

Practical Compressive Sensing with Toeplitz and Circulant Matrices

Wotao Yin^a, Simon Morgan^b, Junfeng Yang^c, Yin Zhang^a

^aRice University, Dept. of Computational and Applied Mathematics, Houston, TX, USA.

^bNew Mexico Consortium, Los Alamos, NM, USA.

^cNanjing University, Department of Mathematics, Nanjing, Jiangsu, China.

ABSTRACT

Compressive sensing encodes a signal into a relatively small number of incoherent linear measurements. In theory, the optimal incoherence is achieved by completely random measurement matrices. However, such matrices are often difficult and costly to implement in hardware realizations.

Random Toeplitz and circulant matrices can be easily (or even naturally) realized in various applications. This paper introduces fast algorithms for reconstructing signals from incomplete Toeplitz and circulant measurements. Computational results are presented to show that Toeplitz and circulant matrices are not only as effective as random matrices for signal encoding, but also permit much faster decoding.

Keywords: Compressive Sensing, Circulant, Toeplitz, Signal Reconstruction, Random Matrix

1. INTRODUCTION

Compressive sensing (CS), an emerging methodology brought to the research forefront by Donoho¹ and Candes, Romberg, and Tao,² acquires a compressible signal from a relatively small set of linear projections. Let \bar{x} denote an n -dimensional real or complex vector that is sparse under a certain basis Ψ , and $b := A\bar{x}$ represent a set of m linear projections of \bar{x} . The basis pursuit problem

$$\text{BP: } \min_x \|\Psi x\|_1 \quad \text{s.t. } Ax = b \quad (1)$$

has been known to return a sparse Ψx . Let k be the number of nonzeros in $\Psi\bar{x}$. It is shown^{3,4} that, when A is a Gaussian random or partial Fourier ensemble, BP returns a solution equal to \bar{x} with high probability from $m = O(k \log(n/k))$ or $O(k \log^4(n))$ linear measurements, respectively.

In CS applications, the acquisition of the linear projections $A\bar{x}$ requires a physical implementation. In most cases, the use of an i.i.d. Gaussian random matrix A is either impossible or overly expensive. This motivates the study of easily implementable CS matrices. Two such matrices are the Toeplitz and circulant matrices, which have been shown to be almost *as effective as* the Gaussian random matrix for CS encoding/decoding. Toeplitz and circulant matrices have the forms, respectively,

$$T = \begin{bmatrix} t_n & t_{n-1} & \cdots & t_1 \\ t_{n+1} & t_n & \cdots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_{2n-1} & t_{2n-2} & \cdots & t_n \end{bmatrix} \quad \text{and} \quad C = \begin{bmatrix} t_n & t_{n-1} & \cdots & t_1 \\ t_1 & t_n & \cdots & t_2 \\ \vdots & \vdots & \ddots & \vdots \\ t_{n-1} & t_{n-2} & \cdots & t_n \end{bmatrix},$$

where every left-to-right descending diagonal is constant, i.e., $T_{i,j} = T_{i+1,j+1}$. If T satisfies the additional property that $t_i = t_{n+i}, \forall i$, it is also a circulant matrix C . Since any (partial) Toeplitz matrix can be extended to a (partial) circulant matrix, our discussions below are based exclusively on circulant matrices. Using Toeplitz, rather than circulant, matrices will require some minor computation overhead.

Send correspondence to the first author at wotao.yin@rice.edu.

1.1 Toeplitz and Circulant Measures

In various physical domains, it is easy to compute $T_\Omega \bar{x}$, where T is either Toeplitz or circulant and T_Ω is its submatrix containing a subset of rows of T . Since the multiplication $T\bar{x}$ is equivalent to the discrete convolution operation $h \star \bar{x}$ for a certain vector h , $T_\Omega \bar{x}$ is the Ω -subsample of the convolution $h \star \bar{x}$.

Tropp *et al.*⁵ described a *random filter* for acquiring a signal \bar{x} : a random vector h , called a random-tap FIR, is convolved with \bar{x} followed by downsampling to yield the compressive measurements b . In this example, the sensing matrix is the circulant matrix induced by h . Haupt *et al.*⁶ introduced the application of random Toeplitz sampling in the channel estimation problem, arising from wireless communication systems, to identify a vector \bar{x} (called impulse responses) that characterizes a discrete linear time-invariant (LTI) system.

Random convolutions can also be applied in some imaging systems in which convolutions either naturally arise or can be physically realized.^{7–11} In spotlight synthetic aperture radar (SAR) imaging, the reflection from a probing pulse can be modeled as a convolution of the probing pulse with a reflectivity profile.⁷ In an optical imaging system, one can apply a so-called spatial light modulator (SLM, which acts like a mask) between the image source \bar{x} and a lens.^{9,10} Besides moving an SLM, random convolutions can be realized by an optical correlator.¹¹ Since a convolution matrix C can be diagonalized by a Fourier transform

$$C = F^* D F, \quad (2)$$

where F is the discrete Fourier matrix of the same size as C , implementing $C\bar{x}$ is equivalent to realizing $F^* D F \bar{x}$, which can be done through optical lens and a static SLM. One can then subsample $F^* D F \bar{x}$ or randomly demodulate it by letting it pass through another SLM before reaching a detector array. An important advantage of the imaging systems described above is that the photodiodes can be swapped out for alternatives, e.g., spectrometers for hyperspectral imagers.

Recently, the use of Toeplitz and circulant matrices has been proposed for compressive MR imaging by Liang *et al.*¹² Our comparative results in Section 4.2 suggest that if this measurement scheme becomes feasible for MRI, then the number of measurements can potentially be considerably smaller than that required by Fourier schemes in a CS setting.

Theoretically, it is shown by Romberg¹¹ that C in (2) is incoherent with any orthonormal basis Ψ with high probability. Let \bar{x} be a signal such that $\bar{x} = \Psi \bar{\alpha}$ and $\text{sign}(\bar{\alpha})$ is uniformly random and has k nonzeros. It is also shown¹¹ that from measurements $b := C_\Omega \bar{x}$ with m randomly chosen locations Ω where $m \geq c(k \log n + \log^3 n)$ for a constant c , ℓ_1 minimization recovers \bar{x} with probability at least $1 - O(n^{-1})$. Bajwa *et al.*¹³ and Haupt *et al.*^{6,14} have results for C_Ω generated from a random vector t and Ω consisting of consecutive indices.

1.2 Contributions

We first empirically demonstrate that random circulant matrices encode most signal as effectively as random Gaussian matrices. In particular, certain random circulant matrices appear universally effective on all tested signals. We then introduce fast algorithms for CS signal reconstruction from Toeplitz and circulant measurements with various forms of regularization and fidelity. The unknown signals can be one, two, or higher dimensional. The reconstruction models involve regularization terms including the ℓ_1 -norm (with or without a sparsifying basis), total variation (TV), as well as their weighted variants, of the unknown signal. The models also involve different data fidelity terms such as exact equality (noiseless data) and ℓ_1 or ℓ_2 square penalty (noisy data). Extensions are also described. Moreover, the algorithms are suitable for implementation on embedded systems and on GPUs. A code for our implementation is available for download at <http://www.caam.rice.edu/~optimization/L1/RecPC>.

2. RANDOM CIRCULANT VS I.I.D. RANDOM MEASUREMENTS

An $m \times n$ general matrix has mn degrees of freedom, but a partial circulant matrix of the same size has at most n degrees of freedom. Hence, a random circulant matrix is generated from much fewer independent random numbers or is “much less random” than an i.i.d. random matrix of the same size. This fact seemingly suggests that a random circulant matrix would yield less incoherent projections, and consequently worse CS recovery.

However, numerical results below will hopefully convince us of a different story: circulant matrices can be equally effective as i.i.d. random matrices.

To obtain a broad picture, we test a variety of circulant matrix types and signal types for CS recovery. Each sensing matrix is generated by $A = PC$ from an $n \times n$ circulant matrix C and an $m \times n$ selection matrix P . The types of C and P are listed in Table 2. There are two major approaches to obtain C : type R generating the first row of C , and type D generating a diagonal complex matrix $D = \text{diag}(d)$ and letting $C = F^*DF$. Type D1, proposed by Romberg,¹¹ has uniform magnitude d_i 's, $|d_i| = 1$, and a randomly generated phase for each d_i . D2 assigns varying magnitudes for the d_i 's according to a “ \wedge ” shaped curve taking values between 0.1 and 2 and also has a randomly generated phase for each d_i ; hence, the resulting matrices C are complex; a real A is formed by stacking the real and imaginary parts of PC , where P selects only $m/2$ rows so that A is $m \times n$.

C and P types for sensing matrix $A = PC$

C-type			P-type		
R1	Circulant:	1st row i.i.d. Gaussian	real [†]	P1	First m rows
R2	1st row given	1st row i.i.d. Bernoulli			
D1	Circulant:	d_i random phase, $ d_i = 1$		P2	Equally-spaced m rows
D2	$C = F^*DF$	d_i random phase			
D3	$D = \text{diag}(d)$	D1 without conjugate symmetry [†]	complex [‡]	P3	Randomly-spaced m rows
I1	i.i.d. random	i.i.d. Gaussian	real		
I2	Non-circulant	i.i.d. Bernoulli			

†: A real matrix $C = F^*DF$ requires a *conjugate symmetric* d of types D1 or D2. An N -dimensional array $x \in \mathbb{C}^{n_1 \times \dots \times n_N}$ is conjugate symmetric (or Hermitian symmetric) if $x(i_1, \dots, i_N) = \text{conj}(x(i'_1, \dots, i'_N))$ where $i'_j := \text{mod}(n_j - i_j + 1, n_j) - 1$.

‡: An $m \times n$ real matrix A is formed by stacking the real and imaginary parts of an $\frac{m}{2} \times n$ matrix PC .

List of signal types

X1	k -sparse Gaussian	i.i.d. Gaussian, randomly supported nonzero
X2	k -sparse Bernoulli	i.i.d. Bernoulli, randomly supported nonzero
X3	k -sparse under DCT	DCT matrix times a vector of X1 type
X4	permuted X3	same as X3 except \bar{x}_i 's are randomly permuted
X5	X1 with noise	X1 plus Gaussian white noise of varying levels
X6	power-law decay	sorted $ \bar{x}_{[i]} = i^{-3}$, \bar{x}_i 's randomly signed and permuted
X7	total variation sparse	$\bar{y}_i = \bar{x}_{i+1} - \bar{x}_i$, \bar{y} is of type X1

We test seven different types of signals listed in the table above. A type-X3 signal can be written as $x = \Psi u$, where Ψ is a discrete cosine transform (DCT) and u is of type X1. Signals of type X6 often arise in signal and image processing. Type X7 is often assumed on piece-wise constant signals and used in image processing.

In all tests, we set the number of measurements to $m = 128$ and signal dimension to $n = 512$. For type X5, we fix $k = 8$ and vary noise standard deviations; $\sigma = 10^{-4} \times 2^i$, $i = 0, 1, \dots, 8$. For each test, 1000 independent tests are run. Types X1 through X6 are recovered by ℓ_1 minimization with constraints $Ax = b$ and type X7 by total variation minimization with the same constraints. Since the involved basis pursuit problems, $\min\{\|x\|_1 : Ax = b\}$, are small-scale linear programs, we use the highly reliable commercial package Gurobi version 2 with default settings. The comparison results are given in Figure 1. On all except type-X3 signals, all the tested circulant matrices perform similarly and are no worse than i.i.d. random matrices (types I1 and I2) in terms of recovery frequencies and recovery errors. On the type-X3 signals, only matrices of types D1/P3 and D3/P3 can be competitive with the i.i.d. matrices; others give worse recovery results to different degrees. Although not all results are shown in the figure due to space limitation, performance differences have only been observed on the type-X3 signals.

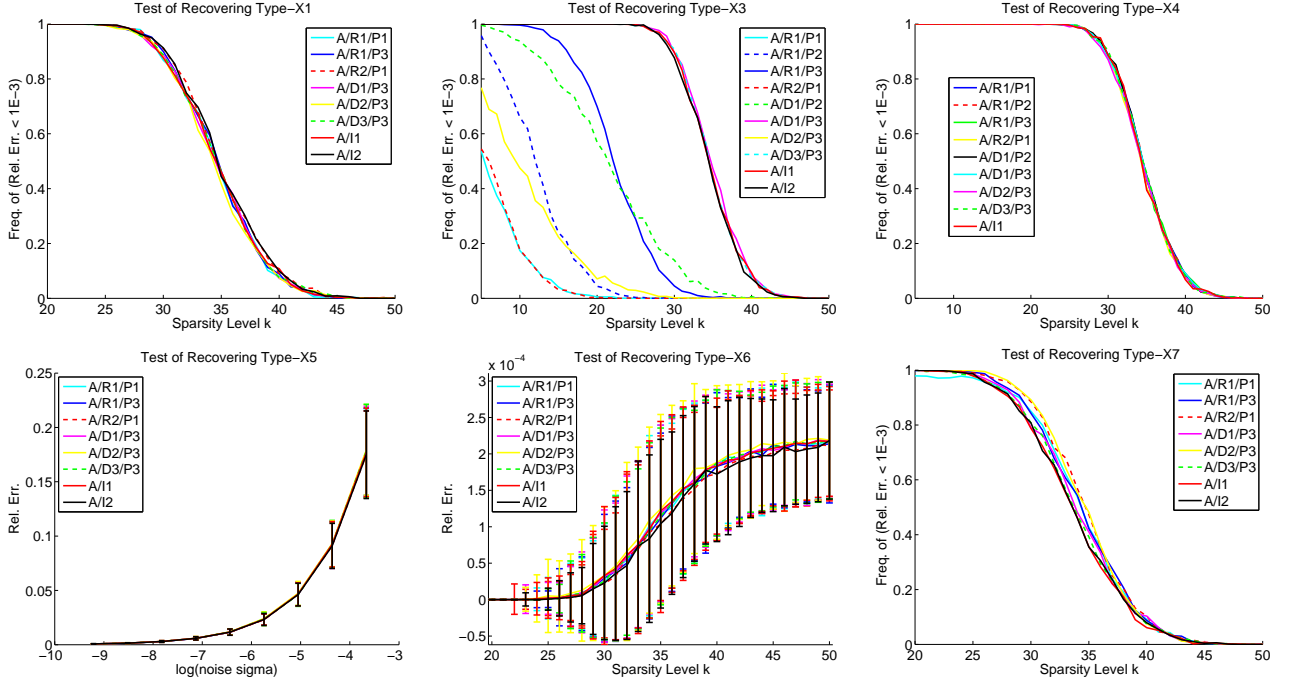


Figure 1. Recovery of 1D signals of types: X1, X3, X4, X5, X6, and X7. The plot for X2 (not shown) is almost identical to that of X1.

To effectively encode type-X3 signals, it is important¹¹ to have uniform magnitude for the d_i 's and random sampling (type P3). Types D1 and D3 are incoherent with any orthonormal sparsifying basis with high probability.

Instead of unit magnitude d_i 's, scrambling the orders of signal coordinates also works. All tested sensing matrices A have the same performance on signals of type X4, which are obtained by randomly permuting the entries of type-X3 signals. However, the physical implementation of random permutations may become an added burden.

3. FAST ALGORITHMS FOR TOEPLITZ AND CIRCULANT MATRICES

Using Toeplitz and circulant sensing matrices allows significantly faster CS reconstruction compared to using i.i.d. random matrices. We introduce fast algorithms in this section for solving a variety of CS reconstruction models with Toeplitz and circulant sensing matrices, including equality fidelity, ℓ_1 and ℓ_2 square penalized fidelity, as well as ℓ_1 and/or total variation regularizations. We also briefly discuss relevant extensions.

3.1 Notation and reconstruction models

For notational simplicity, we assume that the underlying signal is a real, single-channel, 2D image with an $n \times n$ domain, denoted by $x \in \mathbb{R}^{n^2}$. Generalizations and extensions are discussed in subsections 3.3–3.5 below. The discretized total variation of x is defined as $\text{TV}_v(x) = \sum_{i=1}^{n^2} v_i \|D_i x\|$, where, for each i , $D_i \in \mathbb{R}^{2 \times n^2}$ is a local finite difference operator, $v_i \geq 0$ is a weighting parameter, and $\|\cdot\|$ is either the 2-norm (isotropic discretization) or weighted 1-norm (anisotropic discretization). In the following, we let $\|\cdot\|$ be the 2-norm and $\|\cdot\|_{w,1}$ be the weighted 1-norm, i.e., $\|z\|_{w,1} = \sum_i w_i |z_i|$, where w_i 's are nonnegative local weights.

Our algorithmic approach is, with appropriate modifications, applicable to the following models

$$\text{BP} : \quad \min \alpha \text{TV}_v(x) + \beta \|\Psi x\|_{w,1}, \text{ s.t. } Ax = b, \quad (3)$$

$$\text{BPDN} : \quad \min \alpha \text{TV}_v(x) + \beta \|\Psi x\|_{w,1} + \frac{\mu}{2} \|Ax - b\|^2, \quad (4)$$

$$\text{BPDN/L1} : \quad \min \alpha \text{TV}_v(x) + \beta \|\Psi x\|_{w,1} + \mu \|Ax - b\|_1, \quad (5)$$

where $\alpha, \beta \geq 0$ (but at least one is positive) and $\mu > 0$ are scalar parameters. Ψ can be either the identity or an *orthonormal* transform operator such as a wavelet transform. The sensing matrix A is a partial circulant matrix and has the form $A = PC$, where $C \in \mathbb{R}^{n^2 \times n^2}$ is a block-circulant matrix with each block being a circulant matrix and $P \in \mathbb{R}^{m \times n^2}$ is a selection operator. We write $k \in P$ if the k th row is selected by P .

Below we take BPDN as an example to introduce our algorithm. Our algorithms can deal with isotropic and anisotropic discretizations of TV. For simplicity, we only deal with the isotropic TV and uniform weights $v_i \equiv 1$ and $w_i \equiv 1$. The treatment for anisotropic TV and/or local weighted cases are completely analogous. For convenience, we let ψ_i^\top be the i th row of Ψ . Using the above notation, the BPDN model becomes

$$\min \alpha \sum_{i=1}^{n^2} \|D_i x\| + \beta \sum_{i=1}^{n^2} |\psi_i^\top x| + \frac{\mu}{2} \|PCx - b\|^2. \quad (6)$$

We mention that each of the above three models has a “nonnegativity counterpart” in which the underlying signal x is real and satisfies additional constraints $x \geq 0$. The algorithmic approach proposed here can be extended to solve these nonnegativity counterparts with some rather straightforward modifications.

3.2 Augmented Lagrangian and alternating direction methods

In a nutshell, we apply variable-splitting to a problem in the form of $\min\{f(Lx) + g(x)\}$, obtaining $\min\{f(y) + g(x) : Lx - y = 0\}$, and consider its augmented Lagrangian $f(y) + \langle \lambda, Lx - y \rangle + \|Lx - y\|_2^2 + g(x)$. This splitting-variable technique was first applied to TV regularization for image deconvolution.¹⁵ By introducing $\mathbf{y} = [\mathbf{y}_1, \dots, \mathbf{y}_{n^2}]$, where each $\mathbf{y}_i \in \mathbb{R}^2$, and $z, u \in \mathbb{R}^{n^2}$, we transform problem (6) to

$$\min_{\mathbf{y}, z, u, x} \left\{ \alpha \sum_i \|\mathbf{y}_i\| + \beta \sum_i |z_i| + \frac{\mu}{2} \|Pu - b\|^2, u = Cx, \mathbf{y}_i = D_i x, z_i = \psi_i^\top x, i = 1, 2, \dots, n^2 \right\}. \quad (7)$$

For $s, t, \rho \in \mathbb{R}$ and $\mathbf{s}, \mathbf{t}, \nu \in \mathbb{R}^2$, we define

$$\varphi_1(s, t, \rho) = |s| - \rho(s - t) + \frac{\beta_1}{2} \|s - t\|^2 \quad \text{and} \quad \varphi_2(\mathbf{s}, \mathbf{t}, \nu) = \|\mathbf{s}\| - \nu^\top (\mathbf{s} - \mathbf{t}) + \frac{\beta_2}{2} \|\mathbf{s} - \mathbf{t}\|^2, \quad (8)$$

where $\beta_1, \beta_2 > 0$ are parameters*. Then, the augmented Lagrangian function of (7) can be written as

$$\mathcal{L}_{\mathcal{A}}(\mathbf{y}, z, u, x, \lambda) = \alpha \sum_i \varphi_2(\mathbf{y}_i, D_i x, (\lambda_2)_i) + \beta \sum_i \varphi_1(z_i, \psi_i^\top x, (\lambda_1)_i) + \frac{\beta_3}{2} \left\| u - Cx - \frac{\lambda_3}{\beta_3} \right\|^2 + \frac{\mu}{2} \|Pu - b\|^2, \quad (9)$$

where $\beta_3 > 0$ is a penalty parameter and $\lambda = (\lambda_1, \lambda_2, \lambda_3)$ contains the Lagrangian multipliers. For each i , $(\lambda_1)_i \in \mathbb{R}$ and $(\lambda_2)_i \in \mathbb{R}^2$, and $\lambda_3 \in \mathbb{R}^{n^2}$. Given $(\mathbf{y}^k, z^k, u^k, x^k)$ and λ^k , the classical augmented Lagrangian method^{16, 17} (ALM) for (7) iterates as

$$\begin{cases} (\mathbf{y}^{k+1}, z^{k+1}, u^{k+1}, x^{k+1}) \leftarrow \arg \min \mathcal{L}_{\mathcal{A}}(\mathbf{y}, z, u, x, \lambda^k), \\ (\lambda_1)_i^{k+1} \leftarrow (\lambda_1)_i^k - \gamma_1 \beta_1 (z_i^{k+1} - \psi_i^\top x^{k+1}), \forall i, \\ (\lambda_2)_i^{k+1} \leftarrow (\lambda_2)_i^k - \gamma_2 \beta_2 (\mathbf{y}_i^{k+1} - D_i x^{k+1}), \forall i, \\ (\lambda_3)^{k+1} \leftarrow (\lambda_3)^k - \gamma_3 \beta_3 (u^{k+1} - Cx^{k+1}). \end{cases} \quad (10)$$

In the ALM framework, each iteration requires an accurate minimization of $\mathcal{L}_{\mathcal{A}}$ jointly with respect to \mathbf{y}, z, u and x , which can be expensive. It is easy to see that our $\mathcal{L}_{\mathcal{A}}$ is separable with respect to \mathbf{y}, z and u for fixed λ

*Do not confuse β with β_1, β_2 , and β_3 .

and x . To take full advantage of this structure, we apply the *alternating direction method of multipliers*^{18,19} or ADM, which has recently been successfully applied to various signal and image reconstruction applications.²⁰

In our case, each iteration of ADM decreases $\mathcal{L}_A(\mathbf{y}, z, u, x, \lambda^k)$ via one round of alternating minimization with respect to (\mathbf{y}, z, u) and x each, i.e.,

$$\begin{cases} (\mathbf{y}^{k+1}, z^{k+1}, u^{k+1}) \leftarrow \arg \min \mathcal{L}_A(\mathbf{y}, z, u, x^k, \lambda^k), \\ x^{k+1} \leftarrow \arg \min \mathcal{L}_A(\mathbf{y}^{k+1}, z^{k+1}, u^{k+1}, x, \lambda^k), \end{cases} \quad (11)$$

then followed by multiplier updates the same as in (10). For fixed $\lambda = \lambda^k$ and $x = x^k$, since $\mathcal{L}_A(\mathbf{y}, z, u, x^k, \lambda^k)$ is separable with respect to \mathbf{y} , z , and u each, the joint minimization of (\mathbf{y}, z, u) can be carried out in parallel. Specifically, z^{k+1} is determined by

$$z_i^{k+1} \leftarrow \arg \min_{z_i} \varphi_1(z_i, \psi_i^\top x^k, (\lambda_1)_i^k) \triangleq S_1(\psi_i^\top x^k + (\lambda_1)_i / \beta_1, 1 / \beta_1), \quad \forall i, \quad (12)$$

where $S_1(\cdot, 1/\beta_1)$ is the one-dimensional shrinkage operator defined as $(\max\{|\xi| - 1/\beta_1, 0\} \cdot \text{sgn}(\xi))$, where $\text{sgn}(\cdot)$ is the signum function, \mathbf{y}^{k+1} is determined by

$$\mathbf{y}_i^{k+1} \leftarrow \arg \min_{\mathbf{y}_i} \varphi_2(\mathbf{y}_i, D_i x^k, (\lambda_2)_i^k) \triangleq S_2(D_i x^k + (\lambda_2)_i^k / \beta_2, 1 / \beta_2), \quad \forall i, \quad (13)$$

where $S_2(\cdot, 1/\beta_2)$ is known as the two-dimensional shrinkage operator (which can be easily extended to high-dimensional cases with weights) defined as $(\max\{\|\mathbf{s}\| - 1/\beta_2, 0\} \cdot \mathbf{s} / \|\mathbf{s}\|)$, where $0 \cdot (0/0) = 0$ is assumed, and the minimization with respect to u is attained by

$$u^{k+1} \leftarrow (I + (\mu/\beta_3)P^\top P)u = Cx^k + (\lambda_3^k + \mu P^\top b) / \beta_3, \quad (14)$$

where I represents the identity matrix. Since P is a selection matrix, $P^\top P$ is diagonal and thus the solution to (14) can be easily obtained. The computations of \mathbf{y} , z and u are linear in the dimension of x .

In the second step of ADM scheme (11), the minimization with respect to x is a least squares problem with the normal equations

$$Mx^{k+1} = D^\top(\beta_2 y^{k+1} - \alpha \lambda_2^k) + \beta \Psi^\top(\beta_1 z^{k+1} - \lambda_1^k) + C^\top(\beta_3 u^{k+1} - \lambda_3^k), \quad (15)$$

where $M = \alpha\beta_2 D^\top D + \beta\beta_1 \Psi^\top \Psi + \beta_3 C^\top C$ with $D \in \mathbb{R}^{2n^2 \times n^2}$ being the global finite difference operator, and y^{k+1} is a reordering of \mathbf{y}_i^{k+1} , $i = 1, 2, \dots, n^2$. Under the periodic boundary conditions, $D^\top D$ is block-circulant. Since C is a block-circulant matrix and further noting that Ψ is orthonormal, the coefficient matrix M of (15) is diagonalizable by the two-dimensional discrete Fourier transform. Therefore, the solution of (15) involves two fast Fourier transforms (FFTs). Finally, we update the Lagrangian multipliers as described in (10). The entire algorithm for (6) is summarized below, which can be shown to converge for equally valued $\gamma_i \in (0, (\sqrt{5} + 1)/2)$.²¹

ADM: Input problem data P, C, Ψ, b and model parameters $\alpha, \beta \geq 0$ and $\mu > 0$. Given $\beta_1, \beta_2, \beta_3, \gamma_1, \gamma_2, \gamma_3 > 0$, initialize $x = x^0$, $\lambda = \lambda^0$ and set $k = 0$.

While “not converged”, **Do**

- 1) Obtain z^{k+1} , \mathbf{y}^{k+1} , and u^{k+1} according to (12), (13), and (14), respectively;
 - 2) Compute x^{k+1} by solving (15) using FFTs.
 - 3) Update λ according to Lines 2–4 of (10)
-

The best model parameters α , β , and μ are determined by data, and they can be selected using various existing methods such as cross validation. Typically, the algorithms prefer larger γ 's so we recommend fixing $\gamma_1 = \gamma_2 = \gamma_3 = 1.618$, which was used throughout our simulations. One can also fix $\beta_1 = \beta_2 = 10$. The best value of β_3 varies according to the data. A recommended range for β_3 is 1–20.

3.3 Algorithms for other models

The ADM algorithm described above can be modified to solve the BP (3) and BPDN/L1 (5) models. For complex circulant sensing matrices and complex signals, there needs no change but replacing transpose “ \top ” by conjugate transpose “ $*$ ” throughout our derivations above. To use weighted TV term $\sum_i v_i \|D_i x\|$, we simply replace $1/\beta_2$ by v_i/β_2 in the update formula (13) for \mathbf{y}_i^{k+1} . Similarly, to use weighted ℓ_1 function $\sum_i w_i |\psi_i^\top x|$, we replace $1/\beta_1$ by w_i/β_1 in (12) for z_i^{k+1} . To impose $Ax = b$ in BP rather than penalizing it by $\frac{\mu}{2} \|Ax - b\|^2$, we shall remove the term $\frac{\mu}{2} \|Pu - b\|^2$ from model (7) and add $Pu = b$ as constraints. Consequently, the u -subproblem returns u as the projection of $Cx + \lambda_3/\beta_3$ to $\{u : Pu = b\}$, namely, setting u_i to $(Cx + \lambda_3/\beta_3)_i$ for the non-samples $i \notin P$ or to b_i otherwise. The ℓ_1 -fidelity term $\mu \|Ax - b\|_1$ in BPDN/L1 can be used to model impulsive or salt-and-pepper noise in measurements, or unknown corrupted measurements. To use this term, simply replace $\frac{\mu}{2} \|Pu - b\|^2$ by $\mu \|Pu - b\|_1$. The resulting u -subproblem remains separable for each u_i . Let $y_i := (Cx + \lambda_3/\beta_3)_i$. The update becomes $u_i \leftarrow y_i - \text{Proj}_{[-\mu/\beta_3, \mu/\beta_3]}(y_i - b_i)$ for $i \in P$, and $u_i \leftarrow y_i$ otherwise. The ADM approach can also be applied to the nonnegativity counterparts of these three models with some further modifications.

3.4 Extensions to multi-channel, high-dimensional and embedded applications

Multiple channels refer to the multiple frames in a single signal or multiple signals measured at different times or locations. Examples include color and hyperspectral images, video frames, multi-modality medical images, and samples of a signal taken by spatially spread sensors. If the channels share common sparsity patterns, they should be jointly recovered. Each of the three models (3)–(5) can be extended to recover L channels $x^{[1]}, \dots, x^{[L]}$ altogether by using a data fitting term for each $x^{[l]}$ and certain joint-sparsity regularization terms. The variable splitting technique still applies (see the color image deblurring examples^{22,23}). To work with q -dimensional signals, $q > 2$, one can simply use a suitable finite difference operator and q -dimensional shrinkages and FFTs. There are no other changes necessary for component-wise operations. Finally, the component-wise arithmetics and fast Fourier transforms (FFTs) used in the ADM algorithms are suitable for embedded and parallel computation. Multi-CPU/core and GPU implementations of FFTs are readily available.

3.5 Relationships to other CS algorithms

Models (3)–(5) can be solved by some other algorithms applicable to general sensing matrices, but our algorithms are particularly efficient for partial circulant sensing matrices since they are designed for the special structure that both the finite difference operators and block circulant matrices are diagonalized by FFTs, making them scalable to large-scale data.

On the other hand, the proposed algorithms can be utilized as subproblem solvers in some “enhancement algorithms” — algorithms that achieve a better solution quality by solving a sequence of subproblems with varying data, parameters or weights. Such enhancement algorithms include Bregman iterative regularization^{24,25} where one or more augmented Lagrangian iterations (also called Bregman iterations) are taken to improve the quality of BPDN solutions. Other such examples are nonconvex ℓ_p -regularization^{26,27} for $p < 1$ and the iterative support detection method,²⁸ both designed for achieving enhanced CS recoverability.

4. NUMERICAL EXPERIMENTS

4.1 Running times

We report running times of a preliminary implementation of the proposed algorithm for BP (3) on recovering the Shepp-Logan phantom image of varying sizes from noiseless partial circulant measurements (type P3/D1). The code was written in MATLAB with no parallelization or fine tuning yet. The SNRs and times below were obtained on a Lenovo laptop with an Intel P9600 CPU running 32-bit Windows 7 and MATLAB R2009b. The iterations were stopped upon a relative change $\|x^{k+1} - x^k\|/\|x^{k+1}\|$ reaches 10^{-4} . Times of BPDN (4) on noisy samples were similar.

Shepp-Logan Phantom	256 × 256		512 × 512		1024 × 1024	
Sample Rate	5%	10%	5%	10%	5%	10%
SNR	26.5	64.6	64.2	65.0	62.6	68.5
Time in Seconds	5.4	5.5	40	24.6	295.6	153.4

4.2 Simulation results on the Shepp-Logan phantom image

Partial Fourier sensing is a popular choice for sensing 2D images such as MR images. In this subsection, we compare images recovered from partial Fourier samples and partial circulant samples.

We used the Shepp-Logan phantom image (128×128) and solved BP (3) (for noiseless samples) and BPDN (4) (for noisy samples) with total variation only, i.e., $\beta = 0$. In our experiments, we set $\alpha = 1$, $\mu = 10^8$ for noiseless samples, and $\mu = 10^4$ for noisy samples contaminated by white noise of standard deviation 0.01. In each case, we ran three tests, the first with 15 radial lines of samples (about 11.28% of K-space, 924 *complex* subsamples), the second with 924 uniformly random *complex* subsamples of Fourier, and the last with 1848 *real* circulant (type P3/D1) measurements. The simulation results are presented in Figure 2. For partial Fourier measurements, we solved a TV/L2 model of the form $\min \sum_i \|D_i x\| + \frac{\mu}{2} \|P\mathcal{F}x - b\|^2$, where \mathcal{F} represents the 2D Fourier transform, by the algorithm described in.²⁹

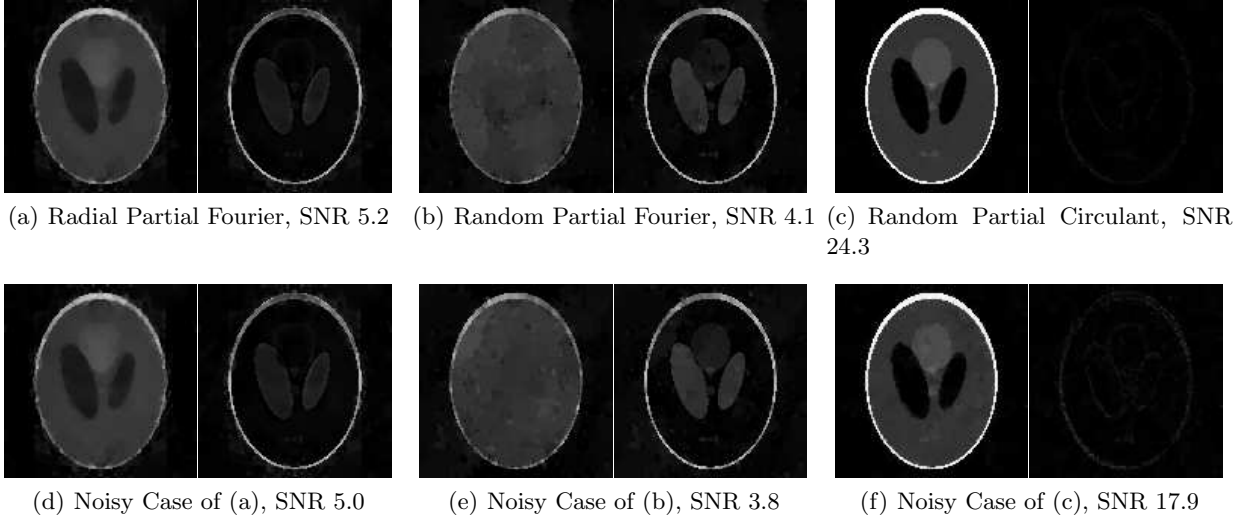


Figure 2. Partial sampling, Fourier versus circulant. Left: recovered images. Right: errors to the groundtruth. (a)–(c): results of noiseless samples. (d)–(f): results of noisy samples. (a) and (d): 15 radial lines (11.28%) of Fourier samples, 924 complex. (b) and (e): 11.28% random Fourier samples, 924 complex. (c) and (f): 11.28% random circulant samples of type P3/D1, 1848 real. The not shown results from 924 complex random circulant samples have SNRs 10.1 (noiseless) and 9.3 (noisy).

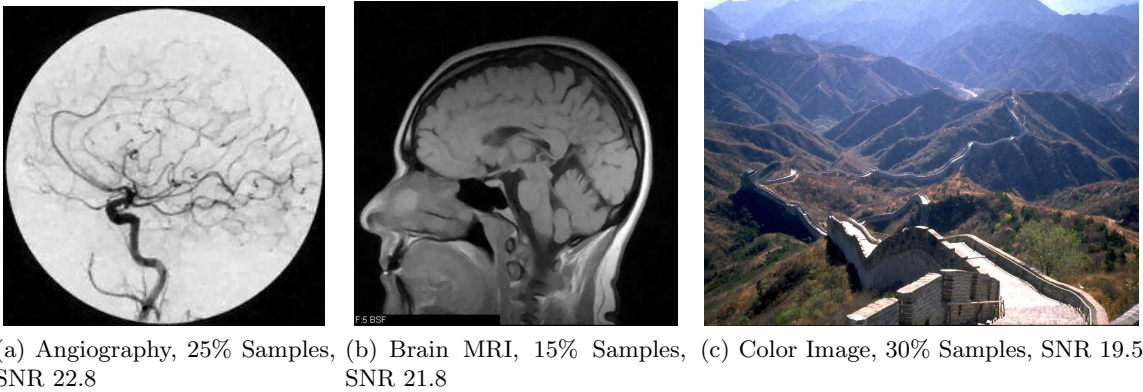


Figure 3. Medical and color images recovered from partial circulant samples.

At the same sample ratio of 11.28%, random circulant samples give better recovered images (Fig.2 (c) and (f)) than those from Fourier samples (Fig.2 (a), (b), (d), and (e)). Images (c) and (f) do not suffer from any

aliasing artifacts of (a) and (d) or the loss of contrast of (b) and (e). Besides total variation reconstruction, we plan to apply the proposed algorithms for the models outlined in section 3.5 above and expect to obtain improved results.

4.3 Medical and color images

We tested recovering medical and color images from partial circulant samples using the model (3). The results are shown in Figure (3). The original images corresponding to the images from left to right have sizes 252×250 , 512×512 , and 1024×768 , respectively.

4.4 Detecting dominant orientations in images without reconstruction

It is rather difficult to learn geometric information of a signal directly from its CS samples without first recovering the signal, in general, because random CS samples are geometrically uncorrelated. We demonstrate that, however, it is possible to estimate edge orientations of images from partial circulant samples obtained from translating a rigid random mask. Figure 4 depicts two 192×192 images with different dominant edge orientations: the dominant orientation of the left image is horizontal, while that of the right image is vertical. Recall that 2D circulant samples are convolutions between the image and a moving mask (under circular boundary conditions). The measures of a random mask moving horizontally over the left image vary much less than those of the same mask moving vertically. We used this principle to decide whether an image being sensed was the one on the left or on the right of figure 4. For demonstrative purposes, we moved a mask $2s$ steps of one pixel each, to take totally $m = 2s + 1$ measures. The $2s$ steps form an L shape, out of which s are on the horizontal arm of L and another s are on its vertical arm. The entries of the mask took values from i.i.d. uniformly distributed random numbers in $(0,1)$. We compared the TV value of all vertical-step measures to that of the horizontal-step measures to determine a dominant orientation according to the smaller TV value. The success rates were 97% for $m = 9$, 99.7% for $m = 19$, and 100% for $m = 49$ out of 10000 independent trials for each m value.

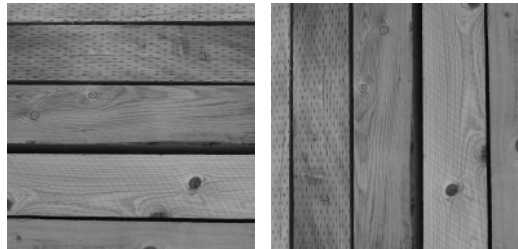


Figure 4.

ACKNOWLEDGMENTS

The authors want to thank Rick Chartrand (Los Alamos), Bingsheng He (Nanjing U.), Kevin Kelly (Rice) for valuable discussions. The work of W. Yin was supported in part by NSF CAREER Award DMS-07-48839, ONR Grant N00014-08-1-1101, the U. S. Army Research Laboratory and the U. S. Army Research Office grant W911NF-09-1-0383, and an Alfred P. Sloan Research Fellowship. The work of Simon Morgan was supported by the DOE. The work of J. Yang was supported in part by the Natural Science Foundation of China grant NSFC-10971095 and the Natural Science Foundation of Jiangsu Province BK2008255. The work of Y. Zhang was supported in part by NSF DMS-0811188 and ONR grant N00014-08-1-1101.

REFERENCES

1. Donoho, D. L., “Compressed sensing,” *Information Theory, IEEE Transactions on* **52**(4), 1289–1306 (2006).
2. Candes, E. J., Romberg, J. K., and Tao, T., “Stable signal recovery from incomplete and inaccurate measurements,” *Communications on Pure and Applied Mathematics* **59**(8), 1207–1223 (2006).
3. Candès, E. and Tao, T., “Near optimal signal recovery from random projections: universal encoding strategies,” *IEEE Transactions on Information Theory* **52**(1), 5406–5425 (2006).
4. Rudelson, M. and Vershynin, R., “Geometric approach to error correcting codes and reconstruction of signals,” *International Mathematical Research Notices* **64**, 4019–4041 (2005).
5. Tropp, J., Wakin, M., Duarte, M., Baron, D., and Baraniuk, R., “Random filters for compressive sampling and reconstruction,” in [*Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*], (2006).

6. Haupt, J., Bajwa, W. U., Raz, G., and Nowak, R. D., "Toeplitz compressed sensing matrices with applications to sparse channel estimation," *Submitted to IEEE Transactions on Information Theory* (2008).
7. Baraniuk, R. and Steeghs, P., "Compressive radar imaging," *IEEE Radar Conference, Waltham, Massachusetts*, 128–133 (2007).
8. Herman, M. A. and Strohmer, T., "High-resolution radar via compressed sensing," *IEEE transactions on signal processing* **57**, 2275–2284 (2009).
9. Marcia, R. and Willett, R., "Compressive coded aperture superresolution image reconstruction," *ICASSP'08*, 833–836 (2008).
10. Marcia, R., Harmany, Z., and Willett, R., "Compressive coded aperture imaging," *SPIE Electronic Imaging'09* (2009).
11. Romberg, J., "Compressive sensing by random convolution," *Submitted to SIAM Journal on Imaging Sciences* (2009).
12. Liang, D., Xu, G., Wang, H., King, K. F., and Ying, L., "Toeplitz random encoding mr imaging using compressed sensing," *IEEE ISBI 2009* (2009).
13. Bajwa, W. U., Haupt, J. D., Raz, G. M., Wright, S. J., and Nowak, R. D., "Toeplitz-structured compressed sensing matrices," *SSP'07* (2007).
14. Rauhut, H., "Circulant and toeplitz matrices in compressed sensing," *In Proc. SPARS'09, Saint Malo, 2009* (2009).
15. Wang, Y., Yang, J., Yin, W., and Zhang, Y., "A new alternating minimization algorithm for total variation image reconstruction," *SIAM Journal on Imaging Sciences* **1**(3), 248–272 (2008).
16. Hestenes, M. R., "Multiplier and gradient methods," *Journal of Optimization Theory and Applications* **4**, 303–320 (1969).
17. Powell, M. J. D., "A method for nonlinear constraints in minimization problems," in [*Optimization*], Fletcher, R., ed., 283–298, Academic Press, New York (1972).
18. Glowinski, R. and Marrocco, A., "Sur l'approximation par elements finis d'ordre un, et la resolution par penalisation-dualite d'une classe de problemes de Dirichlet nonlineaires," *Rev. Francaise d'Aut. Inf. Rech. Oper.* **2**, 41–76 (1975).
19. Gabay, D. and Mercier, B., "A dual algorithm for the solution of nonlinear variational problems via finite-element approximations," *Computers and Mathematics with Applications* **2**, 17–40 (1976).
20. Yang, J. and Zhang, Y., "Alternating direction algorithms for ℓ_1 -problems in compressive sensing," *Rice University CAAM Technical Report TR09-37* (2009).
21. Fortin, M. and Glowinski, R., [*Augmented Lagrangian Methods*], North-Holland, Amsterdam, New York (1983).
22. Yang, J., Yin, W., Zhang, Y., and Wang, Y., "A fast algorithm for edge-preserving variational multichannel image restoration," *SIAM Journal on Imaging Sciences* **2**(2), 569–592 (2008).
23. Yang, J., Zhang, Y., and Yin, W., "An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise," *SIAM Journal on Scientific Computing* **31**(4), 2842–2865 (2009).
24. Osher, S., Burger, M., Goldfarb, D., Xu, J., and Yin, W., "An iterative regularization method for total variation-based image restoration," *SIAM Journal on Multiscale Modeling and Simulation* **4**(2), 460–489 (2005).
25. Yin, W., Osher, S., Goldfarb, D., and Darbon, J., "Bregman iterative algorithms for ℓ_1 -minimization with applications to compressed sensing," *SIAM Journal on Imaging Sciences* **1**(1), 143–168 (2008).
26. Chartrand, R. and Yin, W., "Iteratively reweighted algorithms for compressive sensing," *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 3869–2872 (2008).
27. Chartrand, R., "Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data," *IEEE International Symposium on Biomedical Imaging (ISBI)* (2009).
28. Wang, Y. and Yin, W., "Compressed sensing via iterative support detection," *Rice University CAAM Technical Report TR09-30. Submitted SIAM Journal on Imaging Sciences.* (2009).
29. Yang, J., Zhang, Y., and Yin, W., "A fast alternating direction method for TVL1-L2 signal reconstruction from partial fourier data," *To appear in IEEE Journal of Selected Topics in Signal Processing Special Issue on Compressed Sensing.* (2008).