Tucker (KKT) conditions [33] mandate that the following relationships are satisfied at the optimal point $\mathbf{d}^*, t^*, \boldsymbol{\alpha}^*$:

$$\mathbf{d}^* = \mathbf{z} - \boldsymbol{\alpha}^*/\rho \tag{19}$$

$$\sum_r \alpha_r^* = 1 \tag{20}$$

$$\alpha_r^*(d_r^* - t^*) = 0 \quad \forall i \tag{21}$$

$$\boldsymbol{\alpha}^* \succeq 0 \tag{22}$$

It is clear from Eq. (19) and Eq. (21) that the Lagrange multipliers satisfy the following relationship:

$$\alpha_r^* = \max(\rho(z_r - t^*), 0) \tag{23}$$

Therefore, one only needs to find the m non-zero elements of $\boldsymbol{\alpha}$ where the constraints are active. Assuming that $\mathbf{z}$ is sorted in descending order, we can combine Eq. (23) and Eq. (20) to reveal the following condition:

$$t^* = -1/\rho + \sum_{r=1}^{m} z_r > z_{m+1} \tag{24}$$

One therefore needs only to find the smallest value of m that satisfies Eq. (24) in order to solve Eq. (17). In total, the steps for solving Eq. (16) are summarized in Algorithm 3.

Appendix C
$\mathbf{p}$ Update Step

In the $\mathbf{p}$ update step, we seek the minimizer of Eq. (12) with respect to $\mathbf{p}$ while keeping $\mathbf{u}$, ρ, and $\boldsymbol{\gamma}$ constant. By introducing the auxiliary variable $\mathbf{z} = \mathbf{c} + \boldsymbol{\gamma}/\rho$, this subproblem can be reduced to the following non-convex optimization program:

$$\underset{\boldsymbol{p}}{\text{minimize}}\ I_{Q_p}(\boldsymbol{p}) + \sum_{r=1}^{R} \frac{\rho}{2}\left|\log\det\left(\mathbf{F}(\boldsymbol{p})\boldsymbol{\Sigma}_r\mathbf{F}^H(\boldsymbol{p})\boldsymbol{\Sigma}_w^{-1} + \mathbf{I}_{M,M}\right) - z_r\right|^2 \tag{25}$$

The optimal solution to Eq. (25) is difficult to compute, even when $\mathbf{F}(\cdot)$ takes a simple form. Instead, we consider an approximate update using the proximal gradient, provided that $\mathbf{F}(\mathbf{p})$ is differentiable over Q_p. Assuming this condition holds, the derivative of the $\log\det$ terms can be written as follows:

$$\begin{aligned}
&\frac{\partial}{\partial p_r}\left(\sum_{r=1}^{R}\frac{\rho}{2}\left|\log\det\left(\mathbf{F}(\mathbf{p})\mathbf{\Sigma}_r\mathbf{F}^H(\mathbf{p})\mathbf{\Sigma}_w^{-1}+\mathbf{I}_{M,M}\right)-z_r\right|^2\right)\\
&=\sum_{r=1}^{R}\frac{\rho}{2}\left(\log\det\left(\mathbf{F}(\mathbf{p})\mathbf{\Sigma}_r\mathbf{F}^H(\mathbf{p})\mathbf{\Sigma}_w^{-1}+\mathbf{I}_{M,M}\right)-z_r\right)\\
&\quad\left(\mathbf{F}(\mathbf{p})\mathbf{\Sigma}_r\mathbf{F}^H(\mathbf{p})\mathbf{\Sigma}_w^{-1}+\mathbf{I}_{M,M}\right)^{-1}\\
&\quad\left(\frac{\partial\mathbf{F}(\mathbf{p})}{\partial p_r}\mathbf{\Sigma}_r\mathbf{F}^H(\mathbf{p})+\mathbf{F}(\mathbf{p})\mathbf{\Sigma}_r\frac{\partial\mathbf{F}^H(\mathbf{p})}{\partial p_r}\right)\mathbf{\Sigma}_w^{-1} \qquad (26)
\end{aligned}$$

Expressing this gradient as the vector $\mathbf{g}$, the proximal gradient method updates $\mathbf{p}$ by solving the following optimization program:

$$\underset{\boldsymbol{p}}{\text{minimize}}\; I_{Q_p}(\boldsymbol{p})+\frac{1}{2\kappa}\|\boldsymbol{p}-(\mathbf{z}-\kappa\mathbf{g})\|_{\ell_2}^2 \qquad (27)$$

where the step size κ is computed using an inexact line search method. Our implementation of the design method uses the inexact proximal gradient line search method described in [34]. This line search method ensures that the objective function of Eq. (12) decreases on each iteration.

REFERENCES

[1] 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, Feb. 2006.

[2] E. J. Candes, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[3] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[4] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Nov. 2001.

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

[6] M. Stojnic, "ℓ_2/ℓ_1-optimization in block-sparse compressed sensing and its strong thresholds," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 350–357, Apr. 2010.

[7] Z. Zeinalkhani and A. H. Banihashemi, "Iterative reweighted ℓ_2/ℓ_1 recovery algorithms for compressed sensing of block sparse signals," *IEEE Trans. Signal Process.*, vol. 63, no. 17, pp. 4516–4531, Sep. 2015.

[8] R. Garg and R. Khandekar, "Block-sparse solutions using kernel block rip and its application to group lasso," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 296–304.

[9] Y. C. Eldar and H. Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 505–519, Jan. 2010.

[10] Y. C. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov. 2009.

[11] Y. C. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[12] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3075–3085, Aug. 2009.

[13] Y. Gao and M. Ma, "A new bound on the block restricted isometry constant in compressed sensing," *J. Inequalities Appl.*, vol. 2017, no. 1, 2017, Art. no. 174.

[14] J. Nocedal and S. J. Wright, "Numerical optimization," 2nd ed. New York, NY, USA: Springer, 2006.

[15] G. Li, Z. Zhu, D. Yang, L. Chang, and H. Bai, "On projection matrix optimization for compressive sensing systems," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2887–2898, Jun. 2013.

[16] W. Chen, M. R. Rodrigues, and I. J. Wassell, "Projection design for statistical compressive sensing: A tight frame based approach," *IEEE Trans. Signal Process.*, vol. 61, no. 8, pp. 2016–2029, Apr. 2013.

[17] C. Lu, H. Li, and Z. Lin, "Optimized projections for compressed sensing via direct mutual coherence minimization," *Signal Process.*, vol. 151, pp. 45–55, 2018.

[18] Q. Bao, C. Jiang, Y. Lin, W. Tan, Z. Wang, and W. Hong, "Measurement matrix optimization and mismatch problem compensation for DLSLA 3-D SAR cross-track reconstruction," *Sensors*, vol. 16, no. 8, 2016, Art. no. 1333.

[19] V. Abolghasemi, S. Ferdowsi, B. Makkiabadi, and S. Sanei, "On optimization of the measurement matrix for compressive sensing," in *Proc. 18th IEEE Eur. Signal Process. Conf.*, 2010, pp. 427–431.

[20] R. Obermeier and J. M. Lorenzo, "Sensing matrix design via mutual coherence minimization for electromagnetic compressive imaging applications," *IEEE Trans. Comput. Imag.*, vol. 3, no. 2, pp. 217–229, Jun. 2017.

[21] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Sensing matrix optimization for block-sparse decoding," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4300–4312, Sep. 2011.

[22] L. Zelnik-Manor, K. Rosenblum, and Y. C. Eldar, "Dictionary optimization for block-sparse representations," *IEEE Trans. Signal Process.*, vol. 60, no. 5, pp. 2386–2395, May 2012.

[23] S. Li, Z. Zhu, G. Li, L. Chang, and Q. Li, "Projection matrix optimization for block-sparse compressive sensing," in *Proc. IEEE Int. Conf. Signal Process., Commun. Comput.*, 2013, pp. 1–4.

[24] Z. Li, J. Xie, G. Zhu, X. Peng, Y. Xie, and Y. Choi, "Block-based projection matrix design for compressed sensing," *Chin. J. Electron.*, vol. 25, no. 3, pp. 551–555, 2016.

[25] L. Qin, S. Zhang, X. Guo, and G. Wang, "A novel framework of measurement matrix optimization for block sparse recovery," in *Proc. IEEE Int. Conf. Inf. Automat.*, 2017, pp. 58–64.

[26] W. R. Carson, M. Chen, M. R. Rodrigues, R. Calderbank, and L. Carin, "Communications-inspired projection design with application to compressive sensing," *SIAM J. Imag. Sci.*, vol. 5, no. 4, pp. 1185–1212, 2012.

[27] A. Ashok, L.-C. Huang, and M. A. Neifeld, "Information optimal compressive sensing: Static measurement design," *JOSA A*, vol. 30, no. 5, pp. 831–853, 2013.

[28] Y. Gu, N. A. Goodman, and A. Ashok, "Radar target profiling and recognition based on TSI-optimized compressive sensing kernel," *IEEE Trans. Signal Process.*, vol. 62, no. 12, pp. 3194–3207, Jun. 2014.

[29] L. Wang, M. Chen, M. R. Rodrigues, D. Wilcox, A. R. Calderbank, and L. Carin, "Information-theoretic compressive measurement design," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1150–1164, Jun. 2017.

[30] G. Braun, S. Pokutta, and Y. Xie, "Info-greedy sequential adaptive compressed sensing," in *Proc. 52 Annu. Allerton Conf. Commun., Control, Comput.*, 2014, pp. 858–865.

[31] L. Wang *et al.*, "Signal recovery and system calibration from multiple compressive poisson measurements," *SIAM J. Imag. Sci.*, vol. 8, no. 3, pp. 1923–1954, 2015.

[32] L. Wang, A. Razi, M. Rodrigues, R. Calderbank, and L. Carin, "Nonlinear information-theoretic compressive measurement design," in *Proc. Int. Conf. Mach. Learn.*, 2014, pp. 1161–1169.

[33] H. Kuhn and A. Tucker, "Nonlinear programming," in *Proc. 2nd Berkeley Symp. Math. Statist. Probability*, 1951, pp. 481–491.

[34] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

Authors' photographs and biographies not available at the time of publication.