Signal Recovery in Compressive Sensing via Multiple Sparsifying Bases

U. L. Wijewardhana[*], E. Belyaev[†], M. Codreanu[*], and M. Latva-aho[*]

[*]Centre for Wireless Communications
University of Oulu
Oulu, 90570, Finland
firstname.lastname@oulu.fi

[†]Department of Photonics Engineering
Technical University of Denmark
Lyngby, 2800 Kgs, Denmark
evbel@fotonik.dtu.dk

Abstract

Compressive sensing theory asserts that, under certain conditions, a high dimensional but compressible signal can be recovered from a small number of random linear projections by utilizing computationally efficient algorithms. The *a priori* knowledge of the basis in which the signal of interest is sparse is the key assumption utilized by such algorithms. However, the basis in which the signal is the sparsest is unknown for many natural signals of interest. Instead there may exist multiple bases which lead to a compressible representation of the signal: e.g., an image is compressible in different wavelet transforms. We show that a significant performance improvement can be achieved by utilizing multiple estimates of the signal using sparsifying bases in the context of signal reconstruction from compressive samples. Further, we derive a *customized interior-point method* to jointly obtain multiple estimates of a 2-D signal (image) from compressive measurements utilizing multiple sparsifying bases as well as the fact that the images usually have a sparse gradient.

1 Introduction

According to compressive sensing (CS) framework [1,2], under certain conditions, a high dimensional signal can be reconstructed from a small number of random linear projections by utilizing computationally efficient recovery algorithms. Consider the reconstruction of a 2-D signal $\mathbf{X} \in \mathbb{R}^{N \times N}$ from compressive measurements. The signal of interest $\mathbf{X}$ is acquired utilizing the linear measurement model given by

$$\mathbf{y} = \mathbf{\Phi}\mathbf{x}, \tag{1}$$

where $\mathbf{x} \in \mathbb{R}^{N^2}$ stacks columns of $\mathbf{X}$ into a vector, i.e., $\mathbf{x} = \text{vec}(\mathbf{X})$; $\mathbf{y} \in \mathbb{R}^M$ contains $M < N^2$ linear measurements of $\mathbf{x}$ and $\mathbf{\Phi} \in \mathbb{R}^{M \times N^2}$ denotes the measurement matrix.

In conventional CS recovery algorithms it is assumed that the signal of interest $\mathbf{x}$ has a sparse representation in a known basis [2–4], i.e., $\mathbf{x} = \mathbf{\Psi}\mathbf{w}$, where $\mathbf{\Psi}$ is the $N^2 \times N^2$ representation matrix and $\mathbf{w}$ is the sparse vector containing the transform domain coefficients. Under this basic assumption, the problem of recovering the signal $\mathbf{x}$ from the measurements $\mathbf{y}$ can be formulated as searching for the sparsest vector $\mathbf{w}$ which satisfies $\mathbf{y} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{w}$. Instead of solving this NP-hard combinatorial
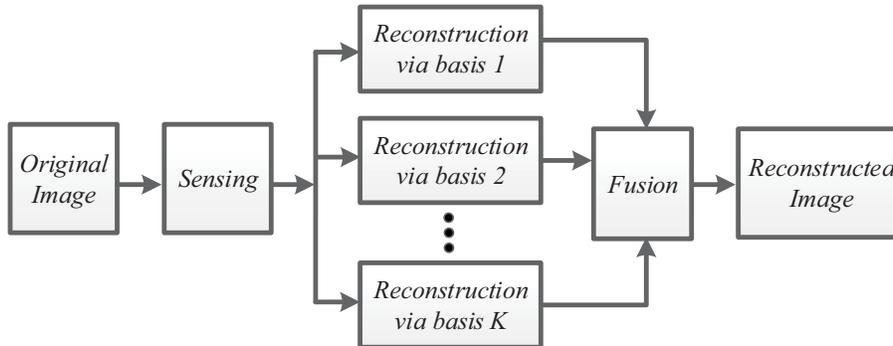
Figure 1: Signal reconstruction from compressive measurements utilizing multiple sparsifying bases.

problem of discovering the sparsest $\mathbf{w}$, most of the existing CS recovery algorithms solve its best convex approximation, the l_1–norm minimization problem [2,5]

$$\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \quad \|\mathbf{w}\|_1 \quad \text{subject to} \quad \mathbf{y} = \mathbf{\Phi}\mathbf{\Psi}\mathbf{w}, \tag{2}$$

with optimization variable $\mathbf{w}$, where $\| \cdot \|_1$ denotes the l_1–norm. Then the signal of interest is estimated as $\hat{\mathbf{x}} = \mathbf{\Psi}\hat{\mathbf{w}}$.

For many natural signals of interest, e.g., images, video sequences, the basis in which the signal is the sparsest is unknown. Instead, usually there exist multiple bases which lead to a compressible representation of the signal, i.e., the energy of the signal is concentrated in a relatively small set of transform domain coefficients [3,6]. Let us refer to them as *sparsifying bases*. In traditional CS recovery algorithms, the basis $\mathbf{\Psi}$ in (2) is chosen from a set of possible sparsifying bases, e.g., discrete wavelet transform (DWT), discrete cosine transform (DCT). Naturally, the quality of reconstruction of the signal of interest varies depending on the choice of the basis $\mathbf{\Psi}$.

So far, most of the effort have been focused on finding a specific basis that provides the sparsest representation of the signal, e.g., via dictionary learning algorithms [7,8], which are generally computationally expensive. However, as pointed out in [9] in a signal denoising application, significant performance improvement can be obtained by combining several medium-sparse representations of the signal as compared to the case in which only the sparsest representation is utilized. In this paper, we show that a similar performance improvement can be also achieved in the context of signal reconstruction from compressive samples by utilizing multiple sparsifying bases. Further, we derive a *customized interior-point method* for recovering 2-D signals (images) from compressive measurements which can utilize multiple sparsifying bases (e.g., wavelet and discrete cosine transforms) as well as the fact that the images usually have a sparse gradient. It is important to point out that, the signal reconstruction method proposed in this paper differs from the signal denoising method introduced in [9] in a fundamental way. Though both methods rely on combining multiple sparse signal representations, the method introduced in [9] uses a randomization technique to generate multiple sparse representations in a *single over-complete dictionary* while our method uses *several distinct sparsifying bases* to obtain multiple sparse representations of the signal of interest.

2   Signal reconstruction via multiple bases

Let us suppose that the signal of interest $\mathbf{x}$ is reasonably compressible in a set of K a-priori known bases $\mathcal{S} = \{\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \ldots, \boldsymbol{\Psi}_K\}$. Then, we perform the signal reconstruction utilizing the idea illustrated in Figure 1. First, we estimate the set of sparse coefficient vectors $\mathcal{W} = \{\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2, \ldots, \hat{\mathbf{w}}_K\}$ from the measurements $\mathbf{y}$ by assuming $\mathbf{x}$ is sparse in each basis $\boldsymbol{\Psi}_k \in \mathcal{S}$. Then the signal $\mathbf{x}$ is reconstructed as $\tilde{\mathbf{x}} = f(\mathcal{W}, \mathcal{S})$, where $f(.)$ is the *fusion function*. We can summarize the key idea as follows:

Algorithm 1 *Signal recovery via multiple sparsifying bases*

1. **Inputs: y**, $\boldsymbol{\Phi}$, $\mathcal{S} = \{\boldsymbol{\Psi}_1, \boldsymbol{\Psi}_2, \ldots, \boldsymbol{\Psi}_K\}$.
2. **Estimate:** $\mathcal{W} = \{\hat{\mathbf{w}}_1, \hat{\mathbf{w}}_2, \ldots, \hat{\mathbf{w}}_K\}$.
3. **Output:** $\tilde{\mathbf{x}} = f(\mathcal{W}, \mathcal{S})$.

In step 2, $\hat{\mathbf{w}}_k \in \mathcal{W}$ can be estimated from measurements $\mathbf{y}$ by solving the l_1–norm minimization problem (2) for each basis $\boldsymbol{\Psi}_k \in \mathcal{S}$. Alternatively, in the sequel we propose a novel method which can utilize multiple sparsifying bases as well as the fact that the images usually have a sparse gradient in order to estimate $\hat{\mathbf{w}}_k \in \mathcal{W}$. In step 3, as an example we can utilize simple averaging as the fusion function $f(\mathcal{W}, \mathcal{S})$ to reconstruct the signal, i.e., $\tilde{\mathbf{x}} = 1/K \sum_{k=1}^{K} \boldsymbol{\Psi}_k \hat{\mathbf{w}}_k$. However, utilization of a more sophisticated fusion function, e.g., by taking into account the sparsity of each $\hat{\mathbf{w}}_k \in \mathcal{W}$, may provide a better recovery performance. Finding the optimal fusion function would be an interesting research area, which is out of the scope of this paper. However, the numerical results show that even a simple averaging of several estimates of the signal significantly improve the reconstruction performance compared to any single estimate.

In general, we need to estimate $\hat{\mathbf{w}}_k \in \mathcal{W}$ in step 2 such that it provides the best sparse representation of the signal $\mathbf{x}$ in $\boldsymbol{\Psi}_k \in \mathcal{S}$. In order to jointly estimate $\hat{\mathbf{w}}_k \in \mathcal{W}$ for the case where the signal of interest is a natural image, we consider solving the following optimization problem:

$$
\begin{aligned}
\text{minimize} \quad & \sum_{k=1}^{K} \lambda_k \|\mathbf{w}_k\|_1 + \text{TV}\left(\sum_{k=1}^{K} \mathbf{X}_k\right) \\
\text{subject to} \quad & \|\mathbf{y} - \boldsymbol{\Phi}\boldsymbol{\Psi}_k\mathbf{w}_k\|_2 \le \delta_k, \quad k = 1, \ldots, K
\end{aligned}
\tag{3}
$$

with the optimization variable $\mathbf{w}_k$ for $k = 1, \ldots, K$. Specifically, we derive a customized interior-point method for solving the problem (3). Here the p–norm of a vector is denoted by $\|\cdot\|_p$, $\text{vec}(\mathbf{X}_k) = \boldsymbol{\Psi}_k\mathbf{w}_k$, and $\text{TV}(\cdot)$ denotes the *total variation* of an image. Problem (3) not only exploits the sparsity of the image in given transforms but also consider that the image has a sparse gradient. The parameter δ_k allows to account for the case of noisy acquisition of the signal. Further, by utilizing the ratio between the regularization parameters λ_k for $k = 1, \ldots, K$, we can encode the belief on sparsity of the signal in bases $\boldsymbol{\Psi}_k \in \mathcal{S}$.

3 Customized interior-point method for CS recovery via multiple bases

Customized methods have been proposed to solve the l_1-norm minimization problem in [10, 11] and TV minimization problem in [10]. However, these methods are not applicable to problem (3) due to the variations in the structure of the Hessian. Thus, in this section, we derive a customized interior-point method for solving problem (3) that compute the search direction of the Newton method efficiently by exploiting the specific structure of the Hessian. It is worth to mention that the implementation of the proposed interior-point method in [12] is inspired from [10], and it can reconstruct medium size images.

3.1 Interior-point method

Let $x_{i,j}$ denotes the pixel in ith row and jth column of an image $\mathbf{X}$. We define the vectors $\mathbf{d}_{ij}^h \in \mathbb{R}^{N^2}$ and $\mathbf{d}_{ij}^v \in \mathbb{R}^{N^2}$ such that

$$(\mathbf{d}_{ij}^h)^{\mathrm{T}}\mathbf{x} = \begin{cases} x_{i,j+1} - x_{i,j} & j < N \\ 0 & j = N \end{cases} \qquad (\mathbf{d}_{ij}^v)^{\mathrm{T}}\mathbf{x} = \begin{cases} x_{i+1,j} - x_{i,j} & i < N \\ 0 & i = N. \end{cases}$$

Then, the total variation of $\mathbf{X}$ can be expressed as $\mathrm{TV}(\mathbf{X}) = \sum_{i=1}^{N}\sum_{j=1}^{N}\|\mathbf{D}_{ij}\mathbf{x}\|_2$, where $\mathbf{D}_{ij} = \begin{bmatrix} \mathbf{d}_{ij}^h & \mathbf{d}_{ij}^v \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{2 \times N^2}$ and $\mathbf{x} = \mathrm{vec}(\mathbf{X})$. Now, by introducing the auxiliary variables $u_{k,n}$ for $k = 1, \ldots, K$, $n = 1, \ldots, N^2$ and v_{ij} for $i, j = 1, \ldots, N$, we equivalently reformulate the problem (3) as

$$
\begin{aligned}
\text{minimize} \quad & \sum_{k=1}^{K}\sum_{n=1}^{N^2}\lambda_k u_{k,n} + \sum_{i=1}^{N}\sum_{j=1}^{N} v_{ij} \\
\text{subject to} \quad & \|\mathbf{y} - \mathbf{A}_k\mathbf{w}_k\|_2 \le \delta_k, & k = 1, \ldots, K \\
& \|\textstyle\sum_{k=1}^{K}\mathbf{D}_{ij}\mathbf{\Psi}_k\mathbf{w}_k\|_2 \le v_{ij}, & i, j = 1, \ldots, N \\
& -u_{k,n} \le w_{k,n} \le u_{k,n}, & k = 1, \ldots, K \quad n = 1, \ldots, N^2
\end{aligned}
\tag{4}
$$

where $\mathbf{A}_k = \mathbf{\Phi}\mathbf{\Psi}_k$ with the optimization variables $\mathbf{w}_k \in \mathbb{R}^{N^2}$, $\mathbf{u}_k = [u_{k,1}, \ldots, u_{k,N^2}]^{\mathrm{T}} \in \mathbb{R}^{N^2}$, and $\mathbf{v} = [v_{11}, \ldots, v_{N1}, \ldots, v_{NN}]^{\mathrm{T}} \in \mathbb{R}^{N^2}$. Next, we introduce the *logarithmic barrier* function for the inequality constraints in problem (4) as

$$
\begin{aligned}
\phi(\mathbf{w}, \mathbf{v}, \mathbf{u}) = & -\sum_{k=1}^{K}\log(\delta_k^2 - \|\mathbf{r}_k\|_2^2) - \sum_{i=1}^{N}\sum_{j=1}^{N}\log(v_{ij}^2 - \|\mathbf{B}_{ij}\mathbf{w}\|_2^2) \\
& - \sum_{k=1}^{K}\sum_{n=1}^{N^2}\Big(\log(u_{k,n} + w_{k,n}) + \log(u_{k,n} - w_{k,n})\Big)
\end{aligned}
\tag{5}
$$

where $\mathbf{w} = [\mathbf{w}_1^{\mathrm{T}} \ \ldots \ \mathbf{w}_K^{\mathrm{T}}]^{\mathrm{T}}$, $\mathbf{u} = [\mathbf{u}_1^{\mathrm{T}} \ \ldots \ \mathbf{u}_K^{\mathrm{T}}]^{\mathrm{T}}$, $\mathbf{r}_k = \mathbf{y} - \mathbf{A}_k\mathbf{w}_k$, and $\mathbf{B}_{ij} = \mathbf{D}_{ij}\mathbf{\Psi}$ with $\mathbf{\Psi} = [\mathbf{\Psi}_1, \ldots, \mathbf{\Psi}_K]$. The log barrier function (5) is defined over $\mathbf{dom}\,\phi = \{(\mathbf{w}, \mathbf{v}, \mathbf{u}) \in \mathbb{R}^{KN^2} \times \mathbb{R}^{N^2} \times \mathbb{R}^{KN^2} | \|\mathbf{r}_k\|_2 < \delta_k, \|\mathbf{B}_{ij}\mathbf{w}\|_2 < v_{ij}, |w_{k,n}| < u_{k,n}, \text{ for } k = 1, \ldots, K, n = 1, \ldots, N^2 \text{ and } i, j = 1, \ldots, N\}$.

For the notational convenience, we define $\mathbf{q} = [\mathbf{w}^{\mathrm{T}} \ \mathbf{v}^{\mathrm{T}} \ \mathbf{u}^{\mathrm{T}}]^{\mathrm{T}}$. Then, for $t > 0$, let $\mathbf{q}_t^\star$ minimize the convex function

$$\psi_t(\mathbf{q}) = t\Big(\sum_{k=1}^{K}\sum_{n=1}^{N^2}\lambda_k u_{k,n} + \sum_{i=1}^{N}\sum_{j=1}^{N} v_{ij}\Big) + \phi(\mathbf{q}). \tag{6}$$

The *central path* associated with problem (4) is defined as the set of *central points* $\mathbf{q}_t^\star$ for $t > 0$. Specifically, the central path leads to an optimal solution of problem (4) as $t \to \infty$, since $\mathbf{q}_t^\star$ is no more than $((2K+1)N^2 + K)/t$–suboptimal [13, Sec. 11.2].

In an interior-point method, we compute a set of central points $\mathbf{q}_t^\star$ for a sequence of increasing values of t until $((2K+1)N^2+K)/t \leq \epsilon$ in order to obtain an ϵ–suboptimal solution for problem (4) [13, Sec. 11.3]. Thus, we can summarize the interior-point method as follows:

Algorithm 2 *Interior-point method for l_1–regularized TV with multiple bases*

    given $t^0 > 0$, $\mu > 1$, $\epsilon > 0$ and strictly feasible $\mathbf{q}^0$.

    initialize $\hat{\mathbf{q}} = \mathbf{q}^0$, $t = t^0$

    repeat

  1. Compute $\mathbf{q}_t^\star$ by minimizing $\psi_t(\mathbf{q})$, starting at $\hat{\mathbf{q}}$.

  2. Update. $\hat{\mathbf{q}} := \mathbf{q}_t^\star$.

  3. **quit** if $((2K+1)N^2 + K)/t < \epsilon$ and output $\hat{\mathbf{q}}$.

  4. Increase t. $t := \mu t$.

At each iteration (except the first one) we compute the central point $\mathbf{q}_t^\star$ starting from the previously computed central point. For this purpose, in step 1, we utilize the Newton's method [13, Sec. 9.5] in order to minimize $\psi_t(\mathbf{q})$, for a fix value of t.

3.2 Efficient solution of the Newton system

In Newton's method, the search direction $\Delta\mathbf{q} = [\Delta\mathbf{w}^\mathrm{T} \ \Delta\mathbf{v}^\mathrm{T} \ \Delta\mathbf{u}^\mathrm{T}]^\mathrm{T}$ is computed as the solution to the Newton system [13, Sec. 9.7]

$$\mathbf{H}\Delta\mathbf{q} = -\mathbf{g}, \tag{7}$$

where $\mathbf{H} = \nabla^2\psi_t(\mathbf{w}, \mathbf{v}, \mathbf{u}) \in \mathbb{R}^{(2K+1)N^2 \times (2K+1)N^2}$ and $\mathbf{g} = \nabla\psi_t(\mathbf{w}, \mathbf{v}, \mathbf{u}) \in \mathbb{R}^{(2K+1)N^2}$ denotes the Hessian and the gradient of ψ_t at $(\mathbf{w}, \mathbf{v}, \mathbf{u})$ respectively. The principle behind a customized interior-point method is *the efficient solution of the Newton system by exploiting the structure of the Hessian*. In this subsection, we present an efficient method for solving the Newton system (7), which is the key to the effectiveness of the customized interior-point method. The Newton's method utilize to compute $\mathbf{q}_t^\star$ in step 1 of *Algorithm 2* can be summarized as follows:

Algorithm 3 *Newton's method*

    given $\epsilon_{\mathrm{nt}} > 0$ and staring point $\mathbf{q}^0 \in \mathbf{dom}\ \psi_t(\mathbf{q})$.

    initialize $\hat{\mathbf{q}} = \mathbf{q}^0$.

    repeat

  1. Compute the search direction $\Delta\mathbf{q}$ by solving Newton system (7) and the decrement as $\eta^2 = -\nabla\psi_t(\hat{\mathbf{q}})^\mathrm{T}\Delta\mathbf{q}$.

  2. **quit** if $\eta^2/2 \leq \epsilon_{\mathrm{nt}}$ and output $\hat{\mathbf{q}}$.

  3. Choose step size s by backtracking line search.

  4. Update. $\hat{\mathbf{q}} := \hat{\mathbf{q}} + s\Delta\mathbf{q}$

Given the search direction $\Delta\mathbf{q}$, the step size in the backtracking line search is taken as $s = \beta^p$, where $p \geq 0$ is the smallest integer that satisfies

$$\psi_t(\mathbf{q} + \beta^p\Delta\mathbf{q}) \leq \psi_t(\mathbf{q}) + \alpha\beta^p\nabla\psi_t(\mathbf{q})\Delta\mathbf{q},$$

where $\nabla\psi_t(\mathbf{q})$ is the gradient of ψ_t at $\mathbf{q}$, $\alpha \in (0, 0.5)$ and $\beta \in (0, 1)$ [13, Sec. 9.2].

To derive an efficient solution to the Newton system (7), first we obtain compact representations of the Hessian and gradient. The Hessian can be expressed as

$$\mathbf{H} = \begin{bmatrix} \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{D}_1 & \mathbf{F}_3 & \mathbf{D}_2 \\ \mathbf{F}_3^{\mathrm{T}} & \mathbf{D}_3 & \mathbf{0} \\ \mathbf{D}_2 & \mathbf{0} & \mathbf{D}_1 \end{bmatrix} \in \mathbb{R}^{(2K+1)N^2 \times (2K+1)N^2}, \tag{8}$$

where $\mathbf{F}_1 = \mathrm{diag}(\mathbf{F}_{1,1}, \ldots, \mathbf{F}_{K,1})$, $\mathbf{D}_1 = \mathrm{diag}(\mathbf{D}_{1,1}, \ldots, \mathbf{D}_{K,1})$, and $\mathbf{D}_2 = \mathrm{diag}(\mathbf{D}_{1,2}, \ldots, \mathbf{D}_{K,2})$ with

$$\mathbf{F}_{k,1} = \frac{2\mathbf{A}_k^{\mathrm{T}}\mathbf{A}_k}{\delta_k^2 - \|\mathbf{r}_k\|_2^2} + \frac{4\mathbf{A}_k^{\mathrm{T}}\mathbf{r}_k\mathbf{r}_k^{\mathrm{T}}\mathbf{A}_k}{(\delta_k^2 - \|\mathbf{r}_k\|_2^2)^2} \in \mathbb{R}^{N^2 \times N^2}$$

$$\mathbf{F}_2 = \sum_{i=1}^{N}\sum_{j=1}^{N}\left[\frac{2\mathbf{B}_{ij}^{\mathrm{T}}\mathbf{B}_{ij}}{(v_{ij}^2 - \|\mathbf{B}_{ij}\mathbf{w}\|_2^2)} + \frac{4\mathbf{B}_{ij}^{\mathrm{T}}(\mathbf{B}_{ij}\mathbf{w})(\mathbf{B}_{ij}\mathbf{w})^{\mathrm{T}}\mathbf{B}_{ij}}{(v_{ij}^2 - \|\mathbf{B}_{ij}\mathbf{w}\|_2^2)^2}\right] \in \mathbb{R}^{KN^2 \times KN^2}$$

$$\mathbf{F}_3 = \begin{bmatrix} \dfrac{-4v_{11}\mathbf{B}_{11}^{\mathrm{T}}\mathbf{B}_{11}\mathbf{w}}{(v_{11}^2 - \|\mathbf{B}_{11}\mathbf{w}\|_2^2)^2} & \cdots & \dfrac{-4v_{NN}\mathbf{B}_{NN}^{\mathrm{T}}\mathbf{B}_{NN}\mathbf{w}}{(v_{NN}^2 - \|\mathbf{B}_{NN}\mathbf{w}\|_2^2)^2} \end{bmatrix} \in \mathbb{R}^{KN^2 \times N^2}$$

$$\mathbf{D}_{k,1} = \mathrm{diag}\left(\frac{2(u_{k,1}^2 + w_{k,1}^2)}{(u_{k,1}^2 - w_{k,1}^2)^2}, \ldots, \frac{2(u_{k,N^2}^2 + w_{k,N^2}^2)}{(u_{k,N^2}^2 - w_{k,N^2}^2)^2}\right) \in \mathbb{R}^{N^2 \times N^2},$$

$$\mathbf{D}_{k,2} = \mathrm{diag}\left(\frac{-4u_{k,1}w_{k,1}}{(u_{k,1}^2 - w_{k,1}^2)^2}, \ldots, \frac{-4u_{k,N^2}w_{k,N^2}}{(u_{k,N^2}^2 - w_{k,N^2}^2)^2}\right) \in \mathbb{R}^{N^2 \times N^2},$$

$$\mathbf{D}_3 = \mathrm{diag}\left(\frac{2(v_{11}^2 + \|\mathbf{B}_{11}\mathbf{w}\|_2^2)}{(v_{11}^2 - \|\mathbf{B}_{11}\mathbf{w}\|_2^2)^2}, \ldots, \frac{2(v_{NN}^2 + \|\mathbf{B}_{NN}\mathbf{w}\|_2^2)}{(v_{NN}^2 - \|\mathbf{B}_{NN}\mathbf{w}\|_2^2)^2}\right) \in \mathbb{R}^{N^2 \times N^2}.$$

Here, we denote the diagonal matrix with diagonal blocks $x_1, \ldots, x_p$ by $\mathrm{diag}(x_1, \ldots, x_p)$. The gradient can be written as $\mathbf{g} = \begin{bmatrix} \mathbf{g}_1^{\mathrm{T}} & \mathbf{g}_2^{\mathrm{T}} & \mathbf{g}_3^{\mathrm{T}} \end{bmatrix}^{\mathrm{T}} \in \mathbb{R}^{(2K+1)N^2}$, where

$$\mathbf{g}_1 = \begin{bmatrix} \dfrac{-2\mathbf{A}_1^{\mathrm{T}}\mathbf{r}_1}{\delta_1^2 - \|\mathbf{r}_1\|_2^2} + \begin{bmatrix} 2w_{1,1}/(u_{1,1}^2 - w_{1,1}^2) \\ \vdots \\ 2w_{1,N^2}/(u_{1,N^2}^2 - w_{1,N^2}^2) \end{bmatrix} \\ \vdots \\ \dfrac{-2\mathbf{A}_K^{\mathrm{T}}\mathbf{r}_K}{\delta_K^2 - \|\mathbf{r}_K\|_2^2} + \begin{bmatrix} 2w_{K,1}/(u_{K,1}^2 - w_{K,1}^2) \\ \vdots \\ 2w_{K,N^2}/(u_{K,N^2}^2 - w_{K,N^2}^2) \end{bmatrix} \end{bmatrix} + \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{2\mathbf{B}_{ij}^{\mathrm{T}}\mathbf{B}_{ij}\mathbf{w}}{(v_{ij}^2 - \|\mathbf{B}_{ij}\mathbf{w}\|_2^2)} \in \mathbb{R}^{KN^2}$$

$$\mathbf{g}_2 = t\mathbf{1} - \begin{bmatrix} 2v_{11}/(v_{11}^2 - \|\mathbf{B}_{11}\mathbf{w}\|_2^2) \\ \vdots \\ 2v_{NN}/(v_{NN}^2 - \|\mathbf{B}_{NN}\mathbf{w}\|_2^2) \end{bmatrix} \in \mathbb{R}^{N^2}$$

$$\mathbf{g}_3 = \begin{bmatrix} \lambda_1 t\mathbf{1} - \begin{bmatrix} 2u_{1,1}/(u_{1,1}^2 - w_{1,1}^2) \\ \vdots \\ 2u_{1,N^2}/(u_{1,N^2}^2 - w_{1,N^2}^2) \end{bmatrix} \\ \vdots \\ \lambda_K t\mathbf{1} - \begin{bmatrix} 2u_{K,1}/(u_{K,1}^2 - w_{K,1}^2) \\ \vdots \\ 2u_{K,N^2}/(u_{K,N^2}^2 - w_{K,N^2}^2) \end{bmatrix} \end{bmatrix} \in \mathbb{R}^{KN^2}.$$

Next, in order to obtain the search direction $\Delta\mathbf{q}$ we solve the Newton system (7) by exploiting the structure of the Hessian $\mathbf{H}$ in (8). In particular, we compute $\Delta\mathbf{w}, \Delta\mathbf{v}$ and $\Delta\mathbf{u}$ efficiently by employing the block elimination procedure [13, App. C.4] instead of using directly the inverse of $\mathbf{H}$, which involves an operation of $\mathcal{O}(P^3)$ flops and memory of $\mathcal{O}(P^2)$ with $P = (2K+1)N^2$. For this purpose, first we express $\Delta\mathbf{w}$ by using the block elimination procedure as $\Delta\mathbf{w} = \mathbf{S}^{-1}\mathbf{z}_1$, where

$$\mathbf{S} = \mathbf{F}_1 + \mathbf{F}_2 + \mathbf{D}_1 - \begin{bmatrix} \mathbf{F}_3 & \mathbf{D}_2 \end{bmatrix} \begin{bmatrix} \mathbf{D}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{F}_3^{\mathrm{T}} \\ \mathbf{D}_2 \end{bmatrix}$$

$$= \mathbf{F}_1 + \mathbf{D}_1 - \mathbf{D}_2\mathbf{D}_1^{-1}\mathbf{D}_2 + \mathbf{F}_2 - \mathbf{F}_3\mathbf{D}_3^{-1}\mathbf{F}_3^{\mathrm{T}} \in \mathbb{R}^{KN^2 \times KN^2} \tag{9}$$

$$\mathbf{z}_1 = -\mathbf{g}_1 + \begin{bmatrix} \mathbf{F}_3 & \mathbf{D}_2 \end{bmatrix} \begin{bmatrix} \mathbf{D}_3 & \mathbf{0} \\ \mathbf{0} & \mathbf{D}_1 \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{g}_2 \\ \mathbf{g}_3 \end{bmatrix}$$

$$= -\mathbf{g}_1 + \mathbf{F}_3\mathbf{D}_3^{-1}\mathbf{g}_2 + \mathbf{D}_2\mathbf{D}_1^{-1}\mathbf{g}_3 \in \mathbb{R}^{KN^2}. \tag{10}$$

Let $\mathbf{S}_1 = \mathbf{F}_1 + \mathbf{D}_1 - \mathbf{D}_2\mathbf{D}_1^{-1}\mathbf{D}_2$, $\mathbf{B}_h$ be a matrix containing $(\mathbf{d}_{ij}^h)^{\mathrm{T}}\mathbf{\Psi}$ as rows and $\mathbf{B}_v$ contains $(\mathbf{d}_{ij}^h)^{\mathrm{T}}\mathbf{\Psi}$ as rows. Then we can express $\mathbf{S}$ as $\mathbf{S} = \mathbf{S}_1 + \mathbf{B}^{\mathrm{T}}\mathbf{C}\mathbf{B}$, where

$$\mathbf{B} = \begin{bmatrix} \mathbf{B}_h \\ \mathbf{B}_v \end{bmatrix} \in \mathbb{R}^{2N^2 \times KN^2} \quad \mathbf{C} = \begin{bmatrix} \mathbf{F}_v^{-1} + \mathbf{\Sigma}_v\mathbf{\Sigma}_{\partial h}^2 & \mathbf{\Sigma}_v\mathbf{\Sigma}_{\partial h}\mathbf{\Sigma}_{\partial v} \\ \mathbf{\Sigma}_v\mathbf{\Sigma}_{\partial v}\mathbf{\Sigma}_{\partial h} & \mathbf{F}_v^{-1} + \mathbf{\Sigma}_v\mathbf{\Sigma}_{\partial v}^2 \end{bmatrix} \in \mathbb{R}^{2N^2 \times 2N^2}$$

with $\mathbf{F}_v = \mathrm{diag}\big((v_{11}^2 - \|\mathbf{B}_{11}\mathbf{w}\|_2^2)/2, \ldots, (v_{NN}^2 - \|\mathbf{B}_{NN}\mathbf{w}\|_2^2)/2\big)$, $\mathbf{\Sigma}_v = \mathrm{diag}\big(-4/(v_{11}^4 - \|\mathbf{B}_{11}\mathbf{w}\|_2^4), \ldots, -4/(v_{NN}^4 - \|\mathbf{B}_{NN}\mathbf{w}\|_2^4)\big)$, $\mathbf{\Sigma}_{\partial h} = \mathrm{diag}(\mathbf{B}_h\mathbf{w})$ and $\mathbf{\Sigma}_{\partial v} = \mathrm{diag}(\mathbf{B}_v\mathbf{w})$ (see [12] for detailed derivation). Now, utilizing the matrix inversion lemma we compute $\Delta\mathbf{w}$ as

$$\Delta\mathbf{w} = \mathbf{S}_1^{-1}\mathbf{z}_1 - \mathbf{S}_1^{-1}\mathbf{B}^{\mathrm{T}}(\mathbf{C}^{-1} + \mathbf{B}\mathbf{S}_1^{-1}\mathbf{B}^{\mathrm{T}})^{-1}\mathbf{B}\mathbf{S}_1^{-1}\mathbf{z}_1. \tag{11}$$

Then, by utilizing the value of $\Delta\mathbf{w}$, we can compute $\Delta\mathbf{v}$ and $\Delta\mathbf{u}$ as

$$\Delta\mathbf{v} = \mathbf{D}_3^{-1}(-\mathbf{g}_2 - \mathbf{F}_3^{\mathrm{T}}\Delta\mathbf{w}), \quad \Delta\mathbf{u} = \mathbf{D}_1^{-1}(-\mathbf{g}_3 - \mathbf{D}_2^{\mathrm{T}}\Delta\mathbf{w}). \tag{12}$$

Since $\mathbf{D}_1, \mathbf{D}_2$ are diagonal and $\mathbf{F}_1$ is block diagonal, the matrix $\mathbf{S}_1$ is a block diagonal matrix containing $N^2 \times N^2$ blocks. Thus, the most computationally expensive operation for the block elimination procedure is the inversion of $\mathbf{C}^{-1} + \mathbf{B}\mathbf{S}_1^{-1}\mathbf{B}^{\mathrm{T}} \in \mathbb{R}^{2N^2 \times 2N^2}$ in (11). Note that for the case of $K = 1$, since $\mathbf{S} \in \mathbb{R}^{N^2 \times N^2}$ it is efficient to inverse $\mathbf{S}$ in (9) instead of utilizing (11). Therefore, the cost of the block elimination procedure is $\mathcal{O}(P^3)$ flops with $P = 2N^2$ ($P = N^2$ for $K = 1$), which is much less compared to that of inverting the matrix $\mathbf{H} \in \mathbb{R}^{(2K+1)N^2 \times (2K+1)N^2}$.
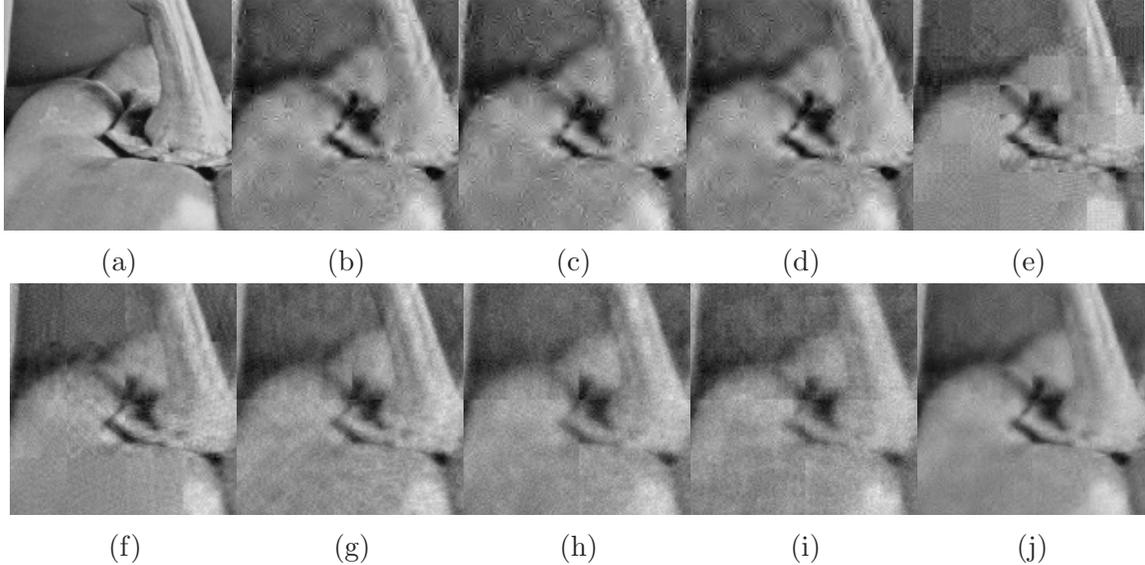
Figure 2: a) Original image b) DWT 'sym6' c) DWT 'bior4.4' d) DWT 'sym8' e) DCT 16×16
f) DCT 32×32 g) DCT 64×64 h) DCT 128×128 i) DCT 256×256 j) Proposed

4   Numerical results

In this section, we evaluate the performance of utilizing multiple bases for compressive sensing reconstruction. Figure 2 presents an example for the improved performance, in terms of the visual quality, of utilizing multiple sparsifying bases in recovering an image from compressive measurements. The figure illustrates a portion of the 'Peppers' image $\mathbf{X}$ reconstructed using different bases as well as using *Algorithm 1* from the compressive measurements $\mathbf{y} = \mathbf{\Phi x}$, where $\mathbf{x}$ is the vectorized version of the 2-D image $\mathbf{X}$. The measurement matrix $\mathbf{\Phi} \in \mathbb{R}^{M \times N^2}$ is generated by selecting random M rows of the 'noiselet transform' matrix [4] at sensing rate $M/N^2 = 0.15$. For each basis, we solve the l_1–norm minimization problem (2) using the NESTA toolbox [11] with the parameters $\mu_f = 10^{-7}$, NESTAmaxiter $= 500$ and $\delta = 0$. Further, in step 3 of *Algorithm 1*, we utilized the simple fusion function $\tilde{\mathbf{x}} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{\Psi}_k \hat{\mathbf{w}}_k$. The figure shows that the reconstructions using only a single basis contain different artifacts which are specific to the given basis, e.g., different ringing artifacts for DWT bases and different blocking artifacts for the DCT bases. The fusion in *Algorithm 1* eliminates these artifacts, and as a result it provides a better subjective visual quality compared to the reconstructions via a single basis.

Table 1 demonstrates the objective quality comparison via Peak Signal-to-Noise Ratio (PSNR) for few example images under different sensing rates. The results show that the basis which achieves the highest PSNR among the considered set of bases depends on the image as well as the sensing rate. For example, the basis DWT 'sym6' achieves the highest PSNR for 'Peppers' image at sensing rate 0.4, while for the same sensing rate the basis DCT 32×32 achieves the highest PSNR for the 'Barbara' image. Similarly, for the same 'Peppers' image at sensing rate 0.25 the basis DCT 16×16 achieves the best PSNR, but at sensing rate 0.4 the basis DWT 'sym6' achieves the best PSNR. However, we can see that the reconstruction via *Algorithm 1* utilizing

Sensing rate	DWT 'sym6'	DWT 'bior4.4'	DWT 'sym8'	DCT 16×16	DCT 32×32	DCT 64×64	DCT 128×128	DCT 256×256	Fused
				Barbara (512×512)					
0.15	20.07	19.77	20.12	22.51	23.15	22.99	22.54	21.86	**23.46**
0.20	21.89	21.63	21.98	24.14	24.70	24.34	23.79	23.00	**25.03**
0.25	23.31	22.91	23.42	25.67	26.12	25.62	24.92	24.09	**26.41**
0.30	24.73	24.32	24.85	27.15	27.43	26.82	26.05	25.10	**27.77**
0.35	26.17	25.70	26.34	28.63	28.74	27.98	27.14	26.09	**29.13**
0.40	27.58	27.11	27.80	29.96	30.01	29.14	28.23	27.15	**30.47**
				Peppers (512×512)					
0.15	24.51	24.02	24.54	25.70	25.92	25.46	24.98	24.70	**27.03**
0.20	26.88	26.42	26.89	27.69	27.62	27.01	26.43	26.13	**28.96**
0.25	29.00	28.67	28.98	29.23	29.01	28.21	27.62	27.28	**30.54**
0.30	30.57	30.24	30.55	30.48	30.10	29.29	28.69	28.32	**31.80**
0.35	31.80	31.54	31.77	31.55	31.10	30.29	29.66	29.26	**32.89**
0.40	32.79	32.60	32.75	32.49	32.01	31.21	30.54	30.18	**33.84**

Table 1: Reconstruction performance comparison in average PSNR (in dB).

multiple bases provides the best PSNR for all images and sensing rates[1].

Table 2 presents the reconstruction performance of fusion via the solution of problem (3) (l_1+TV-fusion) in terms of PSNR. We utilize the proposed interior-point method for solving problem (3) with parameters $\lambda_k = 1, \alpha = 0.02, \beta = 0.3, \mu = 20, \epsilon = 10^{-3}, \epsilon_{nt} = 10^{-3}$ and $\tilde{\mathbf{x}} = \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{\Psi}_k \hat{\mathbf{w}}_k$ as the fusion function. The results show that the reconstruction via the solution of problem (3) even out performs the reconstruction based on the fusion using the estimates from l_1-norm minimization (l_1-fusion). This is due to the fact that problem (3) also take into account that an image has a sparse gradient.

5   Conclusions

We considered the reconstruction of a signal from compressive measurements when the basis in which the signal of interest is the sparsest is unknown. We have shown that fusion of multiple estimates of the signal using different sparsifying bases leads to a better signal recovery compared to reconstructions via any single basis. Further, we derived a customized interior-point method to jointly obtain such multiple estimates of an image from compressive measurements which utilize multiple sparsifying bases as well as the fact that the images usually have a sparse gradient.

6   References

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

---

[1]We refer the interested reader to [12] for more simulation results.

Sensing rate	DWT 'sym6'	DCT 16×16	DCT 32×32	l_1-fusion	l_1+TV-fusion
Peppers (512×512)					
0.20	26.88	27.69	27.62	28.95	**33.01**
0.30	30.57	30.48	30.10	31.91	**34.81**
0.40	32.79	32.49	32.01	33.92	**36.15**
0.50	34.61	34.30	33.71	35.67	**37.36**
Cameraman (256×256)					
0.20	21.97	22.88	22.54	23.84	**27.94**
0.30	24.84	25.44	24.82	26.48	**30.80**
0.40	27.86	27.78	27.11	29.09	**33.50**
0.50	30.65	30.28	29.22	31.59	**36.23**

Table 2: Reconstruction performance comparison between the fusion via l_1-norm minimization and solution of problem (3) in average PSNR (in dB).

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

[3] M. A. Davenport, M. F. Duarte, Y. C. Eldar, and G. Kutyniok, "Introduction to compressed sensing," in *Compressed Sensing: Theory and Applications*. Cambridge University Press, 2012.

[4] E. Candès and J. Romberg, "Sparsity and incoherence in compressive sampling," *Inverse Problems*, vol. 23, no. 3, p. 969, 2007.

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

[6] E. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Trans. Inform. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

[7] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Processing*, vol. 54, no. 11, pp. 4311–4322, Nov 2006.

[8] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 5, 1999, pp. 2443–2446.

[9] M. Elad and I. Yavneh, "A plurality of sparse representations is better than the sparsest one alone," *IEEE Trans. Inform. Theory*, vol. 55, no. 10, pp. 4701–4714, Oct 2009.

[10] E. Candès and J. Romberg, "l1-magic: Recovery of sparse signals via convex programming," Oct. 2005, [Online]. Available: http://statweb.stanford.edu/~candes/l1magic/.

[11] J. Bobin and S. Becker, "NESTA: A fast and accurate first-order method for sparse recovery," Jun. 2011, [Online]. Available: http://statweb.stanford.edu/~candes/nesta/.

[12] U. L. Wijewardhana, E. Belyaev, M. Codreanu, and M. Latva-aho, "Signal recovery in compressive sensing via multiple sparsifying bases: derivations and results," 2016, [Online]. Available: http://www.ee.oulu.fi/~udwijewa/multibases.php.

[13] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, UK: Cambridge University Press, 2004.