# SUM OF OUTER PRODUCTS DICTIONARY LEARNING FOR INVERSE PROBLEMS

*Saiprasad Ravishankar, Raj Rao Nadakuditi, and Jeffrey A. Fessler*

Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA

## ABSTRACT

The data-driven adaptation of synthesis dictionaries has been exploited in many applications in signal processing and imaging. This paper exploits efficient methods for aggregate sparsity penalized dictionary learning by first approximating the matrix of image patches with a sum of sparse rank-one matrices (outer products) and then using a block coordinate descent approach to estimate the unknowns, and applies such sum of outer products methodologies for the problem of dictionary-blind image reconstruction. We propose efficient algorithms for adaptive image reconstruction and provide a convergence analysis for our methods. Our numerical experiments show the promising performance provided by the proposed schemes over recent methods in compressed sensing-based image reconstruction.

***Index Terms***— Sparse representations, Inverse problems, Fast algorithms, Machine learning, Convergence analysis.

## 1. INTRODUCTION

The sparsity of natural signals and images in a transform domain or dictionary has been exploited in many applications. More recently, data-driven adaptation of synthesis dictionaries, called dictionary learning, has been investigated [1–5], and demonstrated to be useful in inverse problem settings [6–9]. Given a collection of signals $\{\mathbf{y}_i\}_{i=1}^N$ that are represented as columns of $\mathbf{Y} \in \mathbb{C}^{n \times N}$, the dictionary learning problem is often formulated as follows [3]:

$$\text{(P0)} \ \min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F^2 \ \text{ s.t. } \ \|\mathbf{x}_i\|_0 \leq s \ \forall\, i, \ \|\mathbf{d}_j\|_2 = 1 \, \forall\, j.$$

Here, $\mathbf{d}_j$ and $\mathbf{x}_i$ denote the columns of the dictionary $\mathbf{D} \in \mathbb{C}^{n \times J}$ and sparse code matrix $\mathbf{X} \in \mathbb{C}^{J \times N}$, respectively, the $\ell_0$ "norm" counts the number of non-zero vector entries, and $s$ is the maximum sparsity level for each signal. Constraining the columns (or atoms) of $\mathbf{D}$ to have unit norm eliminates the scaling ambiguity [10]. Variants of (P0) include replacing the $\ell_0$ "norm" for sparsity with an alternative sparsity criterion, or enforcing additional properties (e.g., incoherence [11]) for the dictionary, or solving an online version of the problem [5]. Various algorithms have been proposed for dictionary learning [2–5, 12, 13]. The K-SVD method [3] for (P0) has been particularly popular [6–8, 14]. However, (P0) is non-convex and NP-hard, and methods such as K-SVD can be computationally expensive and lack proven convergence guarantees.

In this work, following [15], we consider an aggregate sparsity penalized variant of (P0). Specifically, we define $\mathbf{C} = \mathbf{X}^H$ and replace the sparsity constraints in (P0) with the penalty $\|\mathbf{X}\|_0 \triangleq \sum_{i=1}^N \|\mathbf{x}_i\|_0 = \|\mathbf{C}\|_0 = \sum_{j=1}^J \|\mathbf{c}_j\|_0$. Next, similar to prior works [3, 16], we express the matrix $\mathbf{D}\mathbf{C}^H$ as a Sum of OUter Products (SOUP) $\sum_{j=1}^J \mathbf{d}_j \mathbf{c}_j^H$. Then, the dictionary learning problem is

$$\text{(P1)} \ \min_{\{\mathbf{d}_j, \mathbf{c}_j\}} \left\| \mathbf{Y} - \sum_{j=1}^J \mathbf{d}_j \mathbf{c}_j^H \right\|_F^2 + \lambda^2 \sum_{j=1}^J \|\mathbf{c}_j\|_0 \quad \text{s.t. } \ \|\mathbf{d}_j\|_2 = 1, \ \|\mathbf{c}_j\|_\infty \leq L \, \forall\, j$$

where $\lambda^2$ with $\lambda > 0$, is a weight for the penalty that controls the overall sparsity. The $\ell_\infty$ constraints in (P1) prevent pathologies (e.g., unbounded algorithm iterates) due to the objective being non-coercive [17]. In practice, we set $L$ very large, and the constraints are typically inactive. Unlike (P0), Problem (P1) penalizes the number of non-zeros in the entire coefficient matrix, allowing variable sparsity levels across the signals. This is a general and flexible model for images or image patches, and leads to promising performance in our experiments. An alternative to (P1) replaces the $\ell_0$ penalty with a (coercive) $\ell_1$ penalty [16].

In this work, we use (P1) as a regularizer in inverse problem settings, and investigate the problem of aggregate sparsity penalized dictionary-blind image reconstruction, where both the dictionary (for image patches) and the image are estimated from measurements. While prior work [8] employed (P0) as a regularizer for image reconstruction, the algorithm therein (using K-SVD) lacked convergence analysis and was expensive. Here, we present highly efficient (SOUP-based) block coordinate descent approaches to estimate the variables in the proposed dictionary-blind image reconstruction problems. We provide a convergence analysis of our approaches. Our experiments show the promise of our schemes over some prior methods in compressed sensing-based [18, 19] image reconstruction.

## 2. DICTIONARY-BLIND IMAGE RECONSTRUCTION

This section discusses methods for aggregate sparsity penalized dictionary-blind image reconstruction.

### 2.1. Problem Formulations

Dictionary learning can be used to construct data-driven regularizers for inverse problems. The goal in inverse problems is to estimate an unknown (vectorized) signal or image (or volume) $\mathbf{y} \in \mathbb{C}^p$ from its measurements $\mathbf{z} \in \mathbb{C}^m$. We consider the following general regularized linear inverse problem:

$$\min_{\mathbf{y} \in \mathbb{C}^p} \ \|\mathbf{A}\mathbf{y} - \mathbf{z}\|_2^2 + \zeta(\mathbf{y}) \tag{1}$$

where $\mathbf{A}$ is the sensing matrix for the application. For example, in applications such as computed tomography or magnetic resonance imaging, $\mathbf{A}$ is a Radon transform or a Fourier encoding, respectively.

The regularizer $\zeta(\mathbf{y})$ is used in (1) to capture assumed properties of the image $\mathbf{y}$. Here, we focus on data-driven regularizers based on dictionary learning that have gained interest in recent years [6–8]. In particular, we propose the following dictionary-blind image

This work was supported in part by the following grants: ONR grant N00014-15-1-2141, DARPA Young Faculty Award D14AP00086, ARO MURI grants W911NF-11-1-0391 and 2015-05174-05, NIH grant U01 EB01875301, and a UM-SJTU seed grant.

reconstruction problem that uses a regularizer based on (P1):

$$\text{(P2)} \quad \min_{\mathbf{y},\mathbf{D},\mathbf{X}} \nu \left\|\mathbf{A}\mathbf{y}-\mathbf{z}\right\|_2^2 + \sum_{i=1}^{N} \left\|\mathbf{P}_i\mathbf{y}-\mathbf{D}\mathbf{x}_i\right\|_2^2 + \lambda^2 \left\|\mathbf{X}\right\|_0 \quad \text{s.t.} \quad \left\|\mathbf{d}_j\right\|_2 = 1,\, \left\|\mathbf{x}_i\right\|_\infty \leq L \,\forall\, i,j.$$

Here, $\mathbf{P}_i \in \mathbb{R}^{n\times p}$ is an operator that extracts a $\sqrt{n}\times\sqrt{n}$ patch (in case of a 2D image) of $\mathbf{y}$ as a vector, and $\mathbf{D}\in\mathbb{C}^{n\times J}$ is a (unknown) dictionary. A total of $N$ overlapping patches are assumed, and $\nu > 0$ is a weight. We use $\mathbf{Y}$ here to denote the matrix with columns $\mathbf{P}_i\mathbf{y}$ (patches), and $\mathbf{X}$ (with columns $\mathbf{x}_i$) denotes the corresponding sparse representation of $\mathbf{Y}$. Similarly as in (P1), we can approximate the (unknown) patch data matrix $\mathbf{Y}$ using a SOUP representation.

An alternative to Problem (P2) is the following $\ell_1$ norm-based image reconstruction problem, where $\left\|\mathbf{X}\right\|_1 = \sum_{i=1}^{N}\left\|\mathbf{x}_i\right\|_1$ and $\mu > 0$:

$$\text{(P3)} \quad \min_{\mathbf{y},\mathbf{D},\mathbf{X}} \nu \left\|\mathbf{A}\mathbf{y}-\mathbf{z}\right\|_2^2 + \sum_{i=1}^{N} \left\|\mathbf{P}_i\mathbf{y}-\mathbf{D}\mathbf{x}_i\right\|_2^2 + \mu \left\|\mathbf{X}\right\|_1 \quad \text{s.t.} \quad \left\|\mathbf{d}_j\right\|_2 = 1 \,\forall\, j.$$

The goal in (the nonconvex) (P2) or (P3) is to learn $\mathbf{D}$ and the sparse coefficients, and reconstruct $\mathbf{y}$ using only the measurements $\mathbf{z}$.

### 2.2. Algorithms

We adopt iterative block coordinate descent methods for the dictionary-blind image reconstruction problems. The algorithm for (P1) is used for a subproblem for (P2), and is discussed in that setting. The algorithms for (P2) and (P3) alternate between a *dictionary learning step* and an *image update step*. In the dictionary learning step, we minimize the objectives with respect to $(\mathbf{D},\mathbf{X})$ keeping $\mathbf{y}$ fixed. In the image update step, we solve (P2) or (P3) with respect to $\mathbf{y}$. We describe these steps below.

#### *2.2.1. Dictionary Learning Step*

Minimizing (P2) with respect to $(\mathbf{D},\mathbf{X})$ and substituting $\mathbf{X} = \mathbf{C}^H$ directly yields Problem (P1). We update the dictionary and sparse coefficients using an iterative block coordinate descent method (for (P1)) that updates the columns $\mathbf{c}_j$ (of $\mathbf{C}$) and $\mathbf{d}_j$ (of $\mathbf{D}$) sequentially [17]. Specifically, for each $1 \leq j \leq J$, we first solve for $\mathbf{c}_j$ keeping the other variables fixed (*sparse coding step*), and then solve for $\mathbf{d}_j$ (*dictionary atom update step*).

The sparse coding step involves the following problem, where $\mathbf{E}_j \triangleq \mathbf{Y} - \sum_{k\neq j}\mathbf{d}_k\mathbf{c}_k^H$ is a fixed matrix based on the most recent estimates of the variables:

$$\min_{\mathbf{c}_j\in\mathbb{C}^N} \left\|\mathbf{E}_j - \mathbf{d}_j\mathbf{c}_j^H\right\|_F^2 + \lambda^2\left\|\mathbf{c}_j\right\|_0 \quad \text{s.t.} \quad \left\|\mathbf{c}_j\right\|_\infty \leq L. \tag{2}$$

Assuming $L > \lambda$, the solution to Problem (2) (see [17] for proof) is

$$\hat{\mathbf{c}}_j = \min\left(\left|H_\lambda\left(\mathbf{E}_j^H\mathbf{d}_j\right)\right|, L\mathbf{1}_N\right) \odot e^{j\angle\,\mathbf{E}_j^H\mathbf{d}_j} \tag{3}$$

where $\mathbf{1}_N$ is a vector of ones of length $N$, "$\odot$" denotes element-wise multiplication, $\min(\mathbf{a},\mathbf{u})$ denotes element-wise minimum, and for $\mathbf{c}\in\mathbb{C}^N$, $e^{j\angle\mathbf{c}}\in\mathbb{C}^N$ is computed element-wise, with "$\angle$" denoting the phase. The hard-thresholding operator $H_\lambda(\cdot)$ is defined as

$$\left(H_\lambda(\mathbf{b})\right)_i = \begin{cases} 0, & |b_i| < \lambda \\ b_i, & |b_i| \geq \lambda \end{cases} \tag{4}$$

with $\mathbf{b}\in\mathbb{C}^N$, and subscript $i$ indexes vector entries. The solution (3) is unique if and only if $\mathbf{E}_j^H\mathbf{d}_j$ has no entry with magnitude $\lambda$.

Minimizing (P2) (or (P1)) with respect to $\mathbf{d}_j$ yields the following problem:

$$\min_{\mathbf{d}_j\in\mathbb{C}^n} \left\|\mathbf{E}_j - \mathbf{d}_j\mathbf{c}_j^H\right\|_F^2 \quad \text{s.t.} \quad \left\|\mathbf{d}_j\right\|_2 = 1. \tag{5}$$

The solution for (5) is given (cf. [17] for a proof) as follows [16,20]:

$$\hat{\mathbf{d}}_j = \begin{cases} \frac{\mathbf{E}_j\mathbf{c}_j}{\left\|\mathbf{E}_j\mathbf{c}_j\right\|_2}, & \text{if } \mathbf{c}_j \neq 0 \\ \mathbf{v}, & \text{if } \mathbf{c}_j = 0 \end{cases} \tag{6}$$

where $\mathbf{v}$ can be any vector on the $n$-dimensional unit sphere. In particular, here, we set $\mathbf{v}$ to be the first column of the $n\times n$ identity matrix. The solution above is unique if and only if $\mathbf{c}_j \neq 0$.

In the case of (P3), when minimizing with respect to $(\mathbf{D},\mathbf{X})$, we again set $\mathbf{X} = \mathbf{C}^H$, which yields an $\ell_1$ penalized dictionary learning problem [17]. The dictionary and coefficients are then updated using a similar block coordinate descent method as for (P1). In particular, the coefficients $\mathbf{c}_j$ are updated as $\hat{\mathbf{c}}_j = \max\left(\left|\mathbf{E}_j^H\mathbf{d}_j\right| - \frac{\mu}{2}\mathbf{1}_N, 0\right) \odot e^{j\angle\,\mathbf{E}_j^H\mathbf{d}_j}$. The dictionary learning method in this case is a simple extension of the OS-DL method in [16] to the complex-valued setting.

#### *2.2.2. Image Update Step*

Minimizing (P2) (or (P3)) with respect to $\mathbf{y}$ involves the following optimization problem:

$$\min_{\mathbf{y}} \nu\left\|\mathbf{A}\mathbf{y}-\mathbf{z}\right\|_2^2 + \sum_{i=1}^{N}\left\|\mathbf{P}_i\mathbf{y}-\mathbf{D}\mathbf{x}_i\right\|_2^2. \tag{7}$$

This is a least squares problem with the following normal equation:

$$\left(\sum_{i=1}^{N}\mathbf{P}_i^T\mathbf{P}_i \;+\; \nu\,\mathbf{A}^H\mathbf{A}\right)\mathbf{y} = \sum_{i=1}^{N}\mathbf{P}_i^T\mathbf{D}\mathbf{x}_i + \nu\,\mathbf{A}^H\mathbf{z}. \tag{8}$$

When periodically positioned, overlapping image patches (patch overlap stride [8] denoted by $r$) are used, and the patches that overlap the image boundaries 'wrap around' on the opposite side of the image [8], then typically $\sum_{i=1}^{N}\mathbf{P}_i^T\mathbf{P}_i = \beta I$, with $\beta = \frac{n}{r^2}$. In general, the unique solution to (8) can be found using techniques such as conjugate gradients (CG). In several applications (denoising, inpainting, etc.), the solution to (8) can be found efficiently [6, 8]. For example, in single coil compressed sensing MRI [21], where $\mathbf{A} = \mathbf{F}_{\text{u}} \in \mathbb{C}^{m\times p}$ ($m \ll p$), the undersampled Fourier encoding matrix, the measurements $\mathbf{z}$ are samples in Fourier space (k-space) of $\mathbf{y}$, and we assume for simplicity that $\mathbf{z}$ is obtained by subsampling on a Cartesian grid. Denoting by $\mathbf{F}\in\mathbb{C}^{p\times p}$ the full Fourier encoding matrix with $\mathbf{F}^H\mathbf{F} = \mathbf{I}$ (normalized), we get $\mathbf{F}\mathbf{F}_{\text{u}}^H\mathbf{F}_{\text{u}}\mathbf{F}^H$ is a diagonal matrix of ones and zeros, with ones at entries corresponding to sampled k-space locations. Using this in (8) yields the following solution in Fourier space [17] with $\mathbf{S} \triangleq \mathbf{F}\sum_{i=1}^{N}\mathbf{P}_i^T\mathbf{D}\mathbf{x}_i$ and $\mathbf{S}_0 \triangleq \mathbf{F}\mathbf{F}_{\text{u}}^H\mathbf{z}$:

$$\mathbf{F}\mathbf{y}\,(k_1,k_2) = \begin{cases} \frac{\mathbf{S}(k_1,k_2)}{\beta} & ,\,(k_1,k_2)\notin\Omega \\ \frac{\mathbf{S}(k_1,k_2)+\nu\,\mathbf{S}_0(k_1,k_2)}{\beta+\nu} & ,\,(k_1,k_2)\in\Omega \end{cases} \tag{9}$$

where $(k_1,k_2)$ indexes k-space locations (2D coordinates), and $\Omega$ is the subset of k-space sampled. The $\mathbf{y}$ solving (8) is obtained by an inverse FFT of $\mathbf{F}\mathbf{y}$ in (9).

We refer to the algorithms for (P2) and (P3) as SOUP-DILLO (SOUP DIctionary Learning with $\ell_0$ "norm") and SOUP-DILLI image reconstruction algorithms, respectively. In applications such as

(inpainting or) single coil MRI, the cost per outer iteration of the algorithms is dominated by the dictionary learning step, for which (with $J \propto n$) the cost scales as $O(KNn^2)$, with $K$ the number of times $\mathbf{D}$ is updated in dictionary learning. In contrast, recent iterative image reconstruction methods involving K-SVD (e.g., DLMRI [8]) have a worse cost per outer iteration of $O(KNn^3)$.

### 2.3. Convergence Results

Here, we discuss the convergence behavior of the algorithms for (P2) and (P3). Corresponding convergence results for the dictionary learning methods (e.g., for (P1)) and all proofs can be found in [17]. Due to the non-convexity involved, recent results on convergence of block coordinate descent methods [22] do not immediately apply (e.g., the assumptions in [22] such as block-wise quasiconvexity or other conditions do not hold). Here, we discuss the convergence of our algorithms to the critical points [23] in the problems.

Upon substituting $\mathbf{C} = \mathbf{X}^H$, the constraints $\|\mathbf{d}_j\|_2 = 1$, and $\|\mathbf{c}_j\|_\infty \leq L$ in the problems can instead be added as penalties in the costs using barrier functions $\chi(\mathbf{d}_j)$ and $\psi(\mathbf{c}_j)$ that take the value $+\infty$ when the corresponding constraint is violated, and are zero otherwise. Problem (P2) is then written in unconstrained form with objective

$$g(\mathbf{C},\mathbf{D},\mathbf{y}) = \nu \left\|\mathbf{A}\mathbf{y}-\mathbf{z}\right\|_2^2 + \left\|\mathbf{Y} - \textstyle\sum_{j=1}^{J} \mathbf{d}_j \mathbf{c}_j^H\right\|_F^2 + \lambda^2 \sum_{j=1}^{J} \left\|\mathbf{c}_j\right\|_0 + \sum_{j=1}^{J} \chi(\mathbf{d}_j) + \sum_{j=1}^{J} \psi(\mathbf{c}_j). \quad (10)$$

Recall that $\mathbf{Y}$ denotes the matrix with patches $\mathbf{P}_i\mathbf{y}$ for $1 \leq i \leq N$, as its columns. Problem (P3) is also written similarly in an unconstrained form with objective $\tilde{g}(\mathbf{C},\mathbf{D},\mathbf{y})$. Assume that the initialization $(\mathbf{C}^0,\mathbf{D}^0)$ in the algorithms satisfies problem constraints.

**Theorem 1** *Let $\left\{\mathbf{C}^t,\mathbf{D}^t,\mathbf{y}^t\right\}$ denote the iterate sequence generated by the SOUP-DILLO image reconstruction Algorithm with measurements $\mathbf{z} \in \mathbb{C}^m$ and initial $(\mathbf{C}^0,\mathbf{D}^0,\mathbf{y}^0)$. Then we have*

*(i) The objective sequence $\left\{g^t\right\}$ with $g^t \triangleq g\left(\mathbf{C}^t,\mathbf{D}^t,\mathbf{y}^t\right)$ is monotone decreasing, and converges to a finite value, say $g^* = g^*(\mathbf{C}^0,\mathbf{D}^0,\mathbf{y}^0)$.*

*(ii) The iterate sequence is bounded, and all its accumulation points are equivalent in the sense that they achieve the exact same value $g^*$ of the objective.*

*(iii) Each accumulation point $(\mathbf{C},\mathbf{D},\mathbf{y})$ of the iterate sequence satisfies*

$$\mathbf{y} \in \arg\min_{\tilde{\mathbf{y}}} \; g(\mathbf{C},\mathbf{D},\tilde{\mathbf{y}}). \quad (11)$$

*(iv) As $t \to \infty$, $\left\|\mathbf{y}^t - \mathbf{y}^{t-1}\right\|_2$ converges to zero.*

*(v) Suppose each accumulation point $(\mathbf{C},\mathbf{D},\mathbf{y})$ is such that the matrix $\mathbf{B}$ with columns $\mathbf{b}_j = \mathbf{E}_j^H\mathbf{d}_j$ and $\mathbf{E}_j = \mathbf{Y} - \mathbf{D}\mathbf{C}^H + \mathbf{d}_j\mathbf{c}_j^H$, has no entry with magnitude $\lambda$. Then every accumulation point of the iterates is a critical point of $g$. Moreover, $\left\|\mathbf{D}^t - \mathbf{D}^{t-1}\right\|_F \to 0$ and $\left\|\mathbf{C}^t - \mathbf{C}^{t-1}\right\|_F \to 0$ as $t \to \infty$.*

Statements (i) and (ii) establish that for each initial $(\mathbf{C}^0,\mathbf{D}^0,\mathbf{y}^0)$, the iterate sequence in the algorithm for (P3) converges to an equivalence class (common objective value) of accumulation points. Statements (iii) and (iv) establish that the accumulation points are global minimizers of $g$ with respect to $\mathbf{y}$, and $\left\|\mathbf{y}^t - \mathbf{y}^{t-1}\right\|_2 \to 0$. Statement (v) shows that the iterates converge to the critical points of $g$ and $\left\|\mathbf{D}^t - \mathbf{D}^{t-1}\right\|_F \to 0$ and $\left\|\mathbf{C}^t - \mathbf{C}^{t-1}\right\|_F \to 0$. Statement (v) uses a simple condition on the accumulation points of the iterates that is equivalent to assuming that for every $j$, there is a unique minimizer of $g$ with respect to $\mathbf{c}_j$ with all other variables fixed to their values in the accumulation point $(\mathbf{C},\mathbf{D},\mathbf{y})$.

For the algorithm for (P3), the iterate sequence for an initial $(\mathbf{C}^0,\mathbf{D}^0,\mathbf{y}^0)$ converges easily to an equivalence class (corresponding to common cost $\tilde{g}^* = \tilde{g}^*(\mathbf{C}^0,\mathbf{D}^0,\mathbf{y}^0)$) of critical points of $\tilde{g}$.

**Theorem 2** *Let $\left\{\mathbf{C}^t,\mathbf{D}^t,\mathbf{y}^t\right\}$ denote the iterate sequence generated by the SOUP-DILLI image reconstruction Algorithm for (P3) with measurements $\mathbf{z} \in \mathbb{C}^m$ and initial $(\mathbf{C}^0,\mathbf{D}^0,\mathbf{y}^0)$. Then, the iterate sequence converges to an equivalence class of critical points of $\tilde{g}(\mathbf{C},\mathbf{D},\mathbf{y})$. Moreover, $\left\|\mathbf{D}^t - \mathbf{D}^{t-1}\right\|_F \to 0$, $\left\|\mathbf{C}^t - \mathbf{C}^{t-1}\right\|_F \to 0$, and $\left\|\mathbf{y}^t - \mathbf{y}^{t-1}\right\|_2 \to 0$ as $t \to \infty$.*

## 3. NUMERICAL EXPERIMENTS

This section presents numerical results illustrating the convergence and performance of the proposed methods for (P2) and (P3) (called SOUP-DILLO MRI and SOUP-DILLI MRI) for compressed sensing MRI. The usefulness of the dictionary learning methods (e.g., for (P1)) for sparse representation or denoising is demonstrated elsewhere [17, 24].

### 3.1. SOUP-DILLO MRI and SOUP-DILLI MRI Convergence

Here, we consider the complex-valued reference image in Fig. 1(c), and perform 2.5 fold undersampling of the k-space of the reference. We study the behavior of SOUP-DILLO MRI and SOUP-DILLI MRI using $6 \times 6$ image patches with stride $r = 1$ (with patch wrap around), $\nu = 10^6/p$ (with $p$ the number of image pixels), and learning a $36 \times 144$ dictionary (with 1 iteration within dictionary learning) with $\lambda = 0.08$ or $\mu = 0.08$. We set $\mathbf{y}^0 = \mathbf{A}^\dagger\mathbf{z}$, $\mathbf{C}^0 = \mathbf{0}$, and $\mathbf{D}^0$ was a square DCT concatenated with normalized random vectors.

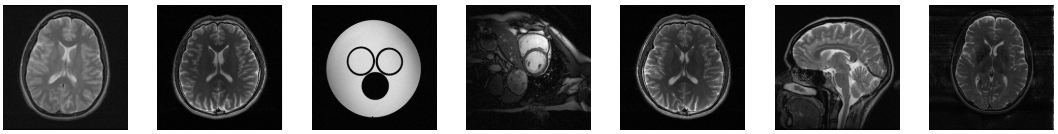
Fig. 2 shows the behavior of the proposed reconstruction methods. The objective function values decreased monotonically and quickly for both methods. The reconstruction PSNR (based on image magnitudes) improved significantly over the iterations. While the initial reconstruction shows large artifacts, these are removed in the SOUP-DILLO MRI result. The $\ell_1$ method achieves a lower PSNR than the $\ell_0$ case. The sparsity fraction for the learned $C$ was $5\%$ for (P2) and $16\%$ for (P3). Although larger values of $\mu$ decrease the eventual sparsity in SOUP-DILLI MRI, we found that the PSNR also degrades for such settings.

Finally, the changes between successive $\mathbf{y}$ iterates (Fig. 2) decreased to small values (similar behavior observed for the $\mathbf{C}$ or $\mathbf{D}$ iterates) for our algorithms, as predicted by Theorems 1 and 2 (a necessary condition for sequence convergence). The dictionary learned in SOUP-DILLO MRI is shown along with the initial dictionary. The real and imaginary parts of the learned $\mathbf{D}$ display novel frequency or edge structures that were learnt from a few k-space measurements.

### 3.2. Dictionary-Blind Compressed Sensing Results

Here, we consider images a-g in Fig. 1 and simulate k-space undersampling. We compare the reconstructions obtained by our methods to those obtained with the K-SVD based DLMRI [8], Sparse MRI [21] and PANO [25]. We used the built-in parameter settings in the public implementations of Sparse MRI [26] and PANO [27], which performed well in our experiments. We used the zero-filling reconstruction as the initial guide image in PANO [27].

For DLMRI [28], we used $6 \times 6$ patches, a $36 \times 144$ $\mathbf{D}$ and ran 45 iterations of the algorithm. The patch stride $r = 1$, and 14400



(a) (b) (c) (d) (e) (f) (g)

**Fig. 1**. MR Data (magnitudes displayed) [17]: (a) and (f) are $512 \times 512$, and the rest are $256 \times 256$. (b) and (g) are rotated by $90^\circ$ for display.

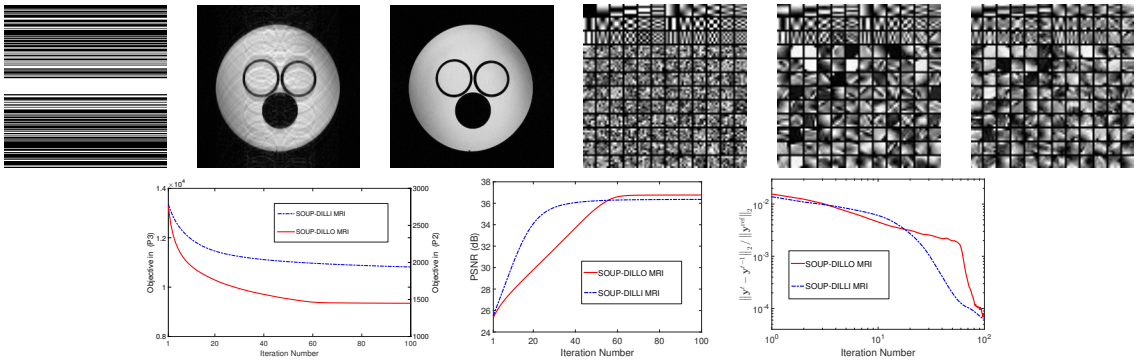



**Fig. 2**. SOUP-DILLO MRI and SOUP-DILLI MRI. Top row (left to right): sampling mask; magnitude of $\mathbf{y}^0$ (24.9 dB); SOUP DILLO MRI reconstruction magnitude (36.8 dB); initial dictionary; real part of learned $\mathbf{D}$ (columns shown as patches) in SOUP-DILLO MRI; along with the imaginary part. Bottom row (left to right): objectives; reconstruction PSNR; and $\left\|\mathbf{y}^t - \mathbf{y}^{t-1}\right\|_2 / \left\|\mathbf{y}^{\text{ref}}\right\|_2$ with $\mathbf{y}^{\text{ref}}$ the reference.

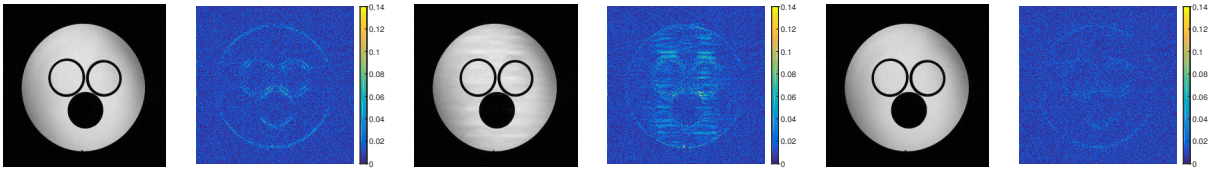


**Fig. 3**. Cartesian sampling with 2.5x undersampling (from Table 1). Reconstructions (magnitudes): DLMRI [8] (first); PANO [25] (third); and SOUP-DILLO MRI (fifth). Reconstruction error maps: DLMRI (second); PANO (fourth); and SOUP-DILLO MRI (sixth).

| Image | UF | Zero-filling | Sparse MRI | PANO | DLMRI | (P3) | (P2) |
|---|---|---|---|---|---|---|---|
| a | 7x | 27.9 | 28.6 | 31.1 | 31.1 | 30.8 | **31.1** |
| b | 2.5x | 27.7 | 31.6 | 41.3 | 40.2 | 38.5 | **42.3** |
| c | 2.5x | 24.9 | 29.9 | 34.8 | 36.7 | 36.6 | **37.3** |
| c | 4x | 25.9 | 28.8 | 32.3 | 32.1 | 32.2 | **32.3** |
| d | 2.5x | 29.5 | 32.1 | 36.9 | 38.1 | 36.7 | **38.4** |
| e | 2.5x | 28.1 | 31.7 | 40.0 | 38.0 | 37.9 | **41.5** |
| f | 5x | 26.3 | 27.4 | 30.4 | 30.5 | 30.3 | **30.6** |
| g | 2.5x | 32.8 | 39.1 | 41.6 | 41.7 | 42.2 | **43.2** |

**Table 1**. Reconstruction PSNRs (dB) for several methods for various images and undersampling factors (UF). We used variable density 2D random sampling [8] for image (f) and variable density Cartesian sampling for other cases. The best PSNRs are marked in bold.

randomly selected patches were used during the learning step (with 20 iterations of K-SVD). Mean-subtraction was not performed for patches prior to the dictionary learning step. A maximum sparsity level ($s = 7$ per patch) is employed together with an error threshold (for sparse coding) during dictionary learning (see [17]). All patches are sparse coded after learning with the same error threshold as in learning and a relaxed maximum sparsity of 14. These settings (all else as per the indications in [28]) worked quite well for DLMRI.

For SOUP-DILLO MRI and SOUP-DILLI MRI, we used settings as in Section 3.1 and ran 45 (outer) iterations of the methods. We found that larger $\lambda$ or $\mu$ values during initial outer iterations led to faster/better aliasing removal. Hence, we vary $\lambda$ from 0.35 to 0.01 over iterations, except for Figs. 1(a), (c), and (f) (noisier), where $\lambda$ varies from 0.35 to 0.04. We used $\mu = \lambda/1.4$, and 5 (inner) dictionary learning iterations for (P2) and 1 such iteration (optimal) for (P3).

Table 1 compares the reconstruction PSNRs for various methods for several cases. The proposed SOUP-DILLO MRI provides large PSNR improvements over DLMRI, PANO, and Sparse MRI. It also outperforms the $\ell_1$ method SOUP-DILLI MRI (by 1.4 dB on average), indicating benefits for the $\ell_0$ approach in practice. SOUP-DILLO MRI (average runtime of 2180 seconds) was also faster than the previous DLMRI (average runtime of 3156 seconds). Fig. 3 shows the reconstructions and error maps (magnitude of the difference between the magnitudes of the reconstructed and reference images) for various methods for an example in Table 1. SOUP-DILLO MRI clearly achieves smaller distortions than the other methods.

## 4. CONCLUSIONS

This paper investigated fast methods for dictionary learning for inverse problems, with convergence analysis. The proposed SOUP-DILLO image reconstruction method outperformed benchmarks involving the K-SVD algorithm, as well as some other recent methods for compressed sensing MRI.

## 5. REFERENCES


[1] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, vol. 381, no. 6583, pp. 607–609, 1996.

[2] K. Engan, S.O. Aase, and J.H. Hakon-Husoy, "Method of optimal directions for frame design," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1999, pp. 2443–2446.

[3] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Transactions on signal processing*, vol. 54, no. 11, pp. 4311–4322, 2006.

[4] M. Yaghoobi, T. Blumensath, and M. Davies, "Dictionary learning for sparse approximations with the majorization method," *IEEE Transaction on Signal Processing*, vol. 57, no. 6, pp. 2178–2191, 2009.

[5] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online learning for matrix factorization and sparse coding," *J. Mach. Learn. Res.*, vol. 11, pp. 19–60, 2010.

[6] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, 2006.

[7] J. Mairal, M. Elad, and G. Sapiro, "Sparse representation for color image restoration," *IEEE Trans. on Image Processing*, vol. 17, no. 1, pp. 53–69, 2008.

[8] S. Ravishankar and Y. Bresler, "MR image reconstruction from highly undersampled k-space data by dictionary learning," *IEEE Trans. Med. Imag.*, vol. 30, no. 5, pp. 1028–1041, 2011.

[9] Y. Wang, Y. Zhou, and L. Ying, "Undersampled dynamic magnetic resonance imaging using patch-based spatiotemporal dictionaries," in *IEEE 10th International Symposium on Biomedical Imaging (ISBI)*, April 2013, pp. 294–297.

[10] R. Gribonval and K. Schnass, "Dictionary identification–sparse matrix-factorization via $l_1$ -minimization," *IEEE Trans. Inform. Theory*, vol. 56, no. 7, pp. 3523–3539, 2010.

[11] D. Barchiesi and M. D. Plumbley, "Learning incoherent dictionaries for sparse approximation using iterative projections and rotations," *IEEE Transactions on Signal Processing*, vol. 61, no. 8, pp. 2055–2065, 2013.

[12] K. Skretting and K. Engan, "Recursive least squares dictionary learning algorithm," *IEEE Transactions on Signal Processing*, vol. 58, no. 4, pp. 2121–2130, 2010.

[13] L. N. Smith and M. Elad, "Improving dictionary learning: Multiple dictionary updates and coefficient reuse," *IEEE Signal Processing Letters*, vol. 20, no. 1, pp. 79–82, Jan 2013.

[14] M. Protter and Michael Elad, "Image sequence denoising via sparse and redundant representations," *IEEE Trans. on Image Processing*, vol. 18, no. 1, pp. 27–36, 2009.

[15] C. Bao, H. Ji, Y. Quan, and Z. Shen, "L0 norm based dictionary learning by proximal methods with global convergence," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 3858–3865.

[16] M. Sadeghi, M. Babaie-Zadeh, and C. Jutten, "Learning overcomplete dictionaries based on atom-by-atom updating," *IEEE Trans. on Signal Process.*, vol. 62, no. 4, pp. 883–891, 2014.

[17] S. Ravishankar, R. R. Nadakuditi, and J. A. Fessler, "Efficient sum of outer products dictionary learning (SOUP-DIL) and its application to inverse problems," 2016, Submitted.

[18] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[19] D. Donoho, "Compressed sensing," *IEEE Trans. Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[20] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the k-svd algorithm using batch orthogonal matching pursuit," `http://www.cs.technion.ac.il/~ronrubin/Publications/KSVD-OMP-v2.pdf`, 2008, Technion - Computer Science Department - Technical Report.

[21] M. Lustig, D. L. Donoho, and J. M. Pauly, "Sparse MRI: The application of compressed sensing for rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–1195, 2007.

[22] P. Tseng, "Convergence of a block coordinate descent method for nondifferentiable minimization," *J. Optim. Theory Appl.*, vol. 109, no. 3, pp. 475–494, 2001.

[23] R. T. Rockafellar and Roger J.-B. Wets, *Variational Analysis, 1st edn.*, Springer-Verlag, 1997.

[24] S. Ravishankar, R. R. Nadakuditi, and J. A. Fessler, "Efficient sum of outer products dictionary learning (SOUP-DIL) - the $\ell_0$ method," 2015, `http://arxiv.org/abs/1511.08842`.

[25] X. Qu, Y. Hou, F. Lam, D. Guo, J. Zhong, and Z. Chen, "Magnetic resonance image reconstruction from undersampled measurements using a patch-based nonlocal operator," *Medical Image Analysis*, vol. 18, no. 6, pp. 843–856, Aug 2014.

[26] M. Lustig, "Michael Lustig home page," `http://www.eecs.berkeley.edu/~mlustig/Software.html`, 2014, [Online; accessed October, 2014].

[27] X. Qu, "PANO Code," `http://www.quxiaobo.org/project/CS_MRI_PANO/Demo_PANO_SparseMRI.zip`, 2014, [Online; accessed May, 2015].

[28] S. Ravishankar and Y. Bresler, "DLMRI - Lab: Dictionary learning MRI software," `http://www.ifp.illinois.edu/~yoram/DLMRI-Lab/DLMRI.html`, 2013, [Online; accessed October, 2014].