ISIT'2013 1569719789

Oblique Pursuits for Compressed Sensing with Random Anisotropic Measurements

Kiryung Lee and Yoram Bresler
Coordinated Science Laboratory and Department of ECE
University of Illinois at Urbana-Champaign
1308 W. Main St., Urbana, IL 61801
Email: {klee81,ybresler}@illinois.edu

Marius Junge
Department of Mathematics
University of Illinois at Urbana-Champaign
1409 W. Green St., Urbana, IL 61801
Email: junge@math.uiuc.edu

***Abstract*—Compressed sensing enables universal, simple, and reduced-cost acquisition by exploiting a sparse signal model. Most notably, recovery of the signal by computationally efficient algorithms is guaranteed for certain random measurement models, which satisfy the so-called isotropy property. However, in real-world applications, this property is often not satisfied. We propose two related changes in the existing framework for the anisotropic case: (i) a generalized RIP called the restricted biorthogonality property (RBOP); and (ii) correspondingly modified versions of existing greedy pursuit algorithms, which we call oblique pursuits. Oblique pursuits provide recovery guarantees via the RBOP without requiring the isotropy property; hence, these recovery guarantees apply to practical acquisition schemes. Numerical results show that oblique pursuits also perform better than their conventional counterparts.**

I. Introduction

Compressed sensing is a new paradigm of data acquisition that exploits a sparse signal model to reduce the amount of data that needs to be acquired to recover the signal of interest. Unlike the conventional paradigm, in which large quantities of data are acquired, often followed by compression, compressed sensing acquires minimally redundant data directly in a universal data-independent way [1]–[3].

We focus on acquisition schemes that take random samples in a certain transform domain. Let $f \in \mathbb{K}^n$ (where $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$) be the unknown signal. The measurement vector $y \in \mathbb{K}^m$ has components

$$y_k = \frac{1}{\sqrt{m}} \phi_{\omega_k}^* f + v_k, \quad \forall k = 1, \ldots, m,$$

where $\Phi = [\phi_1, \ldots, \phi_L] \in \mathbb{K}^{n \times L}$ is a frame representing the transform, $(\omega_k)_{k=1}^m$ are random sampling indices, x^* denotes the Hermitian transpose of vector x, and $v \in \mathbb{K}^m$ denotes additive noise. Compressed sensing aims to reconstruct signals that are (approximately) s-sparse over a dictionary $D \in \mathbb{K}^{n \times n}$ (cf. [2], [4]) from $m (< n)$ compressive measurements. Let $x \in \mathbb{K}^n$ be the coefficient vector of f over D such that $f \approx Dx$ with x being s-sparse. Let $R \in \mathbb{K}^{m \times L}$ be a random sampling operator where the kth row of R is the ω_kth standard basis vector in $\mathbb{R}^L$. Then,

$$\Psi = \frac{1}{\sqrt{m}} R \Phi^* D \tag{1.1}$$

may be regarded as a sensing matrix for s-sparse x that produces m random measurements. Once an estimate $\widehat{x}$ of x is computed, $D\widehat{x}$ provides an estimate of the unknown signal f; hence, we may focus on the recovery of s-sparse x.

We consider a specific model for the random sampling indices $(\omega_k)_{k=1}^m$. Let $(\omega_k)_{k=1}^m$ be i.i.d. copies of a random variable ω distributed on $[L] \triangleq \{1, \ldots, L\}$ according to a distribution p. Denoting the expected value with respect to p by $\mathbb{E}$, the random vector ψ_ω is said to satisfy the *isotropy property* if $\mathbb{E}\psi_\omega \psi_\omega^* = I_n$. Then, $(y_k)_{k=1}^m$ are called random isotropic measurements.

The assumption of isotropy underlies the performance guarantees of compressed sensing algorithms. In the isotropic case, under an additional assumption on the mutual incoherence between the columns of Φ and the columns of D, Ψ satisfies the *restricted isometry property* (RIP) with high probability for $m = O(s \log^4 n)$ [6], [7]. In turn, in the context of the random measurement model in (1.1), the RIP of Ψ enables guaranteed recovery of x from a small number of measurements by computationally efficient algorithms (e.g., [8]–[12]). A "RIPless" guarantee for ℓ_1-norm based solutions [13] has been shown for the same acquisition model. However, while not using the RIP explicitly, the RIPless analysis too depends on the isotropy property of Ψ.

Although the isotropy property of Ψ is a key assumption in deriving the RIP-based (as well as the "RIPless") performance guarantees, it is often not satisfied in real-world applications. To enforce the isotropy property, previous studies on the RIP [6], [7] assumed that p is the uniform distribution on $[L]$, Φ is a tight frame ($\Phi\Phi^* = I_n$), and D is an ortho basis ($D^*D = I_n$). However, this condition is often violated in practice. For example, in the context of imaging, the transform Φ^* is determined by the specific modality. In MRI, Φ^* corresponds to the Fourier transform, which is unitary. But, in CT, Φ^* corresponds to the Radon transform, which is not unitary ($\Phi\Phi^* \neq I_n$). The design of p too depends on the application. In MRI, it is desirable to take more measurements of lower frequency components, which contain more of the signal energy; hence, a variable distribution p is preferred to the uniform distribution [14]. Finally, the dictionary D depends on the signals of interest. Relaxing the orthogonality constraint, one can design a better D for given set of signals. For example, data adaptive transforms without orthogonality empirically provide sparser representations than the block

This work was supported in part by the National Science Foundation under Grants CCF 10-18660, DMS 09-01457, and DMS 12-01886.

This paper presents a selection of the results reported in [5] and all proofs are referred to [5].

DCT [15]. In these scenarios, the isotropy property is not satisfied.

Other work on the anisotropic case: Recently, Rudelson and Zhou [16] and Kueng and Gross [17] studied the guarantee for the ℓ_1-norm based solutions in the anisotropic case. However, they did not propose any change in the solutions to accommodate the anisotropy of random measurements. Furthermore, the theory in these works is not applicable to greedy pursuit algorithms in the anisotropic case, and does not provide guarantees for them.

Contribution: Unlike the existing approaches to the anisotropic case, we propose a way to modify existing greedy pursuit algorithms for the anisotropic context. First, we propose a modified version of RIP called the *restricted biorthogonality property (RBOP)*, which is a property of a pair $(\Psi, \widetilde{\Psi})$ so that $\widetilde{\Psi}^*\Psi x \approx x$ for sparse x. Second, we propose an explicit construction of $\widetilde{\Psi}$ that satisfies the RBOP of $(\Psi, \widetilde{\Psi})$ under a condition similar to that of the RIP in the anisotropic case. Third, we modify the greedy pursuit algorithms to use $\widetilde{\Psi}$ instead of Ψ in key steps. We call the modified greedy algorithms *oblique pursuits*. With random anisotropic measurements, oblique pursuits outperform the original pursuit algorithms and provide performance guarantees via the RBOP.

Organization: The modification of existing algorithms into oblique pursuits is explained in Section II. The conditions that provide the RBOP for random anisotropic measurements are discussed in Section III, followed by performance guarantees of oblique pursuits via the RBOP in Section IV, and numerical studies on empirical performance in Section V. We conclude this paper with a summary in Section VI.

Notation: We will use various notations on a matrix $A \in \mathbb{K}^{m\times n}$. The range space spanned by the columns of A will be denoted by $\mathcal{R}(A)$. The adjoint operator of A will be denoted by A^*. The jth column of A is denoted by a_j and the submatrix of A with columns indexed by $J \subset [n]$ is denoted by A_J. For simplicity, $(A_J)^*$ will be abbreviated as A_J^*.

II. Oblique Pursuits

A. Modifying greedy pursuit algorithms to oblique pursuits

We propose to modify existing greedy pursuit algorithms so that the modified algorithms that we call *oblique pursuits* perform well and provide recovery guarantees in the anisotropic case. Note that the isotropy property is equivalently rewritten as $\mathbb{E}\Psi^*\Psi = I_n$. Instead of this, we assume a much weaker requirement that $\mathbb{E}\Psi^*\Psi$ is only invertible. Then, simply replacing Ψ^* in the greedy pursuit algorithms by $\widetilde{\Psi}^*$ defined by

$$\widetilde{\Psi} \triangleq \Psi(\mathbb{E}\Psi^*\Psi)^{-1} \tag{2.1}$$

provides the desired variants. Our choice of $\widetilde{\Psi}^*$ guarantees that $\mathbb{E}\widetilde{\Psi}^*\Psi = I$, which provides the RBOP of $(\Psi, \widetilde{\Psi})$ for Ψ corresponding to random anisotropic measurements (Section III).[1] We present the motivation for the modified algorithms by analyzing the implication of the replacement of Ψ by $\widetilde{\Psi}$ on key operations in the greedy pursuit algorithms.

[1]Note that the construction of $\widetilde{\Psi}$ in (2.1) requires knowledge of both Ψ and the distribution p of the random sampling indices.

Proxy: One common operation in the greedy pursuit algorithms is to apply Ψ^* to Ψx for sparse x. The resulting vector $\Psi^*\Psi x$ is called "proxy" of x and reveals the information of locations of nonzero elements of x when $\Psi^*\Psi x \approx x$, which holds by the RIP of Ψ.

In the anisotropic case, our choice of $\widetilde{\Psi}^*$ guarantees that $\mathbb{E}\widetilde{\Psi}^*\Psi = I_n$, which provides the desired proxy property $\widetilde{\Psi}^*\Psi x \approx x$ via the RBOP of $(\Psi, \widetilde{\Psi})$; hence, $\widetilde{\Psi}^*\Psi x$ is a good proxy of x.

Orthogonal Matching: OMP [18] and SP [10] include the operation of finding the point $\hat{y}$ nearest to $y = \Psi x$ for sparse x, within the subspace spanned by the columns of Ψ indexed by a given support set J. The outcome of this operation is the orthogonal projection of y onto the subspace $\mathcal{R}(\Psi_J)$. Let $\Pi_J^\perp x$ denote x with elements on J replaced by 0. Then, a proxy of $\Pi_J^\perp x$ is computed as $\Psi^* P_{\mathcal{R}(\Psi_J)}^\perp \Psi x$. In the isotropic case, by the RIP of Ψ and the spectral lemma of the Schur complement, this is a good proxy because $\Psi^* P_{\mathcal{R}(\Psi_J)}^\perp \Psi x \approx \Pi_J^\perp x$ (cf. [19]). The procedure for obtaining the proxy from a given support set J is called "orthogonal matching", which was first introduced in OMP as a way to incrementally refine the support set by one element per iteration.

In the anisotropic case, in order to exploit the RBOP of $\widetilde{\Psi}^*\Psi$, we proceed as follows: (i) replace the orthogonal projection onto $\mathcal{R}(\Psi_J)$ by an oblique projection onto the range space $\mathcal{R}(\Psi_J)$ along the null space $\mathcal{R}(\widetilde{\Psi}_J)^\perp$, which is denoted by $E_{\mathcal{R}(\Psi_J),\mathcal{R}(\widetilde{\Psi}_J)^\perp}$; and (ii) replace Ψ^* by $\widetilde{\Psi}^*$. In fact, the modification (i) also reduces to (ii). Indeed, note that $P_{\mathcal{R}(\Psi_J)} y = \Psi_J \alpha$ for α satisfying

$$\alpha = \arg\min_\alpha \|y - \Psi_J\alpha\|_2^2 = \arg\min_\alpha \|\Psi_J^*(y - \Psi_J\alpha)\|_2^2,$$

and $E_{\mathcal{R}(\Psi_J),\mathcal{R}(\widetilde{\Psi}_J)^\perp} y = \Psi_J\alpha$ for α satisfying

$$\alpha = \arg\min_\alpha \|\widetilde{\Psi}_J^*(y - \Psi_J\alpha)\|_2^2.$$

Therefore, the only difference in computing $P_{\mathcal{R}(\Psi_J)}$ and $E_{\mathcal{R}(\Psi_J),\mathcal{R}(\widetilde{\Psi}_J)^\perp}$ is the replacement of Ψ_J^* by $\widetilde{\Psi}_J^*$. The modified operation is called *oblique matching*.

Signal Estimation: SP [10], CoSaMP [9], and HTP [12] compute an approximation $\hat{x}$ of the sparse x from $y = \Psi x$ such that $\hat{x}$ is supported on a given support set J, and $\Psi\hat{x}$ is closest to y. In fact, the restriction of $\hat{x}$ onto J, denoted by $\hat{x}|_J$, satisfies

$$\hat{x}|_J = \arg\min_\alpha \|\Psi x - \Psi_J\alpha\|_2^2 = \arg\min_\alpha \|\Psi_J(\Psi_J^\dagger\Psi x - \alpha)\|_2^2.$$

By the RIP of Ψ, it follows that $\Psi_J^\dagger\Psi x \approx x_J$ and $\Psi_J^*\Psi_J \approx I_s$, which implies $\|\Psi_J(\Psi_J^\dagger\Psi x - \alpha)\|_2 \approx \|x|_J - \alpha\|_2$. Therefore, $\hat{x}_J \approx x_J$, i.e., $\hat{x}$ is a good approximation of $\Pi_J x$. If J includes the support of x, then $\hat{x}$ is a good approximation of x. Note that $\hat{x}|_J$ is equivalently rewritten as

$$\hat{x}|_J = \arg\min_\alpha \|\Psi_J^*(y - \Psi_J\alpha)\|_2^2.$$

In the anisotropic case, we propose to replace Ψ_J^* by $\widetilde{\Psi}_J^*$. Then, by the RBOP of $(\Psi, \widetilde{\Psi})$, the modified operation computes a good approximation of $\Pi_J x$.

Combining the changes: Existing greedy pursuit algorithms consist of the aforementioned operations. Combining the suggested changes, we obtain appropriately modified algorithms for the anisotropic case: we assign them new names, with the modifier "oblique". For example, SP is extended to *oblique subspace pursuit* (ObSP). CoSaMP, IHT, and HTP are likewise extended to ObCoSaMP, ObIHT, ObHTP, respectively. As an example of the oblique pursuits, we present ObSP in Alg 1, where H_s denotes the hard thresholding operator that keeps only the s-largest entries and sets the other entries to zero.

Algorithm 1: Oblique Subspace Pursuit (ObSP)

while *stop condition not satisfied* **do**
  $\widetilde{J}_{t+1} \leftarrow \operatorname{supp}(x_t) \cup \operatorname{supp}\Big(H_s\big(\widetilde{\Psi}^*(y - \Psi x_t)\big)\Big)$;
  $\widetilde{x} \leftarrow \arg \min_{x, \operatorname{supp}(x) \subset \widetilde{J}_{t+1}} \big\|\widetilde{\Psi}^*_{\widetilde{J}_{t+1}}(y - \Psi x)\big\|_2$;
  $J_{t+1} \leftarrow \operatorname{supp}(H_s(\widetilde{x}))$;
  $x_{t+1} \leftarrow \arg \min_{x, \operatorname{supp}(x) \subset J_{t+1}} \big\|\widetilde{\Psi}^*_{J_{t+1}}(y - \Psi x)\big\|_2$;
  $t \leftarrow t + 1$;
end

Remark 2.1: Schnass and Vandergheynst [20] previously proposed variants of the thresholding algorithm and matching pursuit algorithms that replace Ψ^* by another matrix $\widehat{\Psi}^*$. However, their work differs from ours in several key aspects. First, they only modified the proxy operations; hence, their variant of OMP is structurally different from ObMP, which also replaces the orthogonal matching by the oblique matching. Second, their constructions of $\widehat{\Psi}^*$ were (i) $\widehat{\Psi}^* = \Psi^\dagger$; or (ii) a $\widehat{\Psi}^*$ numerically designed to optimize a version of the Babel function. Both are different from our construction of $\widetilde{\Psi}^*$ in (2.1), and neither one provides a performance guarantee.

B. Explicit form of $\widetilde{\Psi}$ for random anisotropic measurements

If Ψ corresponds to the random anisotropic measurement in (1.1), then $\widetilde{\Psi}$ in (2.1) is equivalently rewritten as

$$\widetilde{\Psi} = \frac{1}{\sqrt{m}} \Lambda^{-1} R (\Phi^*)^\dagger D^{-1} \tag{2.2}$$

where $\Lambda \in \mathbb{K}^{m \times m}$ is a diagonal matrix given by $(\Lambda)_{k,k} = p_{\omega_k}$ for $k = 1, \ldots, m$. Indeed, it is readily verified that for $\widetilde{\Psi}$ defined in (2.2), $\mathbb{E}\widetilde{\Psi}^*\Psi = I_n$.

What is the effect of modifying greedy algorithms to oblique pursuits on their computational cost? In applications, Ψ is often well structured so that applying Ψ and Ψ^* is much cheaper than the usual dense matrix-vector multiplication. For example, in MRI, $R\Phi^*$ and ΦR^* can be efficiently implemented as a nonuniform FFT (NUFFT) [21]. In CT, Φ corresponds to backprojection, for which there also are fast algorithms [22], [23]. This contributes to the low cost of the various greedy pursuit algorithms. However, unlike Ψ^*, which only involves the adjoint operators that preserve the efficient structure, $\widetilde{\Psi}^*$ in (2.2) involves the pseudo inverse. In general, (pseudo)inverses of operators do not necessarily preserve the structure. Nonetheless, in certain scenarios that encompass imaging applications, the computational cost of applying $\widetilde{\Psi}^*$ is as low as that of applying Ψ^*. The cost of applying the diagonal operator Λ^{-1} is negligible. In various modalities, $(\Phi^*)^\dagger$ also admits efficient computation. In MRI, $(\Phi^*)^\dagger = \Phi$; hence, NUFFT is also applicable. In CT, $(\Phi^*)^\dagger$ corresponds to the filtered back projection. Regarding the dictionary, some popular D also have computationally efficient implementation for D^{-1}. For example, wavelet transforms and their inverse have fast implementation. If D applies to nonoverlapping patches (e.g., block DCT [14]), then D^{-1} also applies to nonoverlapping patches and preserves the same block-diagonal structure. Therefore, in modifying the greedy pursuit algorithms to oblique pursuits, we preserve the advantage of low computational cost of the original algorithms.

Remark 2.2: The constructions of $\widehat{\Psi}$ by Schnass and Vandergheynst [20] do not preserve the aforementioned computational advantages of the original algorithm. When Φ^* corresponds to the Fourier transform, their proposed $\widehat{\Psi}^* = \Psi^\dagger$ is a non-structured dense matrix. In particular, if $\Psi^\dagger$ is large (e.g., in imaging applications), neither its computation nor its storage are practical, and the application of $\Psi^\dagger$ using iterative solvers leads to higher cost than the oblique pursuits. Their other construction of $\widehat{\Psi}^*$ that minimizes the Babel function does not preserve structures and is not practical either.

III. RESTRICTED BIORTHOGONALITY PROPERTY

A. General estimate of required number of measurements

We first extend the RIP to a property of a pair of matrices $\Psi, \widetilde{\Psi} \in \mathbb{K}^{m \times n}$ called the *restricted biorthogonality property* (RBOP).

Definition 3.1 (s-RBOC): The *s-restricted biorthogonality constant* $\theta_s(M)$ of $M \in \mathbb{K}^{n \times n}$ is defined as the smallest δ that satisfies

$$|\langle y, Mx \rangle - \langle y, x \rangle| \leq \delta \|x\|_2 \|y\|_2 \tag{3.1}$$

for all s-sparse x, y with common support.
The pair $(\Psi, \widetilde{\Psi})$ satisfies the RBOP of order s with constant c if $\theta_s(\widetilde{\Psi}^*\Psi) < c$.

Remark 3.2: The *restricted isometry constant* (RIC) of Ψ, denoted by $\delta_s(\Psi)$, is defined as the smallest constant δ that satisfies $\big|\|\Psi x\|_2^2 - \|x\|_2^2\big| \leq \delta \|x\|_2^2$ for all s-sparse x. It follows from their definitions that $\theta_s(\Psi^*\Psi) = \delta_s(\Psi)$.

Next, we extend the corresponding RIP analysis of random isotropic measurements [7, Theorem 8.4] to the anisotropic case in the following theorem.

Theorem 3.3: Let $\Psi, \widetilde{\Psi} \in \mathbb{K}^{m \times n}$ be random matrices not necessarily mutually independent, each with i.i.d. rows with elements bounded in magnitude as

$$\max_{k,\ell} |(\Psi)_{k,\ell}| \leq \frac{K}{\sqrt{m}} \quad \text{and} \quad \max_{k,\ell} |(\widetilde{\Psi})_{k,\ell}| \leq \frac{\widetilde{K}}{\sqrt{m}}$$

for $K, \widetilde{K} \geq 1$. Then, $\theta_s(\widetilde{\Psi}^*\Psi) < \delta + \theta_s(\mathbb{E}\widetilde{\Psi}^*\Psi)$ holds with probability $1 - \eta$ provided that

$$m \geq C_1 \delta^{-2} \left(K\sqrt{2 + \theta_s(\mathbb{E}\Psi^*\Psi)} + \widetilde{K}\sqrt{2 + \theta_s(\mathbb{E}\widetilde{\Psi}^*\widetilde{\Psi})} \right)^2 \cdot s(\ln s)^2 \ln n \ln m,$$

$m \geq C_2 \delta^{-2} \widetilde{K} \max(K, \widetilde{K}) s \ln(\eta^{-1})$

for universal constants C_1 and C_2.

Remark 3.4: Rudelson and Zhou [16] derived a symmetric variant of RIP, which holds for $m = O(\kappa K s \log^4 n)$, where κ denotes the condition number of $\mathbb{E}\Psi^*\Psi$, and provides performance guarantees of ℓ_1-based solutions in the anisotropic case. However, this property provides performance guarantee for neither greedy pursuit algorithms nor for oblique pursuits. Kueng and Gross [17] independently showed a weaker version of the RBOP. Unlike Theorem 3.3, which holds *uniformly* for all support of cardinality s, their property holds *locally* for a fixed support with $m = O(\kappa K s \log n)$ (assuming $\widetilde{K} = K$). We can modify Theorem 3.3 to a similar weaker version. In this case, the condition on m is rewritten as $m = O((3\kappa{+}1)/(\kappa{+}1) K s \log^4 n)$. (The different dependence on κ in our and the Kueng/Gross result is due to different assumptions on the eigenvalues of $\mathbb{E}\Psi^*\Psi$.) However, the local version of RBOP of Kueng and Gross is too weak to provide performance guarantees for iterative oblique pursuits.

B. Required number of random anisotropic measurements

With a more detailed specification of the deviation from isotropy, the general estimates of Theorem 3.3 reduce to a more interpretable form.

The isotropy property, $\mathbb{E}\Psi^*\Psi = I_n$, holds under the following ideal assumptions: (i) tight frame: $\Phi\Phi^* = I_n$; (ii) uniform distribution: $p_k = \frac{1}{L}$, $\forall k \in [L]$; (iii) orthonormal basis: $D^*D = I_n$.

In the anisotropic case where the ideal assumptions are violated, RBOP holds for $m = O(s \log^4 n)$, where the universal constant inside the big O notation depends on the degree of the violations. To quantify this statement, we introduce the following metrics that measure the deviation from the ideal assumptions.

- Nonuniform distribution $p \in [0,1]^L$:
$$\nu_{\min} \triangleq L \min_{k \in [L]} p_k \quad \text{and} \quad \nu_{\max} \triangleq L \max_{k \in [L]} p_k. \tag{3.2}$$
- Non-tight frame Φ: WLOG, we assume $\lambda_1(\Phi\Phi^*) - 1 = 1 - \lambda_n(\Phi\Phi^*)$, which implies
$$\delta_n(\Phi^*) = \|\Phi\Phi^* - I_n\| = \frac{\kappa(\Phi\Phi^*) - 1}{\kappa(\Phi\Phi^*) + 1}$$
where $\kappa(\Phi\Phi^*)$ denotes the condition number of $\Phi\Phi^*$.
- Non-ortho basis D: In fact, we only require a small RIC $\delta_s(D)$. An easy upper bound on $\delta_s(D)$ is given by
$$\delta_s(D) \leq \delta_n(D) = \|D^*D - I_n\| = \frac{\kappa(D^*D) - 1}{\kappa(D^*D) + 1}$$
where the last equality holds by the assumption that $\lambda_1(D^*D) - 1 = 1 - \lambda_n(DD^*)$.

Theorem 3.5: Let $\Phi \in \mathbb{K}^{n \times L}$ and $D \in \mathbb{K}^{d \times n}$ be mutually incoherent with constant K, i.e., $\max_{j,k} |\langle \phi_k, d_j \rangle| \leq K$. Let $\Psi, \widetilde{\Psi} \in \mathbb{K}^{m \times n}$ be constructed by (1.1) and (2.1), respectively, using p. Let $\nu_{\min}$ and $\nu_{\max}$ be defined in (3.2). Then, all three inequalities $\theta_s(\widetilde{\Psi}^*\Psi) < \delta$, $\delta_s(\Psi) < \delta + K_1$, and $\delta_s(\widetilde{\Psi}) < \delta + K_1$ hold with probability $1 - \eta$ provided that
$$m \geq C_1 (1 + K_1)^2 K_2^2 \delta^{-2} s (\ln s)^2 \ln n \ln m,$$
$$m \geq C_2 K_2^2 \delta^{-2} s \ln(\eta^{-1})$$
for universal constants C_1 and C_2, where K_1 and K_2 are given in terms of K, $\nu_{\min}$, $\nu_{\max}$, $\delta_s(D)$, and $\delta_n(\Phi^*)$ by
$$K_1 = \max(1 - \nu_{\max}^{-1}, \nu_{\min}^{-1} - 1) + \max(\nu_{\max}, \nu_{\min}^{-1}) \left\{ 1 + \frac{\delta_s(D)}{1 - \delta_s(D)} + \frac{\delta_n(\Phi^*)}{1 - \delta_n(\Phi^*)} + \frac{\delta_s(D)\delta_n(\Phi^*)}{[1 - \delta_s(D)][1 - \delta_n(\Phi^*)]} \right\}$$
$$K_2 = \frac{\|(D^*D)^{-1}\|_{\ell_1^n \to \ell_1^n}}{\nu_{\min}^2} \cdot \left[K + \left(\sup_{\omega \in \Omega} \|\phi_\omega\|_{\ell_2^d} \right) \cdot \frac{\delta_n(\Phi^*)}{1 - \delta_n(\Phi^*)} \cdot \left(\max_{j \in [n]} \|d_j\|_{\ell_2^d} \right) \right].$$

In the isotropic case, i.e., when the ideal assumptions are satisfied, $\nu_{\min} = \nu_{\max} = 1$ and $\delta_s(D) = \delta_n(\Phi^*) = 0$; hence K_2 decreases accordingly and $K_1 = 0$ in this case. Then, Theorem 3.5 reduces to the analogous result in the isotropic case [7, Theorem 8.4], which shows that $\delta_s(\Psi) < \delta$ with $m = O(sK^2\delta^{-2}\ln^4)$. Otherwise, in the anisotropic case, constants K_1 and K_2 increase as a penalty for the violation of the ideal assumptions. In turn, the upper bound on $\delta_s(\Psi)$ becomes more conservative as K_1 increases. As this bound exceeds a certain constant, it fails to provide RIP-based performance guarantees. For example, the recovery guarantee for HTP requires $\delta_{3s}(\Psi) < 0.57$. Nonetheless, the upper bound on $\theta_s(\widetilde{\Psi}^*\Psi)$ remains the same with increasing K_1; hence, the performance guarantees for oblique pursuits (presented in Section IV) are valid in the anisotropic case.

IV. Performance Guarantees

We show that oblique pursuit algorithms admit performance guarantees via the RBOP in the anisotropic case. These performance guarantees are analogous to the corresponding RIP-based performance guarantees of the original pursuit algorithms in the isotropic case. In fact, the modification of the performance guarantees is rather straightforward. Similar to the modification of the original algorithms into oblique pursuits, replacing Ψ^* by $\widetilde{\Psi}^*$ in the RIP result and in its derivation will provide RBOP-based performance guarantees. For example, we present the performance guarantee for ObSP in the following theorem.

Theorem 4.1 (Oblique Subspace Pursuit): Let $(x_t)_{t \in \mathbb{N}}$ be the sequence generated by ObSP. Then
$$\|x_{t+1} - x^\star\|_2 \leq \rho \|x_t - x^\star\|_2 + \tau \|z\|_2 \tag{4.1}$$
where ρ and τ are positive constants given as explicit functions of $\theta_{3s}(\widetilde{\Psi}^*\Psi)$, $\delta_{3s}(\Psi)$, and $\delta_{3s}(\widetilde{\Psi})$. Moreover, ρ, which only depends on $\theta_{3s}(\widetilde{\Psi}^*\Psi)$, satisfies $\rho < 1$, provided that $\theta_{3s}(\widetilde{\Psi}^*\Psi) < 0.325$.

We have derived similar performance guarantees for other greedy algorithms [5]. For example, the recovery by Ob-

CoSaMP, ObIHT, and ObHTP is guaranteed by $\theta_{4s}(\widetilde{\Psi}^*\Psi) < 0.384$, $\theta_{3s}(\widetilde{\Psi}^*\Psi) < 0.5$, and $\theta_{3s}(\widetilde{\Psi}^*\Psi) < 0.577$, respectively. Note that these are only examples. Other greedy algorithms and their RIP-based performance guarantees similarly modify into the corresponding oblique pursuits and RBOP-based performance guarantees.

V. Numerical Results

We compared the empirical performance of oblique pursuits to that of their conventional counterparts and to other methods in the anisotropic case. In this experiment, Φ^* is the DFT matrix, p is the variable density suggested by Lustig et al. [14], and D is a data-adaptive dictionary that applies to non-overlapping patches [15]. This data-adaptive dictionary D provides more sparse representation compared to the block DCT considered by Lustig et al. [14]. Both dictionaries have block diagonal structure but D used in this experiment is nonunitary, with condition number $\kappa(D) = 1.99$. Matrix D^{-1} used in the construction of $\widetilde{\Psi}$ also has the same block diagonal structure.

The input image was a phantom image obtained by s-sparse approximation over the dictionary D of an original brain image with sparsity ratio $s/n = 0.125$. Our goal in this experiment is not to compete with the state of the art of recovery algorithms in CS imaging system; rather, we want to check whether the oblique pursuit algorithms perform competitively with their conventional counterparts in the anisotropic case. This motivates our choice of a simplified test scenario. We also compare the oblique pursuit algorithms to simple zero filling, and to the ℓ_1 analysis formulation [24].

TABLE I

Quality (PSNR in decibels) of images reconstructed from noisy measurements at SNR = 30 decibels. Results averaged over 100 random sampling patterns.

	conventional	oblique
Thres	9.34	31.13
CoSaMP	29.46	32.21
SP	34.74	36.17
IHT	9.34	31.10
HTP	31.58	36.26
ℓ_1-Analysis	30.96	
Zero Filling	31.55	

Table I shows the PSNR of the reconstructed images using the various algorithms with downsampling by factor 3. The oblique pursuit algorithms performed better than their conventional counterparts. In particular, ObSP and ObHTP performed significantly better than zero filling. We observed that thresholding and IHT totally failed in this experiment. In contrast, ObIHT provided a reasonable performance with a fixed step size. Although improvement of the empirical reconstruction performance was not our goal, it turned out fortuitously that the oblique pursuit algorithms, designed to provide recovery guarantees in terms of the RBOP, also show significant improvement in empirical performance.

VI. Conclusion

We propose to modify existing greedy pursuit algorithms to address the anisotropy of random measurements. The modification is based on a generalization of the RIP called the restricted biorthogonality property (RBOP). Unlike RIP, RBOP introduces more degrees of freedom in satisfying the property via design of a $\widetilde{\Psi}$ matched to the given matrix Ψ. In the context of random anisotropic measurements, we propose an explicit construction of $\widetilde{\Psi}$ that preserves the advantage of low computational cost in applying Ψ and Ψ^*. To take advantage of the new RBOP, we extended greedy pursuit algorithms with RIP-based recovery guarantees to new variations – oblique pursuit algorithms, so that they provide RBOP-based recovery guarantees. These recovery guarantees apply with relaxed conditions on the sensing matrices and dictionaries, which are satisfied by practical CS imaging schemes. The extension of greedy pursuit algorithms and their RIP-based recovery guarantees to those based on the RBOP is not restricted to the specific algorithms studied in this paper. Finally, we note that although the oblique pursuit algorithms were designed to provide performance guarantees in the worst-case sense, they also perform better than their conventional counterparts empirically.

References

[1] Y. Bresler, M. Gastpar, and R. Venkataramani, in *Proc. DECI*, Santa Fe, NM, Feb. 1999, pp. 48–48.

[2] D. Donoho, *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[3] E. Candes, J. Romberg, and T. Tao, *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, Feb. 2006.

[4] E. Candes and J. Romberg, *Inverse Problems*, vol. 23, p. 969, 2007.

[5] K. Lee, Y. Bresler, and M. Junge, *arXiv preprint arXiv:1207.2681*, 2012.

[6] M. Rudelson and R. Vershynin, *Comm. Pure Appl. Math.*, vol. 61, no. 8, pp. 1025–1045, 2008.

[7] H. Rauhut, in *Theoretical Foundations and Numerical Methods for Sparse Recovery*, 2010, vol. 9, pp. 1–92.

[8] E. Candes, *Comptes rendus-Mathématique*, vol. 346, no. 9, pp. 589–592, 2008.

[9] D. Needell and J. Tropp, *Appl. Comput. Harmon. Anal.*, vol. 26, no. 3, pp. 301–321, 2009.

[10] W. Dai and O. Milenkovic, *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2230–2249, May 2009.

[11] T. Blumensath and M. Davies, *Appl. Comput. Harmon. Anal.*, vol. 27, no. 3, pp. 265–274, 2009.

[12] S. Foucart, in *Proc. SampTA 2011*, Singapore, 2011.

[13] E. Candes and Y. Plan, *IEEE Trans. Inf. Theory*, vol. 57, no. 11, pp. 7235–7254, Nov. 2011.

[14] M. Lustig, D. Donoho, and J. Pauly, *Magnet. Reson. Med.*, vol. 58, no. 6, pp. 1182–1195, 2007.

[15] S. Ravishankar and Y. Bresler, "Learning sparsifying transforms," *IEEE Trans. Signal Process., in-press*.

[16] M. Rudelson and S. Zhou, *arXiv preprint arXiv:1106.1151*, 2011.

[17] R. Kueng and D. Gross, *arXiv preprint arXiv:1205.1423*, 2012.

[18] Y. Pati, R. Rezaiifar, and P. Krishnaprasad, *Proc. Asilomar Conf. on Signals, Syst., and Comput.*, Pacific Grove, CA, Nov. 1993, pp. 40–44.

[19] K. Lee, Y. Bresler, and M. Junge, *IEEE Trans. Inf. Theory*, vol. 58, no. 6, pp. 3613–3641, Jun. 2012.

[20] K. Schnass and P. Vandergheynst, *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 1994–2002, 2008.

[21] J. Fessler and B. Sutton, *IEEE Trans. Signal Process.*, vol. 51, no. 2, pp. 560–574, 2003.

[22] S. Basu and Y. Bresler, *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1760–1773, 2000.

[23] A. George and Y. Bresler, *SIAM J. Appl. Math.*, vol. 68, no. 2, pp. 574–597, 2007.

[24] E. Candes, Y. Eldar, D. Needell, and P. Randall, *Appl. Comput. Harmon. Anal.*, vol. 31, no. 1, pp. 59–73, 2011.