NOVA

University of Newcastle Research Online

nova.newcastle.edu.au

Hyder, Mashud; Mahata, Kaushik "Sparse sequence recovery via a maximum a posteriori estimation". Published in Proceedings of the 2013 IEEE International Conference on Image Processing, ICIP 2013 (Melbourne, VIC 15-18 September, 2013) p. 489-493 (2013)

Available from: <http://dx.doi.org/10.1109/ICIP.2013.6738101>

© 2013 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting /republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

Accessed from: <http://hdl.handle.net/1959.13/1342277>

SPARSE SEQUENCE RECOVERY VIA A MAXIMUM A POSTERIORI ESTIMATION

Md Mashud Hyder and Kaushik Mahata

SEECS, University of Newcastle, Callaghan, NSW 2308, Australia.

md.hyder@uon.edu.au, kaushik.mahata@newcastle.edu.au

ABSTRACT

A maximum a posteriori (MAP) estimation algorithm is given for reconstructing sparse signals, where a part of the support, and an approximate estimate of the sparse signal are known. This method is useful, e.g., in magnetic resonance image (MRI) sequence, natural video sequences, etc, where it is required to recursively reconstruct a sequence of mutually correlated sparse signals or images. Here we use the last signal as an a priori estimate of the current signal. The priori information is often inaccurate, and we adopt MAP estimation framework to deal with this issue. Simulation studies are performed, and the algorithm is applied to reconstruct MRI image sequences.

Index Terms—Maximum a posteriori, Gaussian mixture model, partially known support, compressive sensing.

1. INTRODUCTION

Compressed sensing (CS) [1] allows us to recover a k -sparse signal $\mathbf{x}$ from a vector valued linear measurement $\mathbf{y}$ of the form

$$\mathbf{y} = \Phi \mathbf{x}, \quad \Phi \in \mathbb{R}^{m \times n}, \quad m < n, \quad (1)$$

provided that (i) $k \leq m/2$, and (ii) any $m \times m$ matrix consisting of any arbitrary m columns of Φ form a basis of $\mathbb{R}^m$ [1]. The support of $\mathbf{x}$ is defined as $\text{supp}(\mathbf{x}) := \{i \in \mathbb{I} : x_i \neq 0\}$, $\mathbb{I} = \{1, 2, \dots, n\}$. Given $\mathbf{y}$ and Φ , the reconstruction of $\mathbf{x}$ requires solving: $\mathbf{x} = \arg \min_{\mathbf{v}} \|\mathbf{v}\|_0$, subject to $\mathbf{y} = \Phi \mathbf{v}$, where $\|\mathbf{v}\|_0$ is simply the number of nonzero components in $\mathbf{v}$. Unfortunately, solving the problem is NP-hard [2]. Basis Pursuit replaces ℓ_0 norm by ℓ_1 -norm [3, 4], which leads a convex problem admitting efficient solution.

In problems like MRI [5, 6] we recursively reconstruct a sequence of highly correlated images, which admit sparse representations under certain transformations, e.g., wavelets, discrete cosine transform (DCT), total variation (TV), etc. In these problems it is natural to use the information present in the signal at time epoch t to estimate that at epoch $t + 1$. This motivates in CS with partially known support [7, 8]. For example, consider the Larynx MRI sequence in Figure 1(a). Let $\bar{\mathbf{x}}(t)$ be the vector of 2-D wavelet coefficients of the image at time t , and let $\mathbf{x}(t)$ be a vector constructed from $\bar{\mathbf{x}}(t)$ by retaining the most significant components of $\bar{\mathbf{x}}(t)$ from its 99.95%-energy support while setting others to zero. Then there is very little difference between the images reconstructed using $\bar{\mathbf{x}}(t)$ and $\mathbf{x}(t)$. Let $\#X$ be the cardinality of a set X . A measure of the difference between $\text{supp}\{\mathbf{x}(t)\}$ and $\text{supp}\{\mathbf{x}(t-1)\}$ is

$$\vartheta(t) = \frac{\#[\text{supp}\{\mathbf{x}(t)\} \setminus \text{supp}\{\mathbf{x}(t-1)\}]}{\#[\text{supp}\{\mathbf{x}(t)\}]}, \quad (2)$$

Research is supported by the Australian research council.

Fig. 1. (a) Larynx MRI video sequence. (b) Larynx MRI images are sparsified using 2-D discrete wavelet transform. While calculating ϑ in (2), we consider 99.95% energy support of $\bar{\mathbf{x}}(t)$ for wavelet domain.

where $X \setminus Y = \{\mathbf{x} \in X : \mathbf{x} \notin Y\}$. Figure 1(b) shows the plot of $\vartheta(t)$ as a function of t for the MRI sequences in Figure 1(a). Note that in wavelet domain the average change of support is around 3% only. Thus, if we have $\mathbf{x}(t-1)$, we can use it as an estimate of $\mathbf{x}(t)$. Recent papers [7, 8, 9] demonstrate that the prior knowledge about the support allows significant reduction in the number of data samples needed for reconstruction. Most CS with partially known support methods allow only a slow variation in the sparsity pattern. This may not hold for low-frame-rate videos. We propose a MAP estimation approach to combat this issue. We use the prior information to set up a *a priori* density function for $\mathbf{x}$, and use it in MAP for reconstruction. We propose numerical strategies to solve the resulting optimization problem. The idea works also when the observed measurements are noisy. The proposed algorithm has been tested in simulations and real-world data.

2. PROBLEM DEFINITION AND RELATED WORK

Assumption 1 An $m \times m$ matrix consisting of any arbitrary m columns of Φ is non-singular [10].

Notation: We denote $\mathbb{X} := \{\mathbf{x} \in \mathbb{R}^n : \Phi \mathbf{x} = \mathbf{y}\}$. Let $T \subset \mathbb{I}$. Then we denote (using Matlab notation) $\Phi_T = \Phi(:, T)$, i.e. Φ_T is the sub-matrix obtained by extracting the columns of Φ corresponding to the indices in T . We define $T^c := \{k \in \mathbb{I} : k \notin T\}$.

$T\}$. Finally for $T_1 \subset \mathbb{I}$ we define $T_1 \setminus T := T_1 \cap T^c$. For a real number z we denote $\lceil z \rceil$ is the smallest integer not less than z .

Consider the system (1). Let $T = \text{supp}(\mathbf{x})$, and $k = \#T$. Note that T is unknown. However, assume that we have an approximate *a priori* estimate $\bar{\mathbf{x}}$ of $\mathbf{x}$, and $\bar{\mathbf{x}}$ is a sparse vector. Let $T_g \subset \mathbb{I}$ be the support of $\bar{\mathbf{x}}$. Hence one can take T_g as an *a priori* estimate of T . Although it is quite likely that $T_g \neq T$. The set corresponding to “misinformation” is $T_e = T_g \setminus T$, i.e., according to the prior information we think $T_e \subset T$, but in reality $T_e \cap T = \emptyset$. Similarly, the unknown part of the support is $T_u := T \setminus T_g$, i.e. according to the prior information we think $T_u \cap T = \emptyset$, but in reality $T_u \subset T$. Let $g = \#(T_g)$, $u = \#(T_u)$. Note that T_g is known, and thus g is known. The error-sets T_u and T_e are unknown. The problem is to estimate sparse $\mathbf{x}$ that satisfies (1). One way to address the problem is to solve [7]

$$\mathbf{x} = \arg \min_{\mathbf{v} \in \mathbb{X}} \|\mathbf{v}_{T_g^c}\|_0, \quad (3)$$

It has been shown in [7] that under Assumption 1, (3) has a unique solution if $g + 2u < m$. Since (3) is an NP-hard problem [2], several alternative methods have been proposed. Least squares CS-residual (LS-CS) [9] is a two step procedure. First, a least squares (LS) estimate of $\mathbf{x}$ is computed using T_g , along with the associated residual $\bar{\mathbf{y}} = \mathbf{y} - \Phi_{T_g}[\Phi_{T_g}^* \Phi_{T_g}^*]^{-1} \Phi_{T_g}^* \mathbf{y}$, where Φ^* denotes the transpose of Φ . In the second step, LS-CS uses a CS algorithm to find a sparse solution $\hat{\mathbf{x}}$ such that $\bar{\mathbf{y}} = \Phi \hat{\mathbf{x}}$. The final estimate of $\mathbf{x}$ is $\bar{\mathbf{x}} + \hat{\mathbf{x}}$, where $\bar{\mathbf{x}}_{T_g} = [\Phi_{T_g} \Phi_{T_g}^*]^{-1} \Phi_{T_g}^* \mathbf{y}$ and $\bar{\mathbf{x}}_{T_g^c} = 0$. Recently proposed modified-CS [7] uses a modified ℓ_1 minimization approach. Modified-CS estimates $\mathbf{x}$ as: $\mathbf{x} = \arg \min_{\mathbf{v} \in \mathbb{X}} \|\mathbf{v}_{T_g^c}\|_1$. Weighted ℓ_1 minimization [11, 8] approach computes $\mathbf{x}$ as: $\mathbf{x} = \arg \min_{\mathbf{v} \in \mathbb{X}} (\|\mathbf{v}_{T_g}\|_1 + \bar{w} \|\mathbf{v}_{T_g^c}\|_1)$, where $\bar{w}$ is a positive constant whose value needs to be determined from the problem in hand. Regularized modified basis pursuit denoising (reg-mod-BPDN) [12], solves

$$\min_{\mathbf{x}} \gamma \|\mathbf{x}_{T_g^c}\|_1 + \frac{1}{2} \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 + \frac{\bar{\gamma}}{2} \|\mathbf{x}_{T_g} - \bar{\mathbf{x}}_{T_g}\|_2^2 \quad (4)$$

where the values of γ and $\bar{\gamma}$ depend on the underlying problem.

3. MAXIMUM A POSTERIORI ESTIMATION (MAP)

3.1. MAP framework

We assume the commonly used measurement model $\mathbf{y} = \Phi \mathbf{x} + \mathbf{e}$, where $\mathbf{e}$ is the measurement noise and the elements of $\mathbf{e}$ are mutually independent and identically distributed as zero mean Gaussian distribution with variance $1/\lambda$. Then the conditional density of $\mathbf{y}$ given $\mathbf{x}$ is given by

$$p(\mathbf{y}|\mathbf{x}) = \left(\frac{\lambda}{2\pi}\right)^m \exp\{-\lambda \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 / 2\}. \quad (5)$$

Now suppose we have some *a priori* knowledge of $\mathbf{x}$ in terms of a priori probability density function $p(\mathbf{x})$. MAP proposes to estimate $\mathbf{x}$ by maximizing the conditional density $p(\mathbf{x}|\mathbf{y})$ given the observed data $\mathbf{y}$ with respect to $\mathbf{x}$. This is same as maximizing the joint log-likelihood function $\log p(\mathbf{y}|\mathbf{x}) + \log p(\mathbf{x})$ [13]. Here we propose a candidate for $p(\mathbf{x})$, which is suitable to represent the prior knowledge available in form of $\bar{\mathbf{x}}$ and T_g . We denote the Gaussian density function as

$$N(\mathbf{x}_i, \nu_i, \varpi) = \frac{1}{\sqrt{2\pi\varpi^2}} \exp\left\{-\frac{(\mathbf{x}_i - \nu_i)^2}{2\varpi^2}\right\} \quad (6)$$

where, mean of $\mathbf{x}_i$ is ν_i . At first, assume that we have no information available about T . Hence $\mathbf{x}$ has two types of components. For $i \in T^c$ we assume that the probability density function of $\mathbf{x}_i$ is $N(\mathbf{x}_i, 0, \mu)$, where μ is a very small number. For $i \in T$ the density function of $\mathbf{x}_i$ is $N(\mathbf{x}_i, 0, \varrho)$, where $\varrho \gg \mu$. Then for every $i \in \mathbb{I}$ there are two possibilities: (i) $i \in T$ with probability k/n ; and (ii) $i \in T^c$ with probability $1 - k/n$. Hence it is appropriate to postulate that $\mathbf{x}_i$ has a Gaussian mixture density $(k/n)N(\mathbf{x}_i, 0, \varrho) + (1 - k/n)N(\mathbf{x}_i, 0, \mu)$.

When a priori estimate T_g of T and $\bar{\mathbf{x}}$ of $\mathbf{x}$ are given, we adapt accordingly. If $i \in T_g$, then we assume that the mean of $\mathbf{x}_i$ is $\bar{\mathbf{x}}_i$ and variance $\bar{\varrho}$. Note that the magnitude of $\bar{\varrho}$ controls the spread of $\mathbf{x}_i$ around $\bar{\mathbf{x}}_i$ for $i \in T_g$. On the other hand, if $i \in T_g^c$ then we assume that each $\mathbf{x}_i$ has a zero mean distribution, and $i \in T$ with probability $q = u/(n - g)$ because only u elements among $n - g$ elements of T_g^c are in T . Then we can postulate that $\mathbf{x}_i$ has an *a priori* probability density function

$$h(\mathbf{x}_i) = \begin{cases} N(\mathbf{x}_i, \bar{\mathbf{x}}_i, \bar{\varrho}), & i \in T_g \\ qN(\mathbf{x}_i, 0, \varrho) + (1 - q)N(\mathbf{x}_i, 0, \mu), & i \in T_g^c. \end{cases} \quad (7)$$

Assuming that the components of $\mathbf{x}$ are mutually independent, we have $p(\mathbf{x}) = \prod_{i=1}^n h(\mathbf{x}_i)$. Combining this with (5) we have

$$-\log p(\mathbf{y}|\mathbf{x}) - \log p(\mathbf{x}) = \frac{\lambda}{2} \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 - \sum_{i=1}^n \ln[h(\mathbf{x}_i)] + \text{const}, \quad (8)$$

where “const” denotes the terms independent of $\mathbf{x}$. Now using (6) and (7) we have

$$-\ln[h(\mathbf{x}_i)] = \begin{cases} \frac{(\mathbf{x}_i - \bar{\mathbf{x}}_i)^2}{2\bar{\varrho}^2}, & i \in T_g \\ \frac{\mathbf{x}_i^2}{2\varrho^2} - \ln\left[1 + r \exp\left\{-\frac{\mathbf{x}_i^2}{2\sigma^2}\right\}\right] + \text{const}, & i \in T_g^c, \end{cases} \quad (9)$$

where $r = \frac{(1-q)\varrho}{q\mu}$, $\frac{1}{\sigma^2} = \frac{1}{\mu^2} - \frac{1}{\varrho^2}$. The expression (9) plays a central role in the sequel. We use (9) for analyzing the MAP cost function and deriving a numerical algorithm to minimize it.

3.2. Solution strategy

Like many optimization problems encountered in CS literature [14, 15], minimizing (8) is a non-convex problem. To deal with the non-convex optimization problem we use the concept of graduated non-convexity (GNC) [16]. Here we construct a sequence of functions $\varphi_j(\mathbf{x})$, $j = 0, 1, 2, \dots, w$ such that: (i) $\varphi_0(\mathbf{x})$ is quadratic; (ii) $|\varphi_{j+1}(\mathbf{x}) - \varphi_j(\mathbf{x})|$ is small in the neighborhood of the minimizer $\mathbf{x}^{(j)}$ of $\varphi_j(\mathbf{x})$; (iii) $\varphi_w(\mathbf{x}) = \frac{\lambda}{2} \|\mathbf{y} - \Phi \mathbf{x}\|_2^2 - \sum_{i=1}^n \ln[h(\mathbf{x}_i)]$ for a user chosen integer w . Because $\varphi_0(\mathbf{x})$ is quadratic, we can compute $\mathbf{x}^{(0)}$ using the standard analytical expression. Then as $|\varphi_1(\mathbf{x}) - \varphi_0(\mathbf{x})|$ is small in the neighborhood of $\mathbf{x}^{(0)}$, by initializing a numerical algorithm to minimize $\varphi_1(\mathbf{x})$ at $\mathbf{x}^{(0)}$ one has a high probability of converging to $\mathbf{x}^{(1)}$. If we continue this process of initializing the numerical algorithm to optimize $\varphi_{j+1}(\mathbf{x})$ at our estimate of $\mathbf{x}^{(j)}$ obtained by numerically optimizing $\varphi_j(\mathbf{x})$, then one can expect that $\mathbf{x}^{(w)}$ is likely to be the minimizer of $\varphi_w(\mathbf{x})$. The sequence of functions $\varphi_j(\mathbf{x})$, $j = 0, 1, 2, \dots, w$ are constructed as

follows. We choose an appropriate real number σ_1 (more details on how the choices are made will follow shortly), and define

$$\wp_0(\mathbf{x}) = \sum_{i \in T_g} \frac{(\mathbf{x}_i - \bar{\mathbf{x}}_i)^2}{2\bar{\rho}^2} + \sum_{i \in T_g^c} \frac{\mathbf{x}_i^2}{2\rho^2} + \frac{\lambda}{2} \|\mathbf{y} - \Phi\mathbf{x}\|_2^2 \quad (10)$$

$${}_j(\mathbf{x}) = \wp_0(\mathbf{x}) - \sum_{i \in T_g^c} \ln \left[1 + r \exp \left\{ -\frac{\mathbf{x}_i^2}{2\sigma_j^2} \right\} \right] \quad (11)$$

$j = 1, 2, 3, \dots, w$; where $\sigma_j = (\sigma/\sigma_1)^{j/w} \sigma_1$, and w is a user chosen integer. The parameter σ_j controls the degree of non-convexity in $\wp_j$. As we increase the value of j from 0 to w , we gradually transform $\hat{\wp}_j$ from a convex function $\hat{\wp}_0$ to our desired likelihood function $\hat{\wp}_w$. If w is sufficiently large, then the change from $\hat{\wp}_{j-1}$ to $\hat{\wp}_j$ is small, and so is the change from $\mathbf{x}^{(j-1)}$ to $\mathbf{x}^{(j)}$. Next we derive an expression for $\mathbf{x}^{(0)}$. Let R be a diagonal matrix such that

$$R_{ii} = \begin{cases} \bar{\rho}^2, & i \in T_g \\ \rho^2, & i \in T_g^c \end{cases}.$$

Recall that $\mathbf{x}^{(0)}$ is the minimum point of $\wp_0$. Hence, if we differentiate (10) with respect to $\mathbf{x}$ and evaluate at $\mathbf{x}^{(0)}$, we must get zero. Hence it can be shown that

$$\mathbf{x}^{(0)} = R^{-1}\Phi^*(I/\lambda + \Phi R^{-1}\Phi^*)^{-1}(\mathbf{y} - \Phi\bar{\mathbf{x}}) + \bar{\mathbf{x}}. \quad (12)$$

We also apply matrix inversion lemma to obtain final expression in (12). The parameter σ_1 controls the degree of non-convexity in $\wp_1$. If we take $\sigma_1 \rightarrow \infty$, then the logarithmic term in (11) tends to $\ln(1+r)$, making $\wp_1$ a quadratic function. In practice, we take $\sigma_1 \geq 5 \max_i |\mathbf{x}_i^{(0)}|$. This ensures $\exp\{-|\mathbf{x}_i^{(0)}|^2/(2\sigma_1^2)\} \geq 0.99$ for all i . Consequently, $\exp\{-|\mathbf{x}_i|^2/(2\sigma_1^2)\} \approx 1$ for all $\mathbf{x}$ satisfying $\|\mathbf{x} - \mathbf{x}^{(0)}\|_2 < \|\mathbf{x}^{(1)} - \mathbf{x}^{(0)}\|_2$ [17].

In the following, we shall explore some properties of $\mathbf{x}^{(j)}$, and develop a numerical algorithm to compute it. Let us define

$$\xi_j(\mathbf{x}_i) = \begin{cases} \frac{1}{\bar{\rho}^2}, & i \in T_g \\ \frac{1}{\bar{\rho}^2} + \frac{r \exp\left(-\frac{\mathbf{x}_i^2}{2\sigma_j^2}\right)}{\sigma_j^2 \left[1 + r \exp\left(-\frac{\mathbf{x}_i^2}{2\sigma_j^2}\right)\right]}, & i \in T_g^c \end{cases} \quad (13)$$

and an $n \times n$ diagonal matrix $W_j(\mathbf{x})$ as

$$W_j(\mathbf{x}) = \text{diag}\{\xi_j(\mathbf{x}_1), \xi_j(\mathbf{x}_2), \dots, \xi_j(\mathbf{x}_n)\}.$$

From (13) it is readily verified that $\xi_j(\mathbf{x}_i) > 0$ for all i . Hence $W_j(\mathbf{x})$ is a positive definite matrix. Now we can verify that

$$\frac{\partial \wp_j(\mathbf{x})}{\partial \mathbf{x}} = W_j(\mathbf{x})\{\mathbf{x} - \bar{\mathbf{x}}\} - \lambda\Phi^*(\mathbf{y} - \Phi\mathbf{x}). \quad (14)$$

Since $\mathbf{x}^{(j)}$ is the minimum point of $\wp_j$, setting $\mathbf{x} = \mathbf{x}^{(j)}$ in (14) we get

$$\mathbf{x}^{(j)} = g_j(\mathbf{x}^{(j)}) \quad (15)$$

where

$$g_j(\mathbf{x}) := \{W_j(\mathbf{x}) + \lambda\Phi^*\Phi\}^{-1}\{\lambda\Phi^*\mathbf{y} + W_j(\mathbf{x})\bar{\mathbf{x}}\}. \quad (16)$$

Also, using matrix inversion lemma:

$$g_j(\mathbf{x}) := W_j^{-1}(\mathbf{x})\Phi^*[I/\lambda + \Phi W_j^{-1}(\mathbf{x})\Phi^*]^{-1}(\mathbf{y} - \Phi\bar{\mathbf{x}}) + \bar{\mathbf{x}}.$$

The equation $\mathbf{x} = g_j(\mathbf{x})$ is nonlinear, and cannot be solved analytically. One possibility is to use a fixed point iteration. However, the

Table 1. MAP with partially known information (PMAP)

1. Set $\mathbf{x} = \bar{\mathbf{x}} + R^{-1}\Phi^*(I/\lambda + \Phi R^{-1}\Phi^*)^{-1}(\mathbf{y} - \Phi\bar{\mathbf{x}})$.
2. Set $\sigma_1 = 5 \max_i |\mathbf{x}_i|$.
3. Set $j = 1$, choose $\eta \in (0, 1)$, $\beta = 0.5$ and $w \geq 15$
- do {
4. Set $\kappa = 1$.
5. while $\wp_j(\kappa g_j(\mathbf{x}) + (1 - \kappa)\mathbf{x}) > \wp_j(\mathbf{x})$ {
- $\kappa = \beta\kappa$.
- } end
6. $\mathbf{x}_o = \mathbf{x}$, and $\mathbf{x} = \kappa g_j(\mathbf{x}) + (1 - \kappa)\mathbf{x}$
7. If $\frac{\|\mathbf{x} - \mathbf{x}_o\|_2}{\|\mathbf{x}_o\|_2} < \eta$ then $j = j + 1$
- } while $j \leq w$.

convergence of the fixed point iteration not guaranteed. Nevertheless, using (14) and (16) we have

$$g_j(\mathbf{x}) - \mathbf{x} = -\{W_j(\mathbf{x}^{(j)}) + \lambda\Phi^*\Phi\}^{-1} \frac{\partial \wp_j(\mathbf{x})}{\partial \mathbf{x}}. \quad (17)$$

The matrix $W_j(\mathbf{x}^{(j)}) + \lambda\Phi^*\Phi$ is positive definite. Hence (17) implies that $\wp_j(\mathbf{x})$ is decreasing along the vector $g_j(\mathbf{x}) - \mathbf{x}$. In fact moving to $g_j(\mathbf{x})$ from $\mathbf{x}$ is same as taking the Newton step associated with some convex-concave procedure to minimize $\wp_j(\mathbf{x})$ [18].

The ‘‘MAP with partially known information’’ (PMAP) optimization strategy is given in Table 1. We assume that the values of ϱ , $\bar{\varrho}$, μ , u and k are known (a direction of assuming values is given below). Using initial $\mathbf{x}^{(0)}$ we calculate σ_1 , and in Step 3 some parameters including w are set. Subsequently, we run an iterative algorithm, which loops through Steps 4-7. For the current value of j , the objective is to minimize $\wp_j$. At the current value of $\mathbf{x}$ we choose $g_j(\mathbf{x}) - \mathbf{x}$ as the decent direction, and in Steps 4, 5 we choose an appropriate step size using a standard backtracking line search. Once an appropriate step size is chosen, we update $\mathbf{x}$, and in Step 7 we check if the relative change in of $\mathbf{x}$ is below η . If yes, then we decide that we are sufficiently close to $\mathbf{x}^{(j)}$, and move on to the next function $\wp_{j+1}$ by incrementing j . Our experimental study demonstrate that having $\eta = 0.02$ and $w = 20$ make a good trade-off between solution accuracy and computation time.

Let us denote the recovered version of $\mathbf{x}(t)$ by $\hat{\mathbf{x}}(t)$. Let $T(t, \tau)$ be the support of most significant components of $\hat{\mathbf{x}}(t)$ where $\tau\%$ -energy resides. Our data-adaptive approach for ‘learning’ the parameters k , ϱ , $\bar{\varrho}$ and u to be used for recovering $\mathbf{x}(t+1)$ from $\mathbf{y}(t+1)$ is as follows. We select an integer valued ‘horizon’ S (typically 2), and set a value of τ (typically 99). Then we set $T_g = T(t, \tau)$ and construct $\bar{\mathbf{x}}$ such that

$$\bar{\mathbf{x}}_i = \begin{cases} \hat{\mathbf{x}}_i(t), & i \in T(t, \tau), \\ 0, & i \in T^c(t, \tau), \end{cases}$$

where $T^c(t, \tau) = \mathbb{I} \setminus T(t, \tau)$. Also we set

$$k = \frac{1}{S} \sum_{j=t-S+1}^t \#\{T(j, \tau)\},$$

$u = \frac{1}{S} \sum_{j=t-S+1}^t \#\{T(j, \tau) \setminus T(j-1, \tau)\}$. We calculate $\bar{\varrho}^2$ as the variance of the change $\hat{\mathbf{x}}_i(j) - \hat{\mathbf{x}}_i(j-1)$ over $j = \{t-S+1, \dots, t-1, t\}$, and $i \in T(j-1, \tau)$. Similarly we take ϱ^2 as the variance of the $\hat{\mathbf{x}}_i(j)$ over $j = \{t-S+1, \dots, t-1, t\}$, and $i \in T(j, \tau) \setminus T(j, \tau/2)$.

Fig. 2. Performance of different algorithms are evaluated for different values of μ_p and m as a function of support change ϑ . (a) $\mu_p = 0.1$, $m = \lceil 0.2n \rceil$, (b) $\mu_p = 0.1$, $m = \lceil 0.25n \rceil$, (c) $\mu_p = 0.2$, $m = \lceil 0.25n \rceil$

4. EXPERIMENTAL RESULTS

We compare the performance of PMAP algorithm with LS-CS [9], reg-mod-BPDN [12], CS-residual, modified-CS-residual (Mod-CS-res) [19] and weighted ℓ_1 [11, 8]. For weighted ℓ_1 we set $\bar{w} = 3$, as recommended in [11]. We take the standard deviation of the measurement noise as 10^{-4} in simulation studies. The artificial data sets are generated by following the procedure given in [12]. The length of vector $\mathbf{x}(1)$ is $n = 512$. We choose a support set $T^{(1)} \subset \mathbb{I}$. The elements of $T^{(1)}$ are chosen uniformly at random among $\binom{n}{k}$ choices. Then we generate $\mathbf{x}(1)$ such that $\mathbf{x}_i(1)$ is a zero-mean Gaussian random variable with variance 1 when $i \in T^{(1)}$, and $\mathbf{x}_i(1)$ is a zero-mean Gaussian random variable with variance μ_e^2 when $i \in \mathbb{I} \setminus T^{(1)}$. Subsequently, for the t -th vector ($t = 2, \dots, M$), we generate the support set $T^{(t)}$ such that $\vartheta(t)$ (see (2)) remains fixed. Then we generate $\mathbf{x}(t)$ such that (i) $[\mathbf{x}(t) - \mathbf{x}(t-1)]_{T^{(t)} \cap T^{(t-1)}}$ is a zero mean Gaussian random vector with covariance matrix $\mu_p^2 I$; (ii) $\mathbf{x}(t)_{T^{(t)} \setminus T^{(t-1)}}$ is a zero mean Gaussian random vector with identity covariance matrix; (iii) $\mathbf{x}(t)_{T^{(t)c}}$ is a zero mean Gaussian random vector with covariance matrix $\mu_e^2 I$. We set $\mu_e = 0.001$. The elements of $\Phi \in \mathbb{R}^{m \times n}$ are mutually independent Gaussian random variables. We take 50% compressive measurements $\mathbf{y}(1)$ and $\mathbf{y}(2)$ of the first two vectors $\mathbf{x}(1)$ and $\mathbf{x}(2)$, and solve normal CS based optimization to recover $\mathbf{x}(1)$ and $\mathbf{x}(2)$. We set $\tau = 99$ and $M = 10$. The normalized root mean squared error [7] defined as $\text{NRMSE} := \|\mathbf{x}_0 - \hat{\mathbf{x}}\|_2 / \|\mathbf{x}_0\|_2$ is used as performance-measure.

For $t > 2$, we consider two different values of m and μ_p . The average size of $T^{(t)}$ is $\lceil 0.1n \rceil$. The support change $\vartheta(t)$ (see (2)) varies between $0.1\#T^{(t)}$ to $0.3\#T^{(t)}$. For each setup we generate 100 different sparse vector sequences $\{\mathbf{x}(t)\}_{t=1}^M$ by using the procedure described above. At first, we set $\mu_p = 0.1$ and $m = \lceil 0.2n \rceil$. The reconstruction performance of algorithms are shown in Figure 2(a). Note that PMAP outperforms reg-mod-BPDN and CS-residual. The performance reg-mod-BPDN and weighted- ℓ_1 are almost similar for lower ϑ . The reg-mod-BPDN outperforms CS-residual with small margin. PMAP performs better until $\vartheta = 0.2$ and its performance degrades quickly when $\vartheta > 0.2$. This does not happen when we increase $m = \lceil 0.25n \rceil$ in Figure 2(b). We found that the NRMSE of PMAP remains almost robust (NRMSE is around 0.0035) with ϑ . The NRMSE of reg-mod-BPDN performs steadily up to $\vartheta = 0.15$ (NRMSE=0.008). Finally, in Figure 2(c), we consider a case when signal amplitude varies considerably with time.

Fig. 3. Reconstructing 32×32 MRI sequence from partial Fourier measurements. $m = \lceil 0.3n \rceil$ for $t = 1, 2$ and $m = \lceil 0.136n \rceil$ for $t > 2$. (a) Larynx sequence. (b) Cardiac sequence.

We set $\mu_p = 0.2$. PMAP still remains robust in the situation.

We then use some real MRI data: Larynx and Cardiac image sequences [7], to evaluate performances of different algorithms. We follow the procedure of [12] to undersample the MRI sequences in Fourier domain. The 2D DWT matrix associated with a two-level Daubechis-4 wavelet is used as a sparsity basis. We set $\tau = 99.5$. The reconstruction performance of different algorithms for the Larynx and Cardiac sequence are summarized in Figure 3(a), 3(b).

5. REFERENCES

- [1] E. J. Candés, J. Romberg, and T. Tao, “Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information,” *IEEE Transactions on Information Theory*, vol. 52, pp. 489–509, Feb. 2006.
- [2] B. K. Natarajan, “Sparse approximate solutions to linear systems,” *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, 1995.
- [3] D. L. Donoho, “Compressed sensing,” *IEEE Transactions on Information Theory*, vol. 52, pp. 1289–1306, Apr. 2006.
- [4] E.J. Candes and T. Tao, “Near-optimal signal recovery from random projections: Universal encoding strategies?,” *Information Theory, IEEE Transactions on*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.
- [5] Hong Jung, Jaeseok Park, Jaeheung Yoo, and Jong Chul Ye, “Radial k-t FOCUSS for high-resolution cardiac cine MRI,” *Magnetic Resonance in Medicine*, vol. 63, no. 1, pp. 68–78, 2010.
- [6] B. Girod, A.M. Aaron, S. Rane, and D. Rebollo-Monedero, “Distributed video coding,” *Proceedings of the IEEE*, vol. 93, no. 1, pp. 71–83, 2005.
- [7] N. Vaswani and Wei Lu, “Modified-CS: Modifying compressive sensing for problems with partially known support,” *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4595–4607, 2010.
- [8] M.A. Khajehnejad, Weiyu Xu, A.S. Avestimehr, and B. Hassibi, “Analyzing weighted ℓ_1 minimization for sparse recovery with nonuniform sparse models,” *Signal Processing, IEEE Transactions on*, vol. 59, no. 5, pp. 1985–2001, may 2011.
- [9] N. Vaswani, “LS-CS-residual (LS-CS): Compressive sensing on least squares residual,” *Signal Processing, IEEE Transactions on*, vol. 58, no. 8, pp. 4108–4120, 2010.
- [10] S.F. Cotter, B.D. Rao, Kjersti Engan, and K. Kreutz-Delgado, “Sparse solutions to linear inverse problems with multiple measurement vectors,” *IEEE Transactions on Signal Processing*, vol. 53, no. 7, pp. 2477–2488, July 2005.
- [11] M. Amin Khajehnejad, Weiyu Xu, A. Salman Avestimehr, and B. Hassibi, “Weighted ℓ_1 minimization for sparse recovery with prior information,” in *Information Theory, 2009. ISIT 2009. IEEE International Symposium on*, July 2009, pp. 483–487.
- [12] Wei Lu and N. Vaswani, “Regularized modified BPDN for noisy sparse reconstruction with partial erroneous support and signal value knowledge,” *Signal Processing, IEEE Transactions on*, vol. 60, no. 1, pp. 182–196, jan. 2012.
- [13] Steven M. Kay, *Fundamentals of statistical signal processing: estimation theory*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.
- [14] J. Trzasko and A. Manduca, “Relaxed conditions for sparse signal recovery with general concave priors,” *IEEE Transactions on Signal Processing*, vol. 57, no. 11, pp. 4347–4354, 2009.
- [15] R. Chartrand, “Exact reconstruction of sparse signals via non-convex minimization,” *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 707–710, Oct. 2007.
- [16] A. Blake and A. Zisserman, *Visual Reconstruction*, MIT Press, Cambridge, MA, 1987.
- [17] H. Mohimani, M. Babaie-Zadeh, and C. Jutten, “A fast approach for overcomplete sparse decomposition based on smoothed ℓ^0 norm,” *IEEE Transactions on Signal Processing*, vol. 57, no. 1, pp. 289–301, Jan. 2009.
- [18] M.M. Hyder and K. Mahata, “Direction-of-arrival estimation using a mixed $\ell_{2,0}$ norm approximation,” *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4646–4655, 2010.
- [19] Wei Lu and N. Vaswani, “Modified compressive sensing for real-time dynamic MR imaging,” in *Image Processing (ICIP), 2009 16th IEEE International Conference on*, nov. 2009, pp. 3045–3048.
- [20] Michael Lustig, David Donoho, and John M Pauly, “Sparse MRI: The application of compressed sensing for rapid MR imaging,” *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1182–95, 2007.