Explicit Shift-Invariant Dictionary Learning

Cristian Rusu, Bogdan Dumitrescu, and Sotirios A. Tsaftaris

***Abstract*—In this letter we give efficient solutions to the construction of structured dictionaries for sparse representations. We study circulant and Toeplitz structures and give fast algorithms based on least squares solutions. We take advantage of explicit circulant structures and we apply the resulting algorithms to shift-invariant learning scenarios. Synthetic experiments and comparisons with state-of-the-art methods show the superiority of the proposed methods.**

***Index Terms*—Dictionary learning, shift-invariant learning, sparse representations.**

I. Introduction

IN THIS letter we study the dictionary learning problem in the context of sparse representations [1]. Given a dataset $\boldsymbol{Y} \in \mathbb{R}^{n\times N}$ and a target sparsity $s \in \mathbb{N}^*$, $s \ll n$, the optimization problem can be stated as:

$$\begin{array}{ll} \underset{\boldsymbol{D},\Gamma}{\text{minimize}} & \|\boldsymbol{Y} - \boldsymbol{D}\Gamma\|_F \\ \text{subject to} & \|\gamma_i\|_0 \le s,\ 1 \le i \le N \\ & \|\boldsymbol{d}_j\|_2 = 1,\ 1 \le j \le m, \end{array} \tag{1}$$

where $\|\gamma_i\|_0$ is the number of non-zero elements of columns γ_i, also known as the ℓ_0 pseudo-norm, $\|\boldsymbol{E}\|_F^2 = \sum_i \sum_j e_{ij}^2$ is the Frobenius norm, columns $\boldsymbol{d}_j$ are called the atoms of the dictionary $\boldsymbol{D} \in \mathbb{R}^{n\times m}$ and the sparse representations matrix is $\boldsymbol{\Gamma} \in \mathbb{R}^{m\times N}$. The dimensions obey $n \le m \ll N$.

To solve this challenging problem, most batch dictionary learning algorithms follow a two-step alternating optimization procedure: keep the dictionary fixed and find the sparse representations in $\boldsymbol{\Gamma}$ using a sparse approximation algorithm, e.g., Orthogonal Matching Pursuit (OMP) [2]; keep the representations $\boldsymbol{\Gamma}$ fixed and update the dictionary $\boldsymbol{D}$, either all at once [3] or one atom at a time [4].

In this letter we study in depth two dictionary structures: circulant and Toeplitz. Keeping the alternate optimization process, we provide efficient solvers that exploit maximally such structures. We apply these structures to the construction of explicit shift-invariant learning algorithms. While other works [5]–[11] treat shift-invariance in an indirect way, the proposed explicit construction offers clear performance advantages in experimental comparisons.

The problem of learning shift-invariant structures has been extensively studied. For example, in [5], [9] the authors propose an extension of K-SVD to accommodate for the possible shifts in the given dataset. Similarly, a variation of the K-SVD algorithm is given in [6], allowing for shift-invariant structures by using a graph-based dictionary and an atom update procedure called Shift-Invariant Singular Value Decomposition (SISVD). The MoTIF learning algorithm [7] is based on the idea of learning translated dictionaries (all possible translations of a few given generating functions, called kernels) such that they are highly correlated in mean with the signals in the training set. In the end, the problem reduces to solving a generalized eigenvalue problem. Finally, in [8] we encounter a shift-invariant dictionary learning method based on tree-structured dictionaries. Applications of these shift-invariant learning methods include audio processing [5], image analysis [7], [8] and feature extraction [11].

Circulant structures were previously considered in the context of nonnegative matrix factorization [12] (NMF). The union of circulants is explicitly used to impose structure and speedup the classical multiplicative update equations of NMF.

Our approach is different from all the above, not only in the overall dictionary learning algorithm, but also in the way we take advantage of the imposed structures.

The letter is organized as follows: Section II presents C–DLA, the proposed circulant dictionary learning algorithm, Section III uses the previously stated results to construct UC–DLA, an algorithm capable of learning a union of circulants, Section IV extends the C–DLA to Toeplitz structures, Section V presents the results and Section VI offers conclusions.

II. Circulant Dictionary Learning

The first class of dictionaries that we consider are the circulants $\boldsymbol{C} \in \mathbb{R}^{n\times n}$. These square matrices are completely defined by the first column $\boldsymbol{c} \in \mathbb{R}^n$; every column is a circular down shift of the first one called kernel: $\boldsymbol{C} = \text{circ}(\boldsymbol{c}) \stackrel{\text{def}}{=} [\boldsymbol{c}\ \boldsymbol{P}\boldsymbol{c}\ \boldsymbol{P}^2\boldsymbol{c} \ldots \boldsymbol{P}^{n-1}\boldsymbol{c}]$ (so-called right circulant), where $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ is the orthonormal permutation matrix that circularly downshifts a column vector by left matrix-vector multiplication. The main property of such matrices is their eigenvalue factorization:

$$\boldsymbol{C} = n^{-1}\boldsymbol{F}^H \Sigma \boldsymbol{F}, \tag{2}$$

where $\boldsymbol{F} \in \mathbb{C}^{n\times n}$ is the orthogonal Fourier matrix ($\boldsymbol{F}\boldsymbol{F}^H = n\boldsymbol{I}_n$), $\boldsymbol{F}^H$ is the conjugate transpose of $\boldsymbol{F}$ and the diagonal matrix $\Sigma = \text{diag}(\boldsymbol{F}\boldsymbol{c}) \in \mathbb{C}^{n\times n}$ has diagonal $\boldsymbol{\sigma} \in \mathbb{C}^n$. Due to the properties outlined before, based on the Fourier transform, a fundamental outcome related to circulant matrices is the fast matrix-vector multiplication:

$$\boldsymbol{C}\gamma = \text{IFFT}(\text{FFT}(\boldsymbol{c}) \odot \text{FFT}(\gamma)), \tag{3}$$

Manuscript received August 13, 2013; revised October 11, 2013; accepted October 18, 2013. Date of of current version November 05, 2013. The work of B. Dumitrescu was supported by Romanian National Authority for Scientific Research, CNCS UEFISCDI, under Project PN-II-ID-PCE-2011-3-0400. The work of S. A. Tsaftaris was supported by the U.S. National Institutes of Health under Grant 2R01HL091989-05. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Petros T. Boufounos.

C. Rusu is with the IMT Institute for Advanced Studies, Lucca, Italy (e-mail: cristian.rusu@imtlucca.it).

S. A. Tsaftaris is with the IMT Institute for Advanced Studie,s Lucca, Italy, and also with the Department of Electrical Engineering and Computer Science, Northwestern University, Evanston, IL 60201 USA (e-mail: s.tsaftaris@imtlucca.it).

B. Dumitrescu is with the Department of Automatic Control and Computers, University Politehnica of Bucharest, Bucharest, Romania (e-mail: bogdan.dumitrescu@acse.pub.ro).

Source code available online at http://prian.imtlucca.it/C_Rusu.html.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2013.2288788

1070-9908 © 2013 IEEE

where $\odot$ denotes the elementwise multiplication.

This section is primarily concerned with minimizing $\|\boldsymbol{Y} - \boldsymbol{C}\boldsymbol{\Gamma}\|_F$, where $\boldsymbol{C}$ is right circulant and $\boldsymbol{Y}, \boldsymbol{\Gamma}$ are fixed.

In the context of dictionary learning algorithms, we rely on the eigenvalue factorization (2) and the fact that all multiplications with orthogonal matrices preserve the matrix norms:

$$\|\boldsymbol{Y} - n^{-1}\boldsymbol{F}^H\Sigma\boldsymbol{F}\boldsymbol{\Gamma}\|_F = \|\tilde{\boldsymbol{Y}} - \Sigma\tilde{\boldsymbol{\Gamma}}\|_F. \tag{4}$$

In this new formulation, the tilde matrices contain the Fourier transforms (of $\boldsymbol{Y}$ and $\boldsymbol{\Gamma}$). Since Σ is a diagonal, the minimization of the Frobenius norm can be done decentralized, by taking the corresponding rows of $\tilde{\boldsymbol{Y}}$ and $\tilde{\boldsymbol{\Gamma}}$.

Furthermore, since we are considering only real dictionaries $\boldsymbol{C}$, we have the symmetry restrictions: $(\sigma_1, \sigma_2, \ldots, \sigma_n) = (a_1, a_2 + b_2 i, a_3 + b_3 i, \ldots, a_3 - b_3 i, a_2 - b_2 i)$ where $\sigma_{n/2+1}$, if n is even, and σ_1 are purely real. Attaching this structure and matching the corresponding rows in (4) in pairs the problem reduces to solving a series of least squares problems involving the rows of $\tilde{\boldsymbol{Y}}$ and $\tilde{\boldsymbol{\Gamma}}$:

$$\begin{pmatrix} \tilde{\boldsymbol{y}}_k^T \\ \tilde{\boldsymbol{y}}_j^T \end{pmatrix} = \begin{pmatrix} a_k + b_k i \\ a_k - b_k i \end{pmatrix} \begin{pmatrix} \tilde{\gamma}_k^T \\ \tilde{\gamma}_j^T \end{pmatrix}, \tag{5}$$

with $j = n - k + 2$ for $k = 2, \ldots, \lceil n/2 \rceil$. When the input data are real, the conjugate symmetric structure is implicitly imposed by the Fourier transforms taken on the columns of $\boldsymbol{Y}$ and $\boldsymbol{\Gamma}$–the corresponding rows (k, j) are complex conjugate. The solution of (5) is given by $a_k + b_k i = \tilde{\gamma}_k^H \tilde{\boldsymbol{y}}_k / \|\tilde{\gamma}_k\|_2^2$.

With these observations, we can proceed to provide a full description of the circulant dictionary learning algorithm. For simplicity, we assume n even herein.

C–DLA: Circulant Dictionary Learning Algorithm: Given the dataset $\boldsymbol{Y} \in \mathbb{R}^{n \times N}$, the number of iterations K and the target sparsity s, design the circulant dictionary $\boldsymbol{C} \in \mathbb{R}^{n \times n}$ and the representations $\boldsymbol{\Gamma} \in \mathbb{R}^{n \times N}$, $\|\gamma_i\|_0 \leq s$, $i = 1, \ldots, N$, such that $\epsilon = \|\boldsymbol{Y} - \boldsymbol{C}\boldsymbol{\Gamma}\|_F$ is reduced.

- **Initialization**: Construct kernel $\boldsymbol{c}$ by extracting the principal component of $\boldsymbol{Y}$. Compute $\tilde{\boldsymbol{Y}} = \text{FFT}(\boldsymbol{Y})$.
- **Iterations**: For $1, \ldots, K$
  1) Produce the sparse representations $\boldsymbol{\Gamma}$ by the OMP.
  2) Compute $a_k + b_k i = \tilde{\gamma}_k^H \tilde{\boldsymbol{y}}_k / \|\tilde{\gamma}_k\|_2^2$, with $k = 1, \ldots, n/2 + 1$ and mirror the conjugate to produce $\boldsymbol{\sigma}$.
  3) Normalize $\boldsymbol{\sigma} = \sqrt{n}\|\boldsymbol{\sigma}\|_2^{-1}\boldsymbol{\sigma}$. Update error term ϵ. ■

The dictionary update and sparse approximation steps can avoid constructing the whole matrix $\boldsymbol{C}$. When using the OMP algorithm, efficient publicly available implementations (such as the OMP-box [13]) use as inputs the projections $\boldsymbol{C}^T\boldsymbol{Y}$ and the Gram matrix $\boldsymbol{G} = \boldsymbol{C}^T\boldsymbol{C}$. Since $\tilde{\boldsymbol{Y}}$ and $\boldsymbol{\sigma}$ are both already computed, evaluating $\boldsymbol{C}^T\boldsymbol{Y}$ reduces to elementwise multiplications and inverse Fourier transforms (see (3)). Note that the symmetric Gram matrix is also circulant, i.e., $\boldsymbol{G} = \text{circ}(\boldsymbol{g})$, with $\boldsymbol{g} = \text{IFFT}(\bar{\boldsymbol{\sigma}} \odot \boldsymbol{\sigma}) = \text{IFFT}(|\boldsymbol{\sigma}|^2)$. Observe that while $\tilde{\boldsymbol{Y}}$ is computed only once, we need to compute $\tilde{\boldsymbol{\Gamma}}$ at every iteration. The structure of the dictionary simplifies also other operations, for example the computation of the mutual coherence becomes: $\mu(\boldsymbol{C}) = \max_{i<j} |G_{ij}| = n\|\boldsymbol{\sigma}\|_2^{-2} \max_{i=2,\ldots,n/2+1} |g_i|$. The final, explicit dictionary is $\boldsymbol{C} = \text{circ}(\text{IFFT}(\boldsymbol{\sigma}))$.

III. Shift-Invariant Dictionary Learning

In this section we explore the application of circulant matrices in the sparse representations context, with a focus on shift-invariant learning. Observe that the simplest shift-invariant dictionary is actually a circulant.

Remark 1: With a circulant $\boldsymbol{C}$, the OMP is shift-invariant. More precisely, denoting $\gamma = \text{OMP}(\boldsymbol{y})$ the solution with $\|\gamma\|_0 = s$ given by OMP for the input vector $\boldsymbol{y}$, then for all $q \in \mathbb{N}$ we have that $\boldsymbol{P}^q\gamma = \text{OMP}(\boldsymbol{P}^q\boldsymbol{y})$.

Proof: We denote with $\boldsymbol{r}$ the residual of the sparse representation obtained with OMP at each of its iterations, which is initially $\boldsymbol{r} = \boldsymbol{y}$ and finally $\boldsymbol{r} = \boldsymbol{y} - \boldsymbol{C}\gamma$. We show by induction that the residual of the shifted problem is always $\boldsymbol{P}^q\boldsymbol{r}$. Initially, this is obviously true. Let us assume that this is also true after a number of iterations. Since $\boldsymbol{P}^q$ is an orthonormal matrix, then we have $\boldsymbol{c}_j^T\boldsymbol{r} = (\boldsymbol{P}^q\boldsymbol{c}_j)^T\boldsymbol{P}^q\boldsymbol{r}$. Hence, if in the selection step of OMP the column $\boldsymbol{c}_{j_0}$ is chosen for the original problem, with $j_0 = \arg\max_j |\boldsymbol{c}_j^T\boldsymbol{r}|$, then $\boldsymbol{P}^q\boldsymbol{c}_{j_0} = \boldsymbol{c}_{(j_0+q) \bmod n}$ is chosen for the shifted problem. Since all selected columns in the shifted problem are the columns selected in the original problem multiplied by (the orthogonal) $\boldsymbol{P}^q$, the least squares problem produces the residual $\boldsymbol{P}^q\boldsymbol{r}$, validating the initial assumption. As $\boldsymbol{P}^q$ is circulant and commutes with $\boldsymbol{C}$, it follows that $\boldsymbol{P}^q\boldsymbol{r} = \boldsymbol{P}^q(\boldsymbol{y} - \boldsymbol{C}\gamma) = \boldsymbol{P}^q\boldsymbol{y} - \boldsymbol{C}\boldsymbol{P}^q\gamma$ and hence $\boldsymbol{P}^q\gamma = \text{OMP}(\boldsymbol{P}^q\boldsymbol{y})$. ■

Since in the previous section we have described an algorithm to learn a single circulant dictionary, we extend the result for the design of unions of circulants. When dealing with a union of circulant matrices $\boldsymbol{C} = [\boldsymbol{C}_1\ \boldsymbol{C}_2\ \ldots\ \boldsymbol{C}_L]$ we also partition the representations $\boldsymbol{\Gamma} = [\boldsymbol{\Gamma}_1^T\ \boldsymbol{\Gamma}_2^T\ \ldots\ \boldsymbol{\Gamma}_L^T]^T$. To solve this problem we deploy a Block Coordinate Relaxation method that involves updating the whole dictionary by successively updating each circulant dictionary $\boldsymbol{C}_j$ one by one. The sparse reconstruction algorithm used is the OMP.

UC–DLA: Union of Circulants Dictionary Learning Algorithm: Given the dataset $\boldsymbol{Y} \in \mathbb{R}^{n \times N}$, the number of circulant blocks L, the number of iterations K and the target sparsity s, design the dictionary $\boldsymbol{C} \in \mathbb{R}^{n \times nL}$, composed of L circulant blocks, and the representations $\boldsymbol{\Gamma} \in \mathbb{R}^{nL \times N}$, $\|\gamma_i\|_0 \leq s$, $i = 1, \ldots, N$, such that $\epsilon = \|\boldsymbol{Y} - \boldsymbol{C}\boldsymbol{\Gamma}\|_F$ is reduced.

- **Initialization**: Construct the dictionary using L randomly selected items from $\boldsymbol{Y}$ as kernels.
- **Iterations**: For $1, \ldots, K$
  1) Apply OMP to construct $\boldsymbol{\Gamma}$ with target sparsity s.
  2) Update each circulant $\boldsymbol{C}_\ell$ with $\ell = 1, \ldots, L$, using the C–DLA applied with target sparsity $s = 1$ for 1 step (avoiding the sparse approximation) on the dataset:

$$\boldsymbol{E}^{(\ell)} = \left(\boldsymbol{Y} - \sum_{i \neq \ell} \boldsymbol{C}_i\boldsymbol{\Gamma}_i\right)^{(\ell)}, \ell = 1, \ldots, L, \tag{6}$$

where $\boldsymbol{E}^{(\ell)}$ is restricted to the data items that use $\boldsymbol{C}_\ell$. ■

The structure of this algorithm is based on invoking C–DLA in a scheme similar to the one used for the creation of unions of orthonormal dictionaries [14].

IV. Extension to Toeplitz Dictionaries

We consider a generalization of the previous circulant result to Toeplitz matrices $\boldsymbol{T} \in \mathbb{R}^{n \times n}$, defined by the first row and column concatenated in $\boldsymbol{t} \in \mathbb{R}^{2n-1}$. Properties of the circulant matrices can be extended to the Toeplitz case by embedding the matrix $\boldsymbol{T}$ into a larger circulant structure $\bar{\boldsymbol{C}} = \text{circ}(\bar{\boldsymbol{c}}) \in \mathbb{R}^{2n \times 2n}$, $\bar{\boldsymbol{c}}_{n+1} = 0$. Of course, $\boldsymbol{T}$ could be embedded in a $(2n-1) \times (2n-1)$ circulant, but we prefer an even size for the convenience of FFT calculations. The equivalent fast matrix-vector multiplication algorithm from (3) now becomes $\boldsymbol{T}\gamma = \boldsymbol{M}\bar{\boldsymbol{C}}\gamma_{\text{ext}}$, where $\boldsymbol{M} = [\boldsymbol{I}_n\ \boldsymbol{0}_n] \in \mathbb{R}^{n \times 2n}$ is the mask and $\gamma_{\text{ext}} \in \mathbb{R}^{2n}$ is the zero-padded γ vector.

Analogously to the circulant case, this section provides an efficient solution to the convex optimization problem:

$$\underset{T}{\text{minimize}} \, \|\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\Gamma}\|_F, \tag{7}$$

where $\boldsymbol{T}$ is Toeplitz and $\boldsymbol{Y}, \boldsymbol{\Gamma}$ are fixed. Using (2) we get:

$$\begin{aligned} \|\boldsymbol{Y} - \boldsymbol{T}\boldsymbol{\Gamma}\|_F &= \|\boldsymbol{Y} - \boldsymbol{M}\bar{\boldsymbol{C}}\boldsymbol{\Gamma}_{\text{ext}}\|_F = \\ \|\boldsymbol{Y} - \boldsymbol{M}\boldsymbol{F}^H\bar{\Sigma}\boldsymbol{F}\boldsymbol{\Gamma}_{\text{ext}}\|_F &= \|\boldsymbol{Y} - \boldsymbol{F}_{1:n}^H\bar{\Sigma}\tilde{\boldsymbol{\Gamma}}_{\text{ext}}\|_F, \end{aligned} \tag{8}$$

where $\tilde{\boldsymbol{\Gamma}}_{\text{ext}}$ is the Fourier transform of the extended $\boldsymbol{\Gamma}$ matrix and $\boldsymbol{F}_{1:n}^H$ represents the first n rows of $\boldsymbol{F}^H$. Considering:

$$\|\boldsymbol{Y} - \boldsymbol{F}_{1:n}^H\bar{\Sigma}\tilde{\boldsymbol{\Gamma}}_{\text{ext}}\|_F = \|\text{vec}(\boldsymbol{Y}) - \text{vec}(\boldsymbol{F}_{1:n}^H\bar{\Sigma}\tilde{\boldsymbol{\Gamma}}_{\text{ext}})\|_2, \tag{9}$$

and observing that expressing matrix multiplications as a linear transformation on $\bar{\Sigma}$ [15] we reach:

$$\text{vec}(\boldsymbol{F}_{1:n}^H\bar{\Sigma}\tilde{\boldsymbol{\Gamma}}_{\text{ext}}) = (\tilde{\boldsymbol{\Gamma}}_{\text{ext}}^T \otimes \boldsymbol{F}_{1:n}^H)\text{vec}(\bar{\Sigma}). \tag{10}$$

The transformation **vec** denotes the vectorization of an input matrix and $\otimes$ denotes the Kronecker product. Equality in (9) holds since the Frobenius norm is elementwise. Denoting the result of this product $\boldsymbol{K} \in \mathbb{C}^{nN \times n^2}$ and taking $\bar{\Sigma}$'s structure in consideration we obtain:

$$\boldsymbol{K}\text{vec}(\bar{\Sigma}) = \boldsymbol{K}_{\text{r}}\bar{\boldsymbol{\sigma}} = \boldsymbol{K}_{\text{r}}\boldsymbol{F}\bar{\boldsymbol{c}}, \tag{11}$$

where $\bar{\Sigma} = \text{diag}(\bar{\boldsymbol{\sigma}})$ and $\boldsymbol{K}_{\text{r}} \in \mathbb{C}^{nN \times 2n}$ consists of the columns of $\boldsymbol{K}$ corresponding to the diagonal entries of $\text{vec}(\bar{\Sigma})$, given by $\tilde{\gamma}_i \otimes \boldsymbol{f}_i$, $i = 1, \ldots, 2n$, where $\tilde{\gamma}_i^T$ is the ith row of $\tilde{\boldsymbol{\Gamma}}$ and $\boldsymbol{f}_i$ is the ith column of $\boldsymbol{F}_{1:n}^H$–we again have a conjugate symmetric structure $\boldsymbol{K}_{\text{r}} = (\boldsymbol{k}_1 \ \boldsymbol{K}_{2:n} \ \boldsymbol{k}_{n+1} \ \boldsymbol{K}_{2:n}^*)$. Thus, (7) reduces to the unconstrained least squares problem:

$$\underset{\bar{\boldsymbol{c}}}{\text{minimize}} \, \|\text{vec}(\boldsymbol{Y}) - \boldsymbol{K}_{\text{r}}\boldsymbol{F}\bar{\boldsymbol{c}}\|_F. \tag{12}$$

Notice that the pseudo-inverse solution becomes $\bar{\boldsymbol{c}} = \boldsymbol{F}^H(\boldsymbol{K}_{\text{r}}^T\boldsymbol{K}_{\text{r}})^{-1}\boldsymbol{K}_{\text{r}}^T\text{vec}(\boldsymbol{Y})$, remember $\bar{c}_{n+1} = 0$, which permits us to avoid the explicit construction of the large matrix $\boldsymbol{K}_{\text{r}}$ and take full advantage of its structure.

When using Toeplitz matrices as dictionaries, it is important to mention that we do have to tend to the atom norms (unlike with the circulant case). That is why, when updating $\boldsymbol{\Gamma}$ using OMP we use the normalized dictionary $\boldsymbol{T}_{\text{norm}}$ (which is not Toeplitz anymore) and then we return the Toeplitz matrix $\boldsymbol{T}$ and the new representations $\boldsymbol{D}_{\text{norms}}^{-1}\boldsymbol{\Gamma}$ where the diagonal matrix $\boldsymbol{D}_{\text{norms}} \in \mathbb{R}^{n \times n}$ contains the norms of the atoms. Due to this explicit, separate normalization step the n^{-1} initial factor from (2) is omitted in (8).

Based on the general alternate optimization learning procedure and (12) a Toeplitz dictionary learning algorithm (T–DLA), analogous to C–DLA, can be obtained (omitted here for brevity). Naturally, the T–DLA is computationally less efficient than C–DLA but it allows for richer structure.

Remark 2: These ideas can be adapted to the circulant case. Consider $\|\boldsymbol{Y} - \boldsymbol{C}\boldsymbol{\Gamma}\|_F = \|\text{vec}(\boldsymbol{Y}) - (\tilde{\boldsymbol{\Gamma}}^T \otimes \boldsymbol{F}^H)\text{vec}(\Sigma)\|_F$. While the C–DLA deals with the frequency component $\boldsymbol{\sigma}$, this approach offers direct access to the kernel $\boldsymbol{c}$. This formulation accommodates for the problem descriptions in [7] where short atoms are translated across long signals. ■

Remark 3: The T–DLA can also produce overcomplete dictionaries $\boldsymbol{T} \in \mathbb{R}^{n \times m}, m > n$. ■

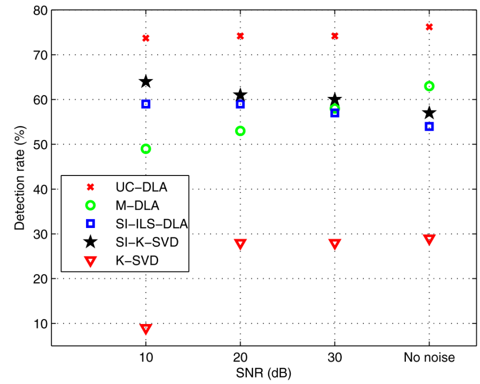

Fig. 1. Average detection rate as a function of noise level for different shift-invariant dictionary learning methods.

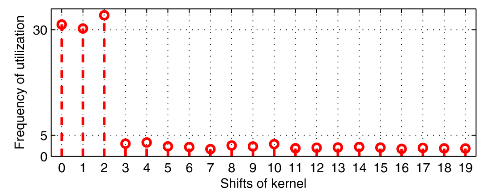

Fig. 2. Average frequency of utilization for all atoms of all the bases. Ideally, the three peeks should be $Ns(Lq)^{-1} \approx 44$.

V. RESULTS

A. Shift-Invariant Learning - Synthetic Example

The goal of UC–DLA is to recover the kernels from an available dataset. To open the possibility for comparison with other shift-invariant methods [9], [10], [11], we create the dataset following the instructions in [9]. This experiment follows the setup: with $n = 20$, $N = 2000$, $s = 3$, randomly produce $L = 45$ kernel columns in $\boldsymbol{R} \in \mathbb{R}^{n \times L}$ allowing for only a maximum of $q = 3$ circular shifts (out of the $n = 20$ possible) and generate the dataset $\boldsymbol{Y} \in \mathbb{R}^{n \times N}$, by taking each $\boldsymbol{y}_j = \sum_{\ell=1}^{L} \alpha_{j\ell}\boldsymbol{P}^{q_{j\ell}}\boldsymbol{r}_\ell, j = 1, \ldots, N$, with $\|\boldsymbol{\alpha}_j\|_0 = s$, $\alpha_{j\ell} \in [-10, 10]$, $q_{j\ell} \in \{0, 1, 2\}$.

We present comparisons of UC–DLA with other shift-invariant methods: SI-K-SVD [9], SI-ILS-DLA [10] and M-DLA [11]. For reference, the K-SVD algorithm is also used to train a dictionary of size 20×135, the $L = 45$ kernels in all $q = 3$ possible shifts. The UC–DLA produces basis $\boldsymbol{C}_\ell \in \mathbb{R}^{n \times n}, \ell = 1, \ldots, L$. To check the detection of a kernel, we take the absolute value dot products between the original kernel and the trained one and check a threshold of 0.99. Above this value the kernel is considered detected and the circulant is attributed to the kernel. The UC–DLA requires approximately 12 seconds with $K = 80$ iterations. The results are averaged over five tests.

Fig. 1 clearly shows that the proposed method substantially outperforms current state-of-the-art algorithms, demonstrating that explicit treatment of shift-invariance has measurable returns (to the tune of 20% with respect to SI-K-SVD in the 30 dB case). Since UC–DLA is not supplied with the correct number of shifts in the dataset ($q = 3$) and it produces the full circulants of $n = 20$ components, Fig. 2 shows on average, over all $L = 45$ circulant bases, how many times each component was selected to take part in any representation – from which an estimation of q can be made. The proposed algorithm correctly identifies the three shifts in the dataset.

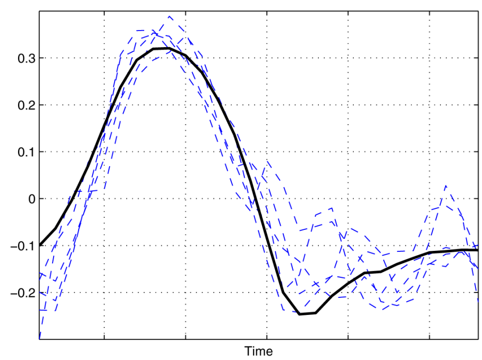

Fig. 3. Example BOLD curves and the circulant kernel.

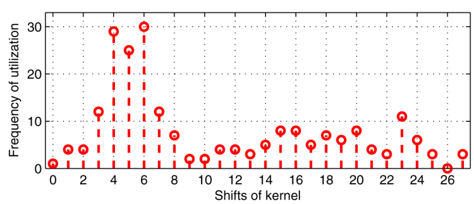

Fig. 4. Frequency of utilization for BOLD time series.

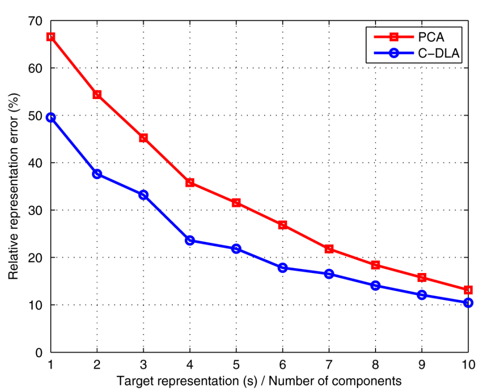

Fig. 5. Representation errors of C–DLA and PCA for the BOLD dataset.

B. Medical Imaging Application

An interesting application arising in the medical imaging arena is the analysis of cardiac phase resolved BOLD MRI time series [16]. These short-length time series are extracted from imaging data of the myocardium and show the evolution of oxygenation in the myocardium as a function of cardiac phase. They typically appear to have a maximum around 1/3 of the time series (coinciding with the location of the end-systole) fading out towards their end (location of end-diastole). Several of such time series can be extracted from several locations of the myocardium in a single MRI study, with the interesting characteristic that several shifts exist between the time series due to known physiological effects [16]. Thus, we are interested to see the capabilities of shift invariant learning, i.e. the C–DLA, in identifying this characteristic curve and its possible shifts that may appear due to physiology.

The available dataset consist of time series extracted from imaging data belonging to 10 subjects. The time series have the mean subtracted and they are normed, interpolated and concatenated columnwise in $\boldsymbol{Y} \in \mathbb{R}^{28\times 240}$–24 per subject, each interpolated to 28 samples. In Fig. 3, the dotted thin lines show some example time series extracted while the solid line denotes the circulant kernel computed using C–DLA. Fig. 4 shows the frequency of utilization of the computed kernel and its shifts with a distinctive dominant group for shifts $\{3,4,5,6,7\}$, thus showing a focused utilization of a few particular shifts. Fig. 5 shows the relative representation errors ($\|\boldsymbol{Y}-\boldsymbol{C}\mathbf{\Gamma}\|_F/\|\boldsymbol{Y}\|_F$) of C–DLA, on the horizontal axis having the sparsity s, as compared to the Principal Component representations (PCA) using s components. Notice that using only one component in each sparse representation with the C–DLA we reach approximately 50% representation error. Even as s increases, C–DLA maintains a performance advance compared to PCA indicating the strong shift structure in the dataset. This type of approach allows for the analysis of the BOLD MRI time series and reaching an interpretable result.

VI. Conclusions

In this letter we present new learning algorithms for the construction of structured dictionaries for sparse representations, namely: circulant and Toeplitz. When learning structured dictionaries, the algorithms presented are fast and have a close form solution, reducing to Fourier transforms, matrix-vector operations and/or least squares problems. Also, we extend the results and produce algorithms able to learn shift-invariant dictionaries. Comparisons with state-of-the-art algorithms show the effectiveness of the proposed methods.

References

[1] A. M. Bruckstein, D. L. Donoho, and M. Elad, "From sparse solutions of systems of equations to sparse modeling of signals and images," *SIAM Rev.*, vol. 51, no. 1, pp. 34–81, 2009.
[2] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, pp. 2231–2242, 2004.
[3] K. Engan, K. Skretting, and J. Hakon-Husoy, "Family of iterative LS-based dictionary learning algorithms, ILS-DLA, for sparse signal representation," *Dig. Signal Process.*, vol. 17, no. 1, pp. 32–49, 2007.
[4] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, 2006.
[5] B. Mailhe, S. Lesage, R. Gribonval, F. Bimbot, and P. Vandergheynst, "Shift-invariant dictionary learning for sparse representations: Extending K-SVD," *EUSIPCO*, 2008.
[6] J. J. Thiagarajan, K. N. Ramamurthy, and A. Spania, "Shift-invariant sparse representation of images using learned dictionaries," *IEEE MLSP*, pp. 145–150, 2008.
[7] P. Jost, P. Vandergheynst, S. Lesage, and R. Gribonval, "MoTIF: An efficient algorithm for learning translation invariant dictionaries," *IEEE ICASSP*, pp. 857–860, 2006.
[8] M. Nakashizuka, H. Nishiura, and Y. Iiguni, "Sparse image representations with shift-invariant tree-structured dictionaries," *IEEE ICIP*, pp. 2145–2148, 2009.
[9] M. Aharon, "Overcomplete dictionaries for sparse representation of signals," Ph.D. dissertation, Technion–Israel Inst. Technol., Haifa, Israel, 2006.
[10] K. Skretting, J. Husoy, and S. Aase, "General design algorithm for sparse frame expansions," *Signal Process.*, vol. 86, pp. 117–126, 2006.
[11] Q. Barthelemy, A. Larue, A. Mayoue, D. Mercier, and J. I. Mars, "Shift & 2D rotation invariant sparse coding for multivariate signals," *IEEE Tran. Signal Process.*, vol. 60, no. 4, pp. 1584–1611, 2012.
[12] V. K. Potluru, S. M. Plis, and V. D. Calhoun, "Sparse shift-invariant NMF," *SSIAI*, pp. 69–72, 2008.
[13] R. Rubinstein, M. Zibulevsky, and M. Elad, Efficient Implementation of the K-SVD Algorithm using Batch Orthogonal Matching Pursuit - Tech. Rep. 2008.
[14] S. Lesage, R. Gribonval, F. Bimbot, and L. Benaroya, "Learning unions of orthonormal bases with thresholded singular value decomposition," *IEEE ICASSP*, vol. 5, pp. 293–296, 2005.
[15] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
[16] S. A. Tsaftaris, X. Zhou, R. Tang, D. Li, and R. Dharmakumar, "Detecting myocardial ischemia at rest with cardiac phase-resolved blood oxygen level-dependent cardiovascular magnetic resonance, circ," *Cardiovasc. Imag.*, vol. 6, no. 2, pp. 311–319, 2013.