Performance of the Delsarte-Goethals Frame on Clustered Sparse Vectors

Marco F. Duarte, *Member, IEEE,* Sina Jafarpour, and A. Robert Calderbank, *Fellow, IEEE*

Abstract

The Delsarte-Goethals frame (DGF) has been proposed for deterministic compressive sensing of sparse and compressible signals. Results in compressive sensing theory show that the DGF enables successful recovery of an overwhelming majority of sufficiently sparse signals. However, these results do not give a characterization of the sparse vectors for which the recovery procedure fails. In this paper we present a formal analysis of the DGF that highlights the presence of clustered sparse vectors within its null space. This in turn implies that sparse recovery performance is diminished for sparse vectors that have their nonzero entries clustered together. Such clustered structure is present in compressive imaging applications, where commonly-used raster scannings of 2-D discrete wavelet transform representations yield clustered sparse representations for natural images. Prior work leverages this structure by proposing specially tailored sparse recovery algorithms that partition the recovery of the input vector into known clustered and unclustered portions. Alternatively, we propose new randomized and deterministic raster scannings for clustered coefficient vectors that improve recovery performance. Experimental results verify the aforementioned analysis and confirm the predicted improvements for both noiseless and noisy measurement regimes.

Submitted February 4, 2012. Revised October 26, 2012. This work was supported by grants ONR N00014-08-1-1110, AFOSR FA9550-09-1-0422, FA9550-09-1-0643, and FA9550-05-0443, and by DARPA under the KECoM program. MFD was also supported by NSF Supplemental Funding DMS-0439872 to UCLA-IPAM, P.I. R. Caflisch. An early version of this work was presented in the International Conference on Sampling Theory and Applications, Singapore, May 2011 [1].

M. F. Duarte is with the Department of Electrical and Computer Engineering, University of Massachusetts, Amherst, MA 01003 USA (e-mail: mduarte@ecs.umass.edu)

S. Jafarpour is with Bina Technologies, Redwood City, CA 94065 USA (e-mail: sina@binatechnologies.com)

A. R. Calderbank is with the Departments of Computer Science, Electrical Engineering, and Mathematics, Duke University, Durham, NC 27708 USA (e-mail: robert.calderbank@duke.edu)

I. Introduction

In compressive sensing (CS) [2, 3], we wish to acquire a signal $f \in \mathbb{R}^{\mathcal{C}}$ by taking its product with a matrix $\Phi \in \mathbb{R}^{N \times \mathcal{C}}$ that we dub the *CS matrix*, obtaining a vector of *measurements* $y = \Phi f \in \mathbb{R}^N$. When $N \ll \mathcal{C}$, this acquisition scheme effectively compresses the signal f. Since in this case the signal recovery problem is ill posed, one must exploit prior information on the signal such as sparsity or compressibility. CS relies on the use of specially tailored signal recovery algorithms based on sparsity to estimate the signal f from the measurements y and the CS matrix Φ. Most work in CS relies on random constructions on the matrix Φ, i.e., the entries of the matrix are drawn independently from a suitable probability distribution such as Gaussian or Rademacher. Such matrices have been shown to provide enough information about a K-sparse signal f through the measurements y when $N = \mathcal{O}(K \log \mathcal{C})$.

The Delsarte-Goethals Frame [4] (DGF) was proposed as a deterministic CS matrix construction that enables efficient recovery of almost all sufficiently sparse signals without the use of randomness in the design of the measurement matrix. The DGF stands out due to its deterministic nature and to a set of accompanying fast signal recovery algorithms [4, 5]. It is also robust to noise, measurement loss, and other practical considerations [6]. However, the DGF must be applied directly on a sparse or compressible vector. That is, the vector f measured using the CS matrix Φ must be sparse or compressible in the canonical domain.

Recent results on average-case performance analysis of sparse recovery [5, 7, 8] have enabled formal performance guarantees for deterministic matrices such as the DGF. More specifically, since the DGF has a small spectral norm (due to being a tight frame) and small coherence (due to the properties of the DG code used to construct the DGF), it can be shown that most sparse signals can be recovered from their DGF measurements. This result relies on the fact that the coherence and spectral norm bounds imply that most sufficiently small column submatrices of the DGF are well conditioned. For applications where all supports of a sparse vector are equally likely, the aforementioned result immediately provides probabilistic performance guarantees [7–11]. However, the result does not explicitly describe the particular submatrices that do not meet such conditioning or, equivalently, the supports for sparse signals for which accurate recovery cannot be guaranteed.

While placing a uniform model over the signal's sparse support allows such recovery results to be valid with high probability, sparse signals from many real-world applications actually exhibit significant structure in the values and locations of the nonzero coefficients [12–14]. Thus, the suitability of the DGF for a particular application depends on whether there exists overlap between the structure present in the

sparse signals observed and the structure that may cause failures in CS recovery from DGF measurements.

In particular, initial results for deterministic compressive imaging (CI) via the DGF presented by Ni et al. [15, 16] show degraded performance compared to that previously published for generic CS. While most natural images are not themselves sparse or compressible, transforms such as the the 2-D discrete wavelet transform (DWT) are suitable to obtain compressible image representations. The DGF can then be applied directly to the wavelet coefficient vector of the image being acquired; this vector is obtained by sorting the coefficients of the 2-D DWT according to a particular raster scanning; it is common for this raster scanning to collect coefficients according to wavelet scale and sort them within each scale by the coordinates of the corresponding wavelet shift . It was hinted in [15] that the DGF may potentially have problems in CI due to the characteristic clustered structure that appears under such standard raster scanning of 2-D DWT coefficients. Ni et al. proposed a two-stage recovery algorithm that addresses these concerns by first estimating the nonzero coefficients concentrated in the coarsest scales of the wavelet transform using linear inversion, and then recovering the rest of the coefficients from the residual measurements using standard CS.

In this paper, we study the behavior of the DGF for such signals with clustered coefficient vectors in additional depth. Specifically, we show that the DGF performs suboptimally for recovery of sparse signals whose largest coefficients are concentrated (or clustered) within the coefficient vector.[1] We prove formally that there exist sparse vectors exhibiting such clustering within the null space of the DGF, which implies that there are clustered sparse vectors that are provably unrecoverable from DGF measurements. Such vectors can be obtained, for example, as the linear combination of two Haar wavelets of matching scale and appropriate shifts; note that this property is independent of the transform used in the compression (in particular, the 2-D DWT used in CI).

We also show that the clustered nature of these vectors in the null space of the DGF severely hampers the recovery of sparse vectors having their nonzero coefficients in a cluster. The 2-D wavelet coefficients of natural images are a prominent example of clustered vectors due to the large magnitude of coarse-scale wavelet coefficients and their concentration within a small portion of the coefficient vector.

To overcome the recovery issues in CI caused by this behavior, we show experimentally an improvement in performance afforded by altering the raster scanning of the coefficient vector so that the clustered structure of the nonzero coefficients is diluted. While a random raster scanning is the easiest approach, it

[1]It is natural to note that the fraction of sparse signals that are clustered is vanishingly small, which is part of the reason we are so interested in the DGF.

would negate many of the attractive features due to the deterministic nature of the DGF implementation. As an alternative, we propose a deterministic alternative raster scanning that is effective at dissipating the clusters that appear in standard raster scanning of 2-D wavelet coefficients, similar to that used in [17]. This modification provides substantially improved performance over the standard use of the DGF [15, 16]. We verify this improvement experimentally for recovery of natural images from both noiseless and noisy measurements.

This paper is organized as follows. Section II provides background and briefly reviews related work. Section III contains our analytical results on the null space of the DGF, and Section IV details compressive imaging as a relevant example application of our analysis. Section V presents validating experimental results, and Section VI concludes the paper.

II. Background and Related Work

A. The Delsarte-Goethals Frame

The Delsarte-Goethals set $DG(m,r)$, $m,r \in \mathbb{Z}$, $r < \frac{m-1}{2}$, is a set containing $2^{m(r+1)}$ binary symmetric matrices of size $m \times m$ with the property that the difference of any two distinct matrices has rank at least $m - 2r$.

The Delsarte-Goethals frame [4] (DGF) Φ is a measurement matrix of size $N = 2^m$ and $\mathcal{C} = 2^m R$, with $R \in [1, 2^{m(r+1)}]$ an integer. Its rows can be indexed using elements $x \in \mathbb{F}_2^m$, represented as binary vectors of length m. Similarly, its columns can be indexed using ordered pairs (P, b), where $P \in DG(m,r)$ and $b \in \mathbb{F}_2^m$. In this way, we label and define the entry of Φ in row x and column (P,b) as $\Phi_{P,b}(x) = i^{xPx^\top + 2bx^\top}$. Here $x^\top$ denotes the transpose of x. Note that all the arithmetic in the expressions $xPx^\top + 2bx^\top$ takes place in the ring of integers modulo 4, since the expression appears as an exponent for $i = \sqrt{-1}$. Given P, b, the vector $xPx^\top + 2bx^\top$ is a codeword in the Delsarte-Goethals code (defined over the ring of integers modulo 4). For a fixed matrix P, the 2^m columns $\{\Phi_{P,b}\}_{b \in \mathbb{F}_2^m}$ form an orthonormal basis Γ_P that can also be obtained by postmultiplying the Walsh-Hadamard basis by the unitary transform $\operatorname{diag}\left[i^{xPx^\top}\right]$. The DGF can then be written as a concatenation of bases $\Phi = [\Gamma_{P_1}\ \Gamma_{P_2}\ \ldots]$, with the matrix aspect ratio $R = \mathcal{C}/N$ determining the number of bases contained in Φ. Note that if we set r to its maximum value $r = \frac{m-1}{2}$, the $DG(m,r)$ code is equivalent to the second-order Reed-Muller code; therefore, the DGF obtained in this case is equivalent to the Reed-Muller sensing matrix of [5]. Additionally, since each $DG(m,r)$ code is a subset of the Reed-Muller code, each DGF is a column submatrix of the Reed-Muller sensing matrix.

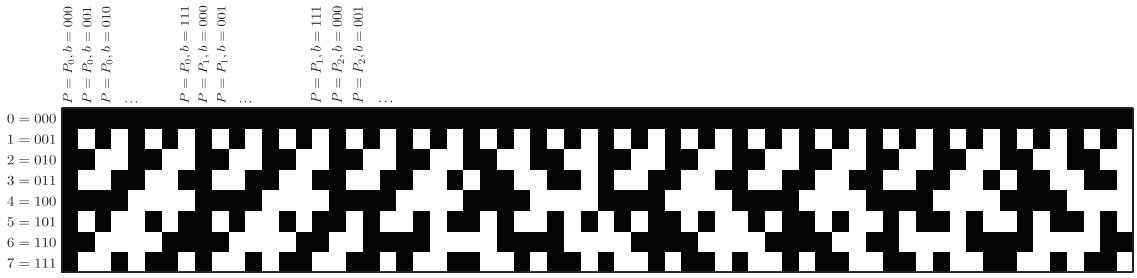

Fig. 1. *Example Delsarte-Goethals frame* Φ *with* $\mathcal{C} = 64$ *and* $N = 8$. *Each row is labeled by a binary vector* $x \in \mathbb{F}_2^3$, *and each column is labeled by a pair* (P, b), *with* $P \in DG(3, 1)$ *and* $b \in \mathbb{F}_2^3$.

The DGF is well suited for CS. In particular, when normalized to obtain unit-norm columns, its worst-case coherence value is $\mu(\Phi) = 2^{r-2}/\sqrt{N}$ and its spectral norm is $\|\Phi\|_2 = \mathcal{C}/N$ (since it is a tight frame). These two properties guarantee that the matrix enables successful recovery of an overwhelming majority of sparse signals [5, 7–10]. In the sequel, we also write $DGF(m, r)$ for the DGF when the dependence on m, r must be made explicit. Figure 1 shows an example DGF for $N = 8$ ($m = 3$) and $\mathcal{C} = 64$ ($R = 8$).

The DGF stands out due to its deterministic nature and to the fast signal recovery algorithms enabled by its construction [4, 5]. However, the DGF must be applied directly on a sparse or compressible vector. That is, the signal f measured using the CS matrix Φ must be sparse or compressible in the canonical domain. While many natural signals do not have this property, transforms used for signal compression are suitable to obtain sparsity or compressibility, with examples for images including the 2-D discrete cosine transform and the 2-D discrete wavelet transform.

B. The Haar Wavelet Basis

The Haar Wavelet Basis provides a set of piecewise constant functions of different support sizes and magnitudes. The first length-N Haar wavelet basis function (where $\mathcal{C}$ is a power of 2) has a constant value throughout its support; it is known as the *Haar scaling function* and denoted by

$$\phi(n) = \sqrt{1/\mathcal{C}},\ 0 \leq n \leq \mathcal{C} - 1.$$

The rest of the Haar wavelet basis functions have only two distinct nonzero values, each covering half of the function's support. The *mother Haar wavelet function* is defined as

$$\psi(n) = \begin{cases} \sqrt{1/\mathcal{C}} & 0 \le n < \mathcal{C}/2, \\ -\sqrt{1/\mathcal{C}} & \mathcal{C}/2 \le n < \mathcal{C}, \\ 0 & \text{otherwise.} \end{cases}$$

The Haar wavelet $\psi_{s,j}$ is labeled according to its scale $s = 0, \ldots, \log_2 \mathcal{C}-1$, and its offset $j = 0, \ldots, 2^s-1$. It is generated from the mother Haar wavelet through dilation and translation:

$$\psi_{s,j}(n) = \sqrt{2^s}\psi(2^s n - \mathcal{C}j).$$

Specifically, for scales $0 \le s < \log_2 \mathcal{C}$ and offsets $0 \le j < 2^s$, the Haar wavelet is defined as follows:

$$\psi_{s,j}(n) = \begin{cases} \sqrt{2^s/\mathcal{C}} & 2^{-s}\mathcal{C}j \le n < 2^{-s}\mathcal{C}(j+1/2), \\ -\sqrt{2^s/\mathcal{C}} & 2^{-s}\mathcal{C}(j+1/2) \le n < 2^{-s}\mathcal{C}(j+1), \\ 0 & \text{otherwise.} \end{cases}$$

This structure for the support of the wavelets (i.e., the location of its nonzero values) is known as a *dyadic structure*: the wavelet's support is clustered and of size $2^{-s}\mathcal{C}$, and the offset of the clustered support is a multiple of its size. For simplicity, we denote by $\mathcal{D}_{s,j}$ the set of indices in the dyadic interval at scale s and offset j:

$$\mathcal{D}_{s,j} = \{2^{-s}\mathcal{C}j, 2^{-s}\mathcal{C}j + 1, \ldots, 2^{-s}\mathcal{C}(j+1) - 1\}. \tag{1}$$

Using this notation, we can write

$$\psi_{s,j}(n) = \sqrt{2^s/\mathcal{C}} \left(\mathcal{I}_{\mathcal{D}_{s+1,2j}}(n) - \mathcal{I}_{\mathcal{D}_{s+1,2j+1}}(n) \right), \tag{2}$$

where $\mathcal{I}_{\mathcal{P}}(n)$ denotes the indicator function for n belonging in the set $\mathcal{P}$; thus the support of $\psi_{s,j}$ is $\mathcal{D}_{s,j}$. It is easy to check that the wavelets defined above have unit norm and are orthogonal to each other, building an orthonormal basis for $\mathbb{R}^{\mathcal{C}}$.

III. Clustered Sparse Vectors in the Null Space of the DGF

In this section, we study the behavior of the DGF when applied to clustered sparse coefficient vectors. Using properties of the Delsarte-Goethals codes, we identify small sets of columns that are linearly dependent and cancel out when added with appropriate signs. We will show that such sets of columns can be partitioned into two subsets of adjacent columns that, after adding the columns within each set and subtracting one set addition from the other, result in perfect cancellation. This, in turn, allows us to

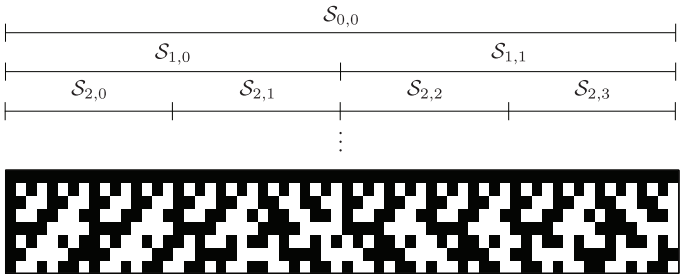

Fig. 2. *Example dyadic column intervals for a DGF of size* $\mathcal{C} = 64$, $N = 8$ *(*$m = 3$*).*

establish the presence of certain clustered sparse vectors in the null space of the DGF. We show that such vectors can be expressed as the superposition of pairs of Haar wavelets, which simplifies their notation.[2]

We begin by defining notation that will simplify the exposition. For each pair (s, j) with $s = 0, \ldots, \log_2 \mathcal{C} - 1$ and $j = 0, \ldots, 2^s - 1$, we let the set $\mathcal{S}_{s,j} \subseteq DG(m, r) \times \mathbb{F}_2^m$ denote the pairs (P, b) that generate the columns of the DGF with indices contained in the dyadic interval $\mathcal{D}_{s,j}$ defined in (1); we call $\mathcal{S}_{s,j}$ a *dyadic column interval*. Some examples are shown in Figure 2. We provide some properties of the column indices (P, b) contained in the subsets $\mathcal{S}_{s,j}$.

Proposition 1. *If* $s \geq \log_2 R$*, then* $\mathcal{S}_{s,j} = \{P_{s,j}\} \times \mathcal{F}_{s,j}$*, where* $P_{s,j}$ *is a fixed matrix in* $DG(m, r)$*, and* $\mathcal{F}_{s,j} = \{a_{s,j}\} \oplus \mathbb{F}_2^{\log_2 \mathcal{C} - s}$*, with* $a_{s,j} \in \mathbb{F}_2^{s - \log_2 R}$ *and* $\oplus$ *denoting pairwise concatenation.*

In words, $\mathcal{F}_{s,j}$ is a subset of $\mathbb{F}_2^m$ whose elements share the $s - \log_2 R$ most significant bits; fluctuations on later bits span the subset. Thus, the subset $\mathcal{S}_{s,j}$ is defined by the matrix $P_{s,j}$ and the "header" $a_{s,j}$ containing the fixed most significant bits of b over $\mathcal{S}_{s,j}$.

We now consider the linear combination of two Haar wavelets projected by the DGF at the same scale $\eta_{s,j_1,j_2,\pm}(x) = \Phi(\psi_{s,j_1} \pm \psi_{s,j_2})$. When this combination is applied to the DGF Φ, the outcome $\eta_{s,j_1,j_2,\pm}(x)$ will correspond to the difference between the sums of columns in two distinct dyadic column subsets. We would like to determine determine which pairs of wavelet superpositions $\psi_{s,j_1} \pm \psi_{s,j_2}$ belong in the null space of Φ, $\mathcal{N}(\Phi)$, by determining values of s, j_1, and j_2 that render the combination $\eta_{s,j_1,j_2,\pm}(x) = 0$ (either for an addition or a subtraction). Each such combination identifies for us a set of $2^{-s+1}\mathcal{C}$ linearly dependent columns in the DGF Φ. The following theorem provides our main result and is proven in Appendix A.

[2]Note that the use in this section of Haar wavelets to express vectors in the null space of the DGF is completely independent of the particular transform used for sparsity in CS; in particular, it is not related to the choice of DWT used to achieve sparsity in CI.

Theorem 1. *Let Φ denote the $DGF(m,r)$ and write $x_{h,l}$ and $x_{f,l}$ for the first and last l entries of the vector $x \in \mathbb{F}_2^m$, $1 \le l \le m$, respectively. For $0 \le s < \log_2 \mathcal{C}$ and $0 \le j_1, j_2 < 2^s$, $j_1 \ne j_2$, we have $\eta_{s,j_1,j_2,-} = 0$ (i.e., $\psi_{s,j_1} - \psi_{s,j_2} \in \mathcal{N}(\Phi)$) if*

- *$s < \log_2 R$ (for all j_1, j_2);*
- *$\log_2 R \le s \le \log_2 R + 2r$, $a_{s,j_1} = a_{s,j_2}$ (that is, the two dyadic column intervals are at the same position within the the bases $\Gamma_{P_{s,j_1}}$, $\Gamma_{P_{s,j_2}}$ containing them) and $x(P_{s,j_1} - P_{s,j_2})x^\top = 0 (\mathrm{mod}\ 4)$ for all $x \in \mathbb{F}_2^m$ such that $x_{f,\log_2 \mathcal{C}-s} = 100\ldots0$; or*
- *$s \ge \log_2 R + 2r$ and $x(P_{s,j_1} - P_{s,j_2})x^\top = 2(a_{s,j_2} - a_{s,j_1})x^\top_{h,s-\log_2 R} \pmod 4$ for all $x \in \mathbb{F}_2^m$ such that $x_{f,\log_2 \mathcal{C}-s} = 100\ldots0$.*

The conditions also hold for $\eta_{s,j_1,j_2,+} = 0$ by adding a term of 2 in the equalities $(\mathrm{mod}\ 4)$ involving x.

The theorem shows that there exist vectors expressed as the sum or difference of two Haar wavelets that are part of the null space of the DGF. These vectors have sparsity $K = 2^{-s+1}\mathcal{C}$, where s is the scale of the two wavelets involved. The proof of Theorem 1 can be extended to real-valued adaptation using (i) zero-diagonal DG matrices P, as used in [15], or (ii) by generating a matrix from the DGF with twice as many rows via the Gray map; see Appendix B for details.

IV. Application: Compressive Imaging

Compressive imaging has emerged as one of the most popular applications of CS, in large part due to the significant amount of compressibility and structure in known sparse representations for natural images, as well as to the feasibility of imaging architectures that compute compressive measurements without requiring the capture of all individual pixels in the image. Options include optical modulators [18] and integrated imaging sensor and analog computation modules [19, 20]; see [14] for a survey.

Perhaps the most popular compressible representation for CI is given by the discrete wavelet transform (DWT). DWT coefficient vectors for natural images exhibit significant structure on the location of the largest-magnitude coefficients within the vector. While our description will focus on the 1-D DWT, the concepts extend to 2-D images in a straightforward fashion. The DWT coefficients are usually raster scanned first by scale, and then by offset, to build a wavelet coefficient vector f_w:

$$f_w := [s_{0,0}\ w_{0,0}\ w_{1,0}\ w_{1,1}\ w_{2,0}\ \ldots\ w_{3,0}\ \ldots]. \tag{3}$$

Here $s_{0,0}$ denotes the scaling coefficient and $w_{s,j}$ denotes the DWT coefficient at scale s and offset j. When a partial-level DWT transform is used, the raster scan begins with the multiple scaling coefficients stacked into a vector, followed by the DWT coefficients ordered as in (3).

A. *Effect of Wavelet Raster Scanning on Vector Structure*

DWT transforms of piecewise smooth signals and natural images have several properties that affect the configuration of their raster-scanned representations. First, the coefficients decay in magnitude as the scale increases, resulting in a concentration of the largest coefficients in the beginning of the vector. Furthermore, the number of 1-D and 2-D DWT coefficients of an image at scale s is equal to 2^{s-1} and $3 \cdot 4^{s-1}$, respectively, implying that the largest coefficients of the signal are clustered within a very small fraction of the coefficient vector.

A study by Ni et al. [15] of the performance of deterministic CI via a DGF shows that the performance is poor when compared to that previously published for generic CS. Leveraging the properties of DWT representation vectors, [15] proposed a two-stage recovery algorithm that improves the performance of CI via the DGF [15]. The first stage estimates the N coefficients at coarsest scales using a submatrix projection that ignores the contribution to the measurements from the rest of the coefficient vector. Due to their location at the beginning of the coefficient vector and the structure of the DGF, this procedure entails a simple projection into the basis Γ_{P_1}. The second stage subtracts the contribution to the measurements of the estimates for the first N coefficients and estimates the remainder of the coefficient vector using a greedy sparse signal recovery algorithm. The incremental nature of the proposed estimation algorithm reduces the number of nonzeros of the signal being estimated, resulting in an improvement over standard CS recovery of the full coefficient vector.

In light of the result of Theorem 1, our contention is that the recovery performance from CS measurements via the DGF suffers due to the clustered structure arising in the DWT coefficient vector. The DWT coefficients for the coarsest scales, which contain the largest magnitudes, are all clustered within a small contiguous portion of the coefficient vector. The projections of this structured vector into the sparse clustered vectors in the null space of the DGF with overlapping supports characterize the loss of information in the coarsest-scale coefficients, as well as the introduction of spurious nonzero coefficient estimates for the corresponding matching support of the null space vector elsewhere in the coefficient vector, and most likely in the finest scales of the DWT. In practice, this loss of information is perceived mostly by the presence of noticeable artifacts of small scale due to the aforementioned spurious nonzero coefficients; the effect of the information loss in the coarsest scale coefficients is not as visibly noticeable, but is manifested via metrics like the recovery signal-to-noise ratio (SNR) or peak SNR.

B. Alternative Raster Scannings

Under our contention, one can expect to see an improvement in the performance of the two-stage algorithm of [15] compared to standard recovery algorithms, given that the first stage of the algorithm in [15] estimates and removes the contribution of the clustered significant coefficients in the image measurements. Once the cluster in the vector to be recovered is dissipated in this fashion, the rest of the vector can be estimated accurately using standard sparse recovery algorithms.

In view of our results, we conjecture that the same improvements in performance can be obtained simply by dissipating the cluster present in the signal's coefficient vector without manipulating the signal to reduce its level of sparsity K. In fact, we may observe better performance in the estimation of the coarse scale coefficients, as the estimates obtained by the algorithm of [15] suffer from interference from the nonzero coefficients that lie outside of the coarsest scales.

A simple approach to alleviate the issues of the DGF with clustered sparse and compressible coefficient vectors is to modify the raster scanning of the coefficients so that the clustered structure is dissipated. It is easy to show that a random raster scanning will dissipate the clustering with high probability; such scrambling or permutation has been previously leveraged in the CS literature [17, 21–23]. In CS practice, such permutation can be obtained by switching the DGF during sensing to a new sensing matrix that features the same columns as the DGF but collected in an arbitrary order. However, a random raster scanning is not completely compatible with the DGF, as it nullifies several useful properties of its deterministic nature, such as scalability in computation to the signal length and implementation simplicity.

As an alternative, we introduce a deterministic raster scanning that evenly spreads the content of coefficient clusters and scales across the bases Γ_{P_i} composing the DGF. We start from the standard raster scanning for DWT coefficients and apply a permutation described by the mapping

$$p(n) = (n \bmod R)N + \lfloor n/R \rfloor, \tag{4}$$

which provides a bijective mapping of the set $\{0, \ldots, \mathcal{C}-1\}$ onto itself. In words, the mapping takes each set of consecutive coefficients of size R and spreads it across the R bases $\{\Gamma_{P_i}\}$; the subsequent blocks are partitioned similarly and ordered lexicographically. The resulting raster scanning separates the 2-D DWT coefficients at a given scale and adjacent offsets by a distance of N entries within the coefficient vector. Furthermore, neighboring coefficients in this raster scan either correspond to different scales or to DWT coefficients for wavelets with offsets whose difference is equal to $R = \mathcal{C}/N$. This partitioning is similar to the linear congruential permutation of [17], except that in (4) the separation between adjacent coefficients N must be a divisor of the length of the coefficient vector $\mathcal{C}$ (i.e., $R = \mathcal{C}/N$), while [17]

requires the separation and the coefficient vector length to be relative primes. Additionally, and in a similar fashion to the earlier random permutation, this deterministic permutation is applied in CS practice by replacing the DGF with a second deterministic matrix obtained by permuting the DGF columns in the prescribed fashion; note that no randomness is leveraged in this case in the sensing matrix design.

V. Experimental results

In this section, we provide experimental evidence to verify the conditions of the DGF null space presented in Section III. First, we experimentally verify the presence of the aforementioned combinations of Haar wavelets in the null space of the DGF. Second, we study the performance of CS via DGF for clustered and non-clustered signals. Third, we study the performance of the DGF for compressive imaging and the effects of random and deterministic permutation in performance, and contrast the improvements against those obtained by alternative approaches from the literature.

A. Pairwise combinations of Haar wavelets in the DGF null space

To verify the conditions given in Theorem 1, our first set of experiments consists of an exhaustive search for pairs of Haar wavelets ψ_{s,j_1} and ψ_{s,j_2} that result in the same measurement vector (up to sign flipping, $y_{s,j_1} = \pm y_{s,j_2}$) after projection via the $DGF(m,r)$ for $r = 1$ and $m = 7, 9, 11, 13$. The corresponding matrices Φ vary from $M = 128$ to 8192 rows. In each case, we identify pairs of wavelets that provide the same measurement vector (up to sign flipping), and tabulate the pairs according to the number of nonzeros (i.e., sparsity level) K of their difference $\eta_{s,j_1,j_2,\pm}(x) = \Phi(\psi_{s,j_1} \pm \psi_{s,j_2}) \in \mathcal{N}(\Phi)$. Tables I to IV show the tabulation for different values of K, $N = 2^m$, and $\mathcal{C}$. We omit data for values of $K > 2N$ (e.g. for values of $s < \log_2 \mathcal{C}/N$), since all combinations of pairs of wavelets of the same scale return a zero-valued measurement vector.

There are several patterns to note in the data collected in the table; we focus on Table I for simplicity when possible.

- For $K = 2N$ (i.e., each vector $\eta_{s,j_1,j_2,\pm} = y_{s,j_1} \pm y_{s,j_2}$ has support covering two bases Γ_{P_1} and Γ_{P_2}), the number of vectors in the null space of Φ is equal to $\binom{\mathcal{C}/N}{2}$. In words, each pairwise difference of Haar wavelets at scale $s = \log_2 \mathcal{C}/N$ is contained in the nullspace $\mathcal{N}(\Phi)$, This implies that the first statement of Theorem 1 is valid for each value of $s = \log_2(C/K)$ and for all pairs of DG matrices (P_1, P_2) involved in the DGF Φ.
- In several cases, the number of vectors in the null space for $K' < 2N$ is twice the number for $K = 2K'$, meaning that the second statement of Theorem 1 remains valid for each pair of DG

K	$N/8$	$N/4$	$N/2$	N	$2N$
$\mathcal{C} = 2N$	0	8	4	2	1
$\mathcal{C} = 4N$	0	48	24	12	6
$\mathcal{C} = 8N$	0	96	112	56	28
$\mathcal{C} = 16N$	0	192	480	240	120
$\mathcal{C} = 32N$	256	896	1984	992	496
$\mathcal{C} = 64N$	512	1792	3968	4032	2016
$\mathcal{C} = 128N$	1024	7680	16128	16256	8128

TABLE I

COUNT OF VECTORS $\psi_{s,j_1} \pm \psi_{s,j_2}$ IN THE NULL SPACE OF THE DGF Φ WITH $N = 128$ ROWS AND VARYING NUMBER OF COLUMNS N. THE VECTORS ARE TABULATED BY THEIR SPARSITY $K = 2^{1-s}\mathcal{C}$.

matrices (P_1, P_2) and for all possible values of $a_{s,j}$, $s = \log_2(\mathcal{C}/K) + 2$.

- For each value of N and $\mathcal{C}$, this doubling relationship eventually breaks down when K is sufficiently small (i.e., s is sufficiently large). Our conjecture here is that the condition on the pair of matrices (P_1, P_2) in the second statement of Theorem 1 stops being valid for an increasing number of pairs as the condition on the null space of the matrix difference $(P_1 - P_2)$ becomes more stringent. In fact, we were not able to find any null space vectors with sparsity $K < N/8$ for the matrices we tested.
- The particular choice of matrices $P \in DG(m, r)$ involved in the DGF Φ can reduce the number of vectors $\eta_{s,j_1,j_2,\pm}(x)$ that is contained in the null space for each value of s. For example, notice in Table I that when up to 16 DG matrices are used to build the DGF Φ, there are no vectors of sparsity $K = N/8 = 16$ in the null space; however, when the number of DG matrices used goes to 32, there appear 256 null space vectors with sparsity $K = 16$. Our conjecture is that since each pair of matrices (P_1, P_2) that hold the second condition of Theorem 2 would give rise to 16 null space vector combinations, there should be 16 pairs of matrices (P_1, P_2) within the 32 DG matrices used that have the aforementioned conditions. In fact, each pair of wavelets found in this case contains one wavelet supported in the first half of the vector and another supported on the second half. The number of null space vectors also doubles each time that $\mathcal{C}$ doubles as well.

K	$N/4$	$N/2$	N	$2N$
$\mathcal{C} = 2N$	0	4	2	1
$\mathcal{C} = 4N$	0	24	12	6
$\mathcal{C} = 8N$	32	112	56	28
$\mathcal{C} = 16N$	64	224	240	120

TABLE II

COUNT OF VECTORS $\psi_{s,j_1} \pm \psi_{s,j_2}$ IN THE NULL SPACE OF THE DGF Φ WITH $N = 512$ ROWS AND VARYING NUMBER OF COLUMNS N. THE VECTORS ARE TABULATED BY THEIR SPARSITY $K = 2^{1-s}\mathcal{C}$.

K	$N/8$	$N/4$	$N/2$	N	$2N$
$\mathcal{C} = 2N$	16	8	4	2	1
$\mathcal{C} = 4N$	32	48	24	12	6
$\mathcal{C} = 8N$	64	224	112	56	28
$\mathcal{C} = 16N$	128	960	480	240	120

TABLE III

DISTRIBUTION OF VECTORS $\eta_{s,j_1,j_2,\pm}$ IN THE NULL SPACE OF THE DGF Φ WITH $N = 128$ ROWS AND VARYING NUMBER OF COLUMNS N. THE VECTORS ARE TABULATED BY THEIR SPARSITY $K = 2^{1-s}\mathcal{C}$.

B. Effect of clustered structure on recovery performance

The second sets of experiments studies the performance of CS recovery from DGF measurements for clustered sparse signals. We use the ℓ_1-norm minimization algorithm from the ℓ_1-magic toolbox [24] and test the DGF with $N = 128$ and several values of $\mathcal{C}$ and K. For each combination of parameters we execute 100 iterations for two different sparse signal models. A *standard* sparse signal model selects the support of the signal (i.e., the indices for its K nonzero values) uniformly at random from all $\binom{N}{K}$ possibilities. In contrast, a *clustered* signal model selects the support of the signal via a two-stage approach: first, we select uniformly at random one of the clustered supports for the signals of sparsity $K_S = 32$ that is contained in the null space, as identified in Section V-A; second, we select the K indices of the nonzero uniformly at random from all the $\binom{K_S}{K}$ subsets of the support chosen in the previous step. In both cases, the nonzero values of the signal are then drawn independently from a $\mathcal{N}(0, 1)$ standard normal distribution.

We compute the average normalized recovery error for each pair of values $(K, \mathcal{C})$ and for each signal model, in order to evaluate the effect of clustering in CS performance. The results, shown in Fig. 3,

K	$N/8$	$N/4$	$N/2$	N	$2N$
$\mathcal{C} = 2N$	16	8	4	2	1
$\mathcal{C} = 4N$	32	48	24	12	6
$\mathcal{C} = 8N$	64	224	112	56	28

TABLE IV

DISTRIBUTION OF VECTORS $\eta_{s,j_1,j_2,\pm}$ IN THE NULL SPACE OF THE DGF Φ WITH $N = 512$ ROWS AND VARYING NUMBER OF COLUMNS N. THE VECTORS ARE TABULATED BY THEIR SPARSITY $K = 2^{1-s}\mathcal{C}$.

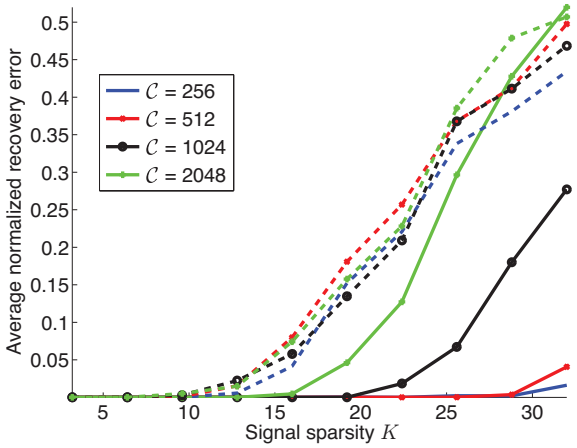

Fig. 3. *Performance of CS recovery from DGF measurements for signals with and without clustered structure for* $N = 128$ *and several values of* $\mathcal{C}$ *and* K*. The solid lines represent the average performance for signals drawn randomly from a* standard *(unstructured) sparsity model, while the dashed lines represent the average performance for signals drawn randomly from a* clustered *sparse model. There is a noticeable gap in performance between the two models; the gap vanishes as* $\mathcal{C}$ *increases due to the fixed value for* N*, and recovery failing for all (clustered and standard) sparse signals.*

display the performance gap between the two models for the same signal sparsity K. For example, in the case where $\mathcal{C} = 512$, we have accurate recovery of signals from the standard model with sparsity up to $K = 29$; in contrast, signals from the clustered model must have sparsity $K \leq 13$ to achieve exact recovery. Similarly, the average normalized error for recovery of signals of sparsity $K = 32$ and length $N = 512$ jumps from 0.04 for the standard sparse model to 0.5 for the clustered sparse model.

The gap in performance between the two sparse signal models becomes smaller as the value of $\mathcal{C}$ increases; this is due to the fact that we have fixed the number of measurements of N, which reduces the maximum sparsity K for which we can still achieve accurate recovery for the uniformly distributed support

signal model. In contrast, the recovery performance for clustered sparse signals stays roughly constant as N increases, signifying that the sparse signal structure, rather than the number of measurements taken, is the core reason for the decreased recovery performance observed for the clustered signal model.

C. Compressive imaging

In our third set of experiments, we validate the improvements in performance of deterministic CI afforded by the raster scannings proposed in Section IV-B. Our experiments use the MRI image of size $\mathcal{C} = 512 \times 512$ used in [15]. We use the Daubechies-8 DWT to obtain compressible coefficients for the image, and we set the number of measurements to $N = 65536$, providing a compression ratio of 25%. We test basis pursuit (BP) [25] solved via ℓ_1-magic [24], iterative hard thresholding (IHT) [26], and the two-step approximation (TSA) algorithm from [15] with BP for the second stage, all using the real-valued $DGF(16, 0)$. We also test BP and IHT under the proposed raster scannings described in Section IV-B. In all cases, the algorithms (both BP and IHT) are set to converge once the normalized magnitude of the residual is sufficiently small, i.e., $\|y - \Phi f\|_2/\|y\|_2 < 10^{-3}$; in BP, this is specified as an inequality constraint on f. First, we compare the quality and computational cost of the different approaches, measured by signal-to-noise ratio (SNR) of the reconstruction and computation time; these experiments were performed in a computer with a Xeon X5680 CPU and 8GB of memory, running Matlab for Linux, and using a single core. The results, shown in Table V, show that under similar computational cost, the proposed matrix designs provide significant improvement in the quality of recovery. Representative example recovered images are shown in Figure 4. We also compare the quality of recovery for the different raster scanning approaches when independent and identically distributed Gaussian noise of variance σ^2 is added to the measurements for several values of σ. The resulting SNRs, shown in Table VI, show once again that the proposed raster scannings significantly improve the performance of image recovery.

To test the generality of the improvements in CI, we perform a comparison of the performance of the DGF with and without random column permutation on a subset of the "Miscellaneous" natural image dataset available at [27]. All images are of size 512×512 and we use the real-valued $DGF(16, 0)$ to obtain $N = 65536$ compressive measurements. We use the ASP solver for BP [28]; the algorithm is set to converge once the relative magnitude of the measurement-domain error is sufficiently small, i.e., $\|y - \Phi \widehat{f}\|_2 < \|y - \Phi f_K\|_2$, where f_K denotes the optimal K-sparse approximation of the image's wavelet coefficients f with $K = \mathcal{C}/10$. Table VII shows the performance, measured by the recovery SNR, for the different images without and with a random permutation being applied to the columns of the DGF. Except for a single example, all images exhibit similar or improved recovery performance when a random

Algorithm	SNR (dB)	Time (s)
TSA	21.74	1008
BP + Random Raster	23.60	820
IHT + Random Raster	22.25	804
BP + Deterministic Raster	23.52	822
IHT + Deterministic Raster	20.98	813

TABLE V

PERFORMANCE AND COMPUTATIONAL COST OF SIGNAL RECOVERY ALGORITHMS. THE USE OF THE PROPOSED RASTER SCANNINGS SIGNIFICANTLY IMPROVES IMAGE RECOVERY PERFORMANCE.

Algorithm	$\sigma = 0.01$	$\sigma = 0.05$	$\sigma = 0.1$
TSA	21.64	19.51	16.37
BP + Random Raster	23.41	20.71	18.18
IHT + Random Raster	22.24	20.82	17.20

TABLE VI

PERFORMANCE OF SIGNAL RECOVERY ALGORITHMS FOR NOISY MEASUREMENTS MEASURED IN SNR (DB). THE USE OF RANDOM RASTER SCANNING SIGNIFICANTLY IMPROVES THE PERFORMANCE UNDER THE NOISY MEASUREMENT REGIME.

permutation is applied to the columns of the DGF, emphasizing the role of structure in the coefficient vector in the performance of CS via the DGF.

VI. CONCLUSIONS AND FURTHER WORK

In this paper, we have provided analytical and empirical evidence on the structure of the null space of the Delsarte Goethals frame (DGF). Specifically, we have shown the existence of highly-structured, clustered sparse vectors in $\mathcal{N}(\Phi)$, and demonstrated the effect of such structure in the null space on recovery performance from DGF measurements for clustered sparse signals. We also connected our analytical results with a previously-observed performance gap when the DGF is used in CI applications, and showed that this gap can be closed by dissipating the clustered structure that is common on 2-D wavelet coefficient vectors of natural images under commonly used raster scannings.

While the results we show here might at first make a case against the use of the DGF in CS, we highlight the fact that the number of sparse signal supports for which the described behavior appears is significantly smaller than the total number of possible supports; this in turn agrees with prior results on

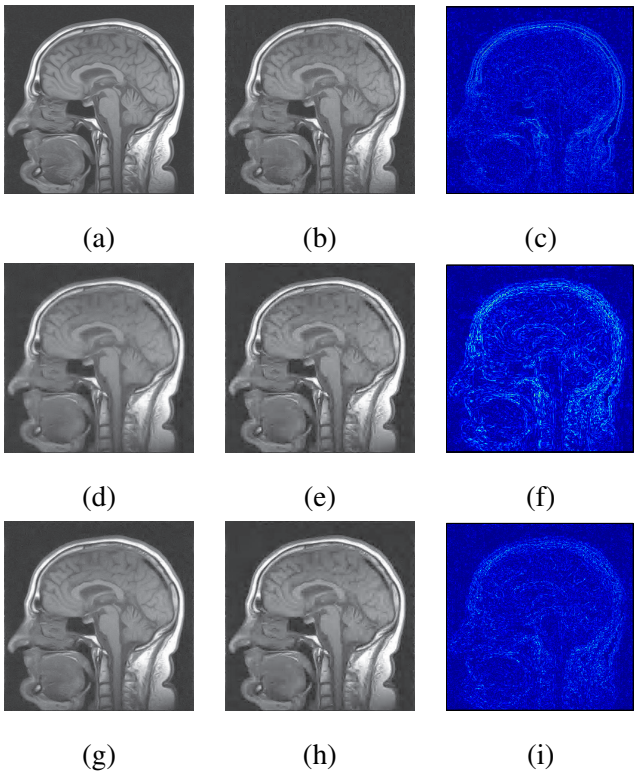

Fig. 4. Comparison of CS recovery performance of a $\mathcal{C} = 512 \times 512$ image from $N = 65536$ DGF measurements. (a) Original image. (b) Two-step approximation [15]. (c) Difference image of (a), (b). (d-e) BP and IHT with standard DGF. (f) Difference image of (a), (d). (g-h) BP and IHT using random raster scanning. (i) Difference image of (a), (g).

the successful performance of DGF measurements in CS for the overwhelming majority of sufficiently sparse signals [4, 5]. In a sense, this combined suitability for most sparse signals is what has rendered the DGF of such interest to the CS community.

We have additionally proposed possible coefficient permutation techniques that dissipate the presence of clusters, which is effectively equivalent to permuting the columns of the original DGF and provides an alternative deterministic CS matrix for applications featuring clustering in sparse vectors. Our results also highlight the opportunity for the design of deterministic CS matrices that may be suited to particular types of structure. Examples include tree sparsity, clustered sparsity, and refractory sparsity [12–14, 29].

Finally, we note that while our analysis was restricted to clustered signals, it may be possible to construct arbitrary sparse vectors in the null space of the DGF using more sophisticated methods, such as specially-designed code indicator vectors that vanish under a carefully constructed linear projection. Such methods have the promise of providing designs for new coding theory-inspired deterministic CS matrices for which we can provide a complete null space characterization. This in turn would enable a

Image Name	No Perm.	Rand. Perm.	Improvement
Cameraman	25.80	28.93	3.13
House	33.21	41.67	8.46
Airplane (F-16)	23.95	21.59	-2.36
Sailboat on Lake	27.08	29.18	2.10
Girl (Lena)	39.78	47.18	7.40
Couple	28.04	31.64	3.60
Mandrill	17.70	18.52	0.82
Peppers	17.48	17.71	0.23
Man	31.09	45.61	14.52
Bridge	33.97	37.44	4.47
Girl (Tiffany)	12.43	12.36	-0.07

TABLE VII

PERFORMANCE OF COMPRESSIVE IMAGING VIA DGF WITH NO PERMUTATION AND RANDOM PERMUTATION OF THE IMAGE'S WAVELET COEFFICIENTS. IN ALMOST ALL CASES, THE PERFORMANCE OF RANDOM PERMUTATION IS SIMILAR OR IMPROVED OVER THE BASELINE APPROACH.

more precise study of recovery performance from deterministic CS measurements. We leave this line of research for future work.

APPENDIX A
PROOF OF THEOREM 1

A. Dyadic Column Sums of the DGF

We begin by considering the projections of an individual Haar wavelet $y_{s,j} = \Phi\psi_{s,j}$, which can be rewritten using (2) as

$$y_{s,j}(x) = \sqrt{2^{s+1}/\mathcal{C}} \left(\sum_{(P,b)\in\mathcal{S}_{s+1,2j}} \Phi_{P,b}(x) - \sum_{(P,b)\in\mathcal{S}_{s+1,2j+1}} \Phi_{P,b}(x) \right). \quad (5)$$

Thus, we focus on the sums of columns of the DGF Φ over dyadic column intervals, which we term *dyadic sum* and denote as

$$S_{s,j}(x) := \sum_{(P,b)\in\mathcal{S}_{s,j}} \Phi_{P,b}(x). \quad (6)$$

This new notation simplifies (5) to

$$y_{s,j}(x) = \sqrt{2^{s+1}/\mathcal{C}}(S_{s+1,2j}(x) - S_{s+1,2j+1}(x)).$$

We begin by stating some properties of these subsets $\mathcal{S}_{s,j}$.

Proposition 2. *If $s > \log_2(\mathcal{C}/N) = \log_2 R$, then $\mathcal{S}_{s,j} = \{P_{s,j}\} \times \mathcal{F}_{s,j}$, where $P_{s,j}$ is a fixed matrix in $DG(m,r)$, and $\mathcal{F}_{s,j} = a_{s,j} \oplus \mathbb{F}_2^{\log_2 \mathcal{C} - s}$, with $a_{s,j} \in \mathbb{F}_2^{s-\log_2 R}$.*

In words, $\mathcal{F}_{s,j}$ is a subset (and a subgroup) of $\mathbb{F}_2^m$ whose elements share the $s - \log_2 R$ most significant bits; fluctuations on later bits span the subset. Thus, for $s \geq \log_2(\mathcal{C}/N) = \log_2 R$, the subset $\mathcal{S}_{s,j}$ is defined by the matrix $P_{s,j}$ and the "header" $a_{s,j}$ containing the fixed most significant bits of b over $\mathcal{S}_{s,j}$.

Proposition 3. *If $s \leq \log_2(\mathcal{C}/N) = \log_2 R$, then $\mathcal{S}_{s,j} = \mathcal{P}_{s,j} \times \mathbb{F}_2^m$, where $\mathcal{P}_{s,j}$ is a subset of $DG(m,r)$ containing $2^{-s}\mathcal{C}$ +matrices.*

Proof sketch. Since the offset and size of the interval $\mathcal{S}_{s,j}$ is a multiple of $N = 2^m$, the selected columns include the set of bases Γ_P for a subset of matrices P used in the construction of the DGF (as defined in Section II-A). □

Armed with this properties, we can prove the following results on $S_{s,j}(x)$. The first lemma states that almost all dyadic sums at coarse scales vanish.

Lemma 1. *For $s \leq \log_2(\mathcal{C}/N)$,*

$$S_{s,j}(x) = \begin{cases} |\mathcal{S}_{s,j}| & x = 0, \\ 0 & x \neq 0. \end{cases}$$

Proof. We write

$$\begin{aligned} S_{s,j}(x) &= \sum_{(P,b)\in\mathcal{S}_{s,j}} \Phi_{P,b}(x) = \sum_{P\in\mathcal{P}_{s,j}} \sum_{b\in\mathbb{F}_2^m} i^{xPx^\top + 2bx^\top} \\ &= \sum_{P\in\mathcal{P}_{s,j}} i^{xPx^\top} \sum_{b\in\mathbb{F}_2^m} i^{2bx^\top}. \end{aligned}$$

For $x \neq 0$ the second sum is equal to zero, rendering $S_{s,j}(x) = 0$. When $x = 0$, the second sum is equal to N and the first sum is equal to $|\mathcal{P}_{s,j}|$, and so we have $S_{s,j}(x) = |\mathcal{P}_{s,j}|N = |\mathcal{S}_{s,j}|$, proving the lemma. □

The next lemma states that for finer scales many dyadic sums vanish and the nonzero sums have a fixed magnitude.

Lemma 2. *For $s \geq \log_2(\mathcal{C}/N)$, $S_{s,j}(x)$ can only take one of these values: $\{0, 2^{-s}\mathcal{C}, -2^{-s}\mathcal{C}\}$.*

Proof. We write

$$\begin{aligned} S_{s,j}(x) &= \sum_{(P,b)\in\mathcal{S}_{s,j}} \Phi_{P,b}(x) = \sum_{b\in\mathcal{F}_{s,j}} i^{xP_{s,j}x^\top + 2bx^\top} \\ &= i^{xP_{s,j}x^\top} \sum_{b\in\mathbb{F}_2^{\log_2 \mathcal{C}-s}} i^{2a_{s,j}x_{h,s-\log_2 R} + 2bx^\top_{f,\log_2 \mathcal{C}-s}} \qquad (7) \\ &= i^{xP_{s,j}x^\top + 2a_{s,j}x^\top_{h,s-\log_2 R}} \sum_{b\in\mathbb{F}_2^{\log_2 \mathcal{C}-s}} i^{2bx^\top_{f,\log_2 \mathcal{C}-s}}. \qquad (8) \end{aligned}$$

In (7) we partitioned the vector x into its first $s - \log_2 R$ bits and its last $\log_2 \mathcal{C} - s$ bits so that we can factor out the component corresponding to the "header" $a_{s,j}$; note that $s - \log_2 R + \log_2 \mathcal{C} - s = \log_2 \mathcal{C} - \log_2(\mathcal{C}/N) = \log_2 N = m$ so that all bits are used. For $x_{f,\log_2 \mathcal{C}-s} \neq 0$, the sum in (8) is equal to zero, rendering $S_{s,j}(x) = 0$. When $x_{f,\log_2 \mathcal{C}-s} = 0$, the sum is equal to $2^{-s}\mathcal{C}$. Finally, note that the exponent on the first term is even and yields ± 1, proving the lemma. □

We are finally ready to prove Theorem 1. We begin by writing

$$\begin{aligned} \eta_{s,j_1,j_2,\pm}(x) &= y_{s,j_1} \pm y_{s,j_2} \\ &= \sqrt{\frac{2^{s+1}}{\mathcal{C}}} \left[(S_{s+1,2j_1}(x) - S_{s+1,2j_1+1}(x)) \right. \\ &\quad \left. \pm (S_{s+1,2j_2}(x) - S_{s+1,2j_2+1}(x)) \right]. \qquad (9) \end{aligned}$$

Lemmas 1 and 2 show that the value of a dyadic sum of a given scale can take only two or five distinct values, respectively, depending on the scale s used. When $s < \log_2(\mathcal{C}/N)$, Lemma 1 provides $\eta_{s,j_1,j_2,\pm}(x) = 0$ for all $j_1,\ j_2$, and $x \in \mathbb{F}_2^m$; this gives the first condition in the theorem. When $s \geq \log_2(\mathcal{C}/N)$, the sums from (8) involved in (9) vanish for all x with $x_{f,\log_2 \mathcal{C}-s} \neq 0$, and so we have

$$\eta_{s,j_1,j_2,\pm}(x) = 0 \text{ if } x_{f,\log_2 \mathcal{C}-s} \neq 0.$$

If $x_{f,\log_2 \mathcal{C}-s-1} = 0$, all terms inside the sum in (8) for each of the four sums S involved in (9) are equal to one and so each of the sums in (8) are equal to $2^{-s-1}\mathcal{C}$. With this information, denoting bit b of x by x_b, plugging (8) in (9), and factoring out all common bits from the two headers $a_{s+1,2j_i}, a_{s+1,2j_i+1}$

for $i = 1, 2$, we obtain

$$\begin{aligned}\eta_{s,j_1,j_2,\pm}(x) &= \sqrt{2^{-s-1}\mathcal{C}}\left[i^{xP_{s,j_1}x^\top + 2a_{s,j_1}x_{h,s-\log_2 R}^\top}\left((-1)^{0\cdot x_{s-\log_2 R+1}} - (-1)^{1\cdot x_{s-\log_2 R+1}}\right)\right.\\ &\quad \left.\pm i^{xP_{s,j_2}x^\top + 2a_{s,j_2}x_{h,s-\log_2 R}^\top}\left((-1)^{0\cdot x_{s-\log_2 R+1}} - (-1)^{1\cdot x_{s-\log_2 R+1}}\right)\right]\\ &= \begin{cases}\sqrt{2^{-s+1}\mathcal{C}}\left[i^{xP_{s,j_1}x^\top + 2a_{s,j_1}x_{h,s-\log_2 R}^\top} \pm i^{xP_{s,j_2}x^\top + 2a_{s,j_2}x_{h,s-\log_2 R}^\top}\right] & \text{if } x_{s-\log_2 R+1} = 1,\\ 0 & \text{if } x_{s-\log_2 R+1} = 0.\end{cases}\end{aligned} \tag{10}$$

We then have that $\eta_{s,j_1,j_2,-}(x) = 0$ for all $x \in \mathbb{F}_2^m$ if for every $x \in \mathbb{F}_2^m$ such that $x_{f,\log_2 \mathcal{C}-s} = 100\ldots0$ we have

$$\begin{aligned}xP_{s,j_1}x^\top + 2a_{s,j_1}x_{h,s-\log_2 R}^\top &= 2a_{s,j_2}x_{h,s-\log_2 R}^\top\\ &\quad + xP_{s,j_2}x^\top \pmod 4\end{aligned} \tag{11}$$

(so that the two terms in (10) cancel out). Similarly, we have that $\eta_{s,j_1,j_2,+}(x) = 0$ for all $x \in \mathbb{F}_2^m$ if for all $x \in \mathbb{F}_2^m$ such that $x_{f,\log_2 \mathcal{C}-s} = 100\ldots0$ we have

$$\begin{aligned}xP_{s,j_1}x^\top + 2a_{s,j_1}x_{h,s-\log_2 R}^\top &= xP_{s,j_2}x^\top\\ &\quad + 2a_{s,j_2}x_{h,s-\log_2 R}^\top\\ &\quad + 2 \pmod 4.\end{aligned} \tag{12}$$

These conditions match the third case stated in the theorem. We now focus on $\eta_{s,j_1,j_2,-}(x)$ and study a couple of special cases for (11).

- If $P_{s,j_1} = P_{s,j_2}$, then condition (11) requires $2a_{s,j_1}\bar{x}^\top = 2a_{s,j_2}\bar{x}^\top$ for all $\bar{x} \in \mathbb{F}_2^{s-\log_2 R}$. This is only possible if $a_{s,j_1} = a_{s,j_2}$, which is a contradiction with $j_1 \neq j_2$ and $P_{s,j_1} = P_{s,j_2}$.
- If $P_{s,j_1} \neq P_{s,j_2}$ and $a_{s,j_1} = a_{s,j_2}$, then condition (11) requires

  $$xP_{s,j_1}x^\top = xP_{s,j_2}x^\top.$$

  This implies that for $\eta_{s,j_1,j_2,\pm}$ to be part of the null space of Φ, we must have $(P_{s,j_1} + P_{s,j_2})$ with rank at most $m - s + \log_2(\mathcal{C}/N)$. Since the DGF class of matrices forces this rank to be at least $m - 2r$, we may have $\eta_{s,j_1,j_2,\pm} = 0$ (i.e., sums of wavelets at scale s are in the null space) only if $s \leq \log_2(\mathcal{C}/N) + 2r$.

This gives the second condition in the theorem; the conditions for $\eta_{s,j_1,j_2,+}(x)$ follow similarly from (12). □

l	$g(l)$
1	$[1\ 1]$
i	$[1\ -1]$
-1	$[-1\ -1]$
$-i$	$[-1\ 1]$

TABLE VIII
GRAY MAP TABLE.

APPENDIX B
REAL-VALUED VERSIONS OF THE DGF

In certain cases, we desire a CS matrix with real-valued entries; there are two possible approaches to adapt the DGF to a real-valued CS matrix. We now show that for each one of these options, the argument in the proof of Theorem 1 can be easily adapted, and therefore the same vectors are in the null space of the real-valued versions of the DGF.

First, one can restrict the matrices $P \in DG(m,r)$ to the subset of the DG set of matrices with zero-valued diagonal entries. With such a restriction, the term $xPx^\top = 2\sum_{0\le i<j<2^m} x_i x_j P_{i,j}$ is an even number, rendering the entries of Φ real-valued. In this case, Theorem 1 can be applied without change, since the result is not dependent on the particular choice of matrices P.

Alternatively, one can create a CS matrix having twice as many rows as the DGF by applying the Gray map $g : \{1,-1,i,-i\} \to \{-1,1\}^2$ to the entries of Φ, given in Table VIII. The Gray map has the property that the norm of the difference between any two powers of i is equal to the norm of the difference of their Gray map image vectors. The new Gray-mapped CS matrix, which we denote by Φ^G, has 2^{m+1} rows and $2^{m(r+1)}$ columns indexed by $(h,x) \in \mathbb{F}_2 \times \mathbb{F}_2^m$ and $(P,b) \in DG(m,r) \times \mathbb{F}_2^m$, respectively. By defining the equivalence $P = P' + d_P$, where d_P is the extracted diagonal of P and P' is the remainder of P, the entries of Φ^G can be written as follows:

$$\Phi^G_{P,b}(h,x) = i^{xP'x^\top + 2bx^\top + d_P x^\top + (2h-1)\epsilon(d_P x^\top)},$$

where $\epsilon(l) = l \bmod 2$, i.e., ϵ_1 extracts the least significant bit of the binary representation of l. Thus we have replaced $xPx^\top$ from the definition of the complex DGF with $xP'x^\top + d_P x^\top + (2h-1)\epsilon(d_P x^\top)$ for the real-valued DGF. Since the dependence of $\Phi^G_{P,b}(h,x)$ on b is the same as that of $\Phi_{P,b}(x)$, it is easy to see that the proof of Theorem 1 is extendable to dyadic sums of the Gray-mapped DGF, as follows.

Corollary 2. *Let Φ^G denote the real-valued Gray mapped version of the $DGF(m,r)$ and write $x_{h,l}$ and $x_{f,l}$ for the first and last l entries of the vector $x \in \mathbb{F}_2^m$, $1 \le l \le m$, respectively. For $0 \le s < \log_2 \mathcal{C}$ and $0 \le j_1, j_2 < 2^s$, $j_1 \ne j_2$, we have $\psi_{s,j_1} + \psi_{s,j_2} \in \mathcal{N}(\Phi^G)$ if*

- $s < \log_2 R$ *(for all j_1, j_2);*
- $\log_2 R \le s \le \log_2 R + 2r$, $a_{s,j_1} = a_{s,j_2}$, $x(P'_{s,j_1} - P'_{s,j_2})x^\top + (d_{P'_{s,j_1}} - d_{P'_{s,j_2}})x^\top = 0 (\mathrm{mod}\ 4)$, *and* $(d_{P'_{s,j_1}} - d_{P'_{s,j_2}})x^\top = 0 (\mathrm{mod}\ 2)$ *for all $x \in \mathbb{F}_2^m$ such that $x_{f,\log_2 \mathcal{C}-s} = 100\ldots0$; or*
- $s \ge \log_2 R$ *and* $x(P'_{s,j_1} - P'_{s,j_2})x^\top + (d_{P'_{s,j_1}} - d_{P'_{s,j_2}})x^\top = 2(a_{s,j_2} - a_{s,j_1})x^\top_{h,s-\log_2 R} \ (\mathrm{mod}\ 4)$, *and* $(d_{P'_{s,j_1}} - d_{P'_{s,j_2}})x^\top = 0 (\mathrm{mod}\ 2)$ *for all $x \in \mathbb{F}_2^m$ such that $x_{f,\log_2 \mathcal{C}-s} = 100\ldots0$.*

The conditions also hold for $\eta_{s,j_1,j_2,-} = 0$ by adding a term of 2 in the equalities $(\mathrm{mod}\ 4)$ involving x.

REFERENCES

[1] M. F. Duarte, S. Jafarpour, and R. Calderbank, "New wavelet coefficient raster scannings for compressive imaging," in *Int. Conf. Sampling Theory and Applications (SAMPTA)*, Singapore, May 2011.

[2] D. L. Donoho, "Compressed sensing," *IEEE Trans. Info. Theory*, vol. 52, no. 4, pp. 1289–1306, September 2006.

[3] E. J. Candès, "Compressive sampling," in *Proc. International Congress of Mathematicians*, Madrid, Spain, 2006, vol. 3, pp. 1433–1452.

[4] R. Calderbank and S. Jafarpour, "Reed-Muller sensing matrices and the Lasso," in *Int. Conf. Sequences and their Applications (SETA)*, Paris, France, Sep. 2010, pp. 442–463.

[5] R. Calderbank, S. Howard, and S. Jafarpour, "Construction of a large class of deterministic sensing matrices that satisfy a statistical isometry property," *IEEE J. Selected Topics in Signal Processing*, vol. 4, no. 2, pp. 358–374, Apr. 2010.

[6] V. Kostina, M. F. Duarte, S. Jafarpour, and R. Calderbank, "The value of redundant measurement in compressed sensing," in *IEEE Int. Conf. Acoustics, Speech, and Signal Proc. (ICASSP)*, Prague, Czech Republic, May 2011, pp. 3656–3659.

[7] J. A. Tropp, "On the conditioning of random subdictionaries," *Appl. Comp. Harmonic Analysis*, vol. 25, no. 1, pp. 1–24, July 2008.

[8] E. J. Candès and Y. Plan, "Near-ideal model selection by ℓ_1 minimization," *Annals of Statistics*, vol. 37, no. 5A, pp. 2145–2177, Oct. 2009.

[9] J. A. Tropp, "Norms of random submatrices and sparse approximation," *C. R. Acad. Sci. Paris, Ser. I*, vol. 346, no. 23–24, pp. 1271–1274, 2008.

[10] W. Bajwa, R. Calderbank, and S. Jafarpour, "Why Gabor frames? Two fundamental measures of coherence and their role in model selection," *J. Communication Networks*, vol. 12, no. 4, pp. 289–307, Aug. 2010.

[11] L. Applebaum, W. U. Bajwa, M. F. Duarte, and R. Calderbank, "Multiuser detection in asynchronous on-off random access channels using Lasso," *Physical Comm.*, vol. 5, no. 2, pp. 129–147, June 2012.

[12] R. G. Baraniuk, V. Cevher, M. F. Duarte, and C. Hegde, "Model-based compressive sensing," *IEEE Trans. Info. Theory*, vol. 56, no. 4, pp. 1982–2001, Apr. 2010.

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

[14] M. F. Duarte and Y. C. Eldar, "Structured compressed sensing: From theory to applications," *IEEE Trans. Signal Proc.*, vol. 59, no. 9, pp. 4053–4085, Sep. 2011.

[15] K. Ni, S. Datta, P. Mahanti, S. Roudenko, and D. Cochran, "Efficient deterministic compressed sensing for images with chirps and Reed-Muller sequences," *SIAM J. Imaging Science*, vol. 4, no. 3, pp. 931–953, 2011.

[16] K. Ni, P. Mahanti, and S. Roudenko, "Stability of efficient deterministic compressed sensing for images with chirps and Reed-Muller sequences," *Preprint*, 2010.

[17] L. Gan, T. T. Do, and T. D. Tran, "Fast compressive imaging using scrambled block Hadamard ensemble," in *European Signal Processing Conf. (EUSIPCO)*, Laussanne, Switzerland, Aug. 2008.

[18] M. F. Duarte, M. A. Davenport, D. Takhar, J. N. Laska, T. Sun, K. F. Kelly, and R. G. Baraniuk, "Single pixel imaging via compressive sampling," *IEEE Signal Proc. Mag.*, vol. 25, no. 2, pp. 83–91, March 2008.

[19] R. Robucci, J. Gray, L. K. Chiu, J. K. Romberg, and P. Hasler, "Compressive sensing on a CMOS separable-transform image sensor," *Proc. IEEE*, vol. 98, no. 6, pp. 1089–1101, June 2010.

[20] L. Jacques, P. Vandergheynst, A. Bibet, V. Majidzadeh, A. Schmid, and Y. Leblebici, "CMOS compressed imaging by random convolution," in *IEEE Int. Conf. Acoustics, Speech, and Signal Proc. (ICASSP)*, Taipei, Taiwan, Apr. 2009, pp. 1113–1116.

[21] T. T. Do, L. Gan, N. H. Nguyen, and T. D. Tran, "Fast and efficient compressive sensing using structurally random matrices," *IEEE Trans. Signal Proc.*, vol. 60, no. 1, pp. 139–154, Jan. 2012.

[22] E. Candès, J. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure and Applied Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[23] M. F. Duarte, "Fast reconstruction from random incoherent projections," Tech. Rep. TREE-0507, Rice Univ. ECE Dept., Houston, TX, May 2005.

[24] E. J. Candès and J. Romberg, "ℓ_1-MAGIC: Recovery of sparse signals via convex programming," http://www.l1-magic.org, Oct. 2005.

[25] S. Chen, D. L. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientific Computing*, vol. 20, no. 1, 1998.

[26] T. Blumensath and M. E. Davies, "Iterative hard thresholding for compressed sensing," *Appl. Comput. Harmonic Analysis*, vol. 27, no. 3, pp. 265–274, Nov. 2008.

[27] University of Southern California Signal and Image Processing Institute, "The USC-SIPI image database," http://sipi.usc.edu/database/.

[28] M. Friedlander and M. Saunders, "ASP: A set of Matlab functions for solving basis pursuit-type problems," 2010.

[29] C. Hegde, M. F. Duarte, and V. Cevher, "Compressive sensing recovery of spike trains using a structured sparsity model," in *Workshop on Signal Proc. with Adaptive Sparse Structured Representations (SPARS)*, Saint Malo, France, Apr. 2009.