

EXPLOITING THE CONVEX-CONCAVE PENALTY FOR TRACKING: A NOVEL DYNAMIC REWEIGHTED SPARSE BAYESIAN LEARNING ALGORITHM

Yu Wang^{*}, David Wipf[†], Wei Chen^{*} and Ian J. Wassell^{*},

^{*} Computer Laboratory, University of Cambridge, UK

[†] Microsoft Research, Beijing, China

yw323@cam.ac.uk, davidwip@microsoft.com, wc253@cam.ac.uk, ijw24@cam.ac.uk

ABSTRACT

We propose a novel dynamic reweighted ℓ_2 ($\text{DR}\ell_2$) algorithm in the regime of dynamic compressive sensing. Our analysis shows that aiming to solve a Type II optimization problem, $\text{DR}\ell_2$ is effectively minimizing a ‘convex-concave’ penalty in the coefficients that transitions from a convex region to a concave function using knowledge of past estimations. $\text{DR}\ell_2$ thus provides superior reconstruction performance compared with state-of-the-art dynamic CS algorithms.

1. INTRODUCTION

Most recent work on dynamic compressive sensing (CS) [1] has focused on sparse signal estimation by employing a sparsity-inducing prior distribution in the ‘source space’ (Type I approach). Among these, multiple measurement vector (MMV) [2] approach is widely used to model a sequence of temporally correlated signals. MMV model assumes that the support (indices corresponding to the non-zero values in the signal) of the correlated sparse signal vectors are identical in the source space. However, the support of a practical time-varying signal changes as time elapses rather than constant [3], even if it is signaturred as a slow dynamic process.

In view of the success of sparse Bayesian learning (SBL) approaches [4] [5], we are interested in developing Type II algorithms that operate and optimize in the ‘parameter space’, to reconstruct a sequence of slowly time-varying signals in the presence of support innovations. In [6], Wipf et al. prove that Type II approaches seek optimal values in the parameter space while it is equivalently minimizing an implicit ‘nonseparable’ cost function in the source space. In this way, a nonseparable penalty introduces dependencies between entries in the source space and hence offers a tighter approximation to the ℓ_0 norm in promoting sparsity. The Type II approach is also employed in [7], where the common sparsity assumption is imposed by introducing a constant temporal correlation to ensure that the MMV model is satisfied.

In comparison, we propose a novel CS approach, the dynamic reweighted ℓ_2 ($\text{DR}\ell_2$) algorithm. We show that the proposed algorithm benefits from an embedded ‘convex-

concave’ penalty which is not present in conventional CS. Simulation results show that $\text{DR}\ell_2$ achieves better performance than the state-of-the-art algorithms in reconstructing signals having slow support changes. We also emphasize that the proposed algorithm is free from the MMV model and that it does not require any support detection schemes. Thus, $\text{DR}\ell_2$ relaxes the common support assumption and further avoids the error introduced by the threshold required in [8] [9]. The notation $\text{diag}(\mathbf{X})$ extracts the diagonal entries from matrix \mathbf{X} to form a vector and $\text{Diag}(\mathbf{x})$ constructs a diagonal matrix using the elements from vector \mathbf{x} . X_{ii} is the i^{th} diagonal element of matrix \mathbf{X} .

2. A NOVEL STATISTICAL DYNAMIC MODEL WITH PRIOR KNOWLEDGE

In the regime of CS reconstruction, we consider recovering the sparse signal \mathbf{x}_t at time t using the sampling system:

$$\mathbf{y}_t = \Phi \mathbf{x}_t + \mathbf{n}_t, \quad (1)$$

where $\mathbf{y}_t \in \mathbb{R}^N$ denotes the current compressive measurements, and the known sensing matrix $\Phi \in \mathbb{R}^{N \times M}$, ($N < M$) projects the sparse signal $\mathbf{x}_t \in \mathbb{R}^M$ from the source space into measurement space. Here, for simplicity, Φ is assumed to be constant during the dynamic process.

2.1. Sparse Bayesian Learning (SBL) Model

In the original SBL formulation [4], the signal prior is given by the Gaussian distribution:

$$p(\mathbf{x}_t; \boldsymbol{\gamma}) = \prod_{i=1}^M (2\pi\gamma_i)^{-1/2} \exp\left(-\frac{x_{t,i}^2}{2\gamma_i}\right). \quad (2)$$

The vector $\boldsymbol{\gamma} = [\gamma_1, \gamma_2, \dots, \gamma_M]^T$ consists of the hyperparameters γ_i that control the prior variance of each signal entry $x_{t,i}$. The noise \mathbf{n}_t is parameterized with the variance σ^2 . The likelihood of the complete data set is given by:

$$p(\mathbf{y}_t | \mathbf{x}_t, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left\{-\frac{1}{2\sigma^2} \|\mathbf{y}_t - \Phi \mathbf{x}_t\|_2^2\right\}. \quad (3)$$

The posterior density of \mathbf{x}_t is also Gaussian. Using $\mathbf{\Gamma} = \text{Diag}(\boldsymbol{\gamma})$, the posterior mean for \mathbf{x}_t is obtained as $\boldsymbol{\mu}_t$, with the estimation covariance $\boldsymbol{\Sigma}_x$:

$$\boldsymbol{\mu}_t = (\boldsymbol{\Phi}^T \boldsymbol{\Phi} + \sigma^2 \mathbf{\Gamma}^{-1})^{-1} \boldsymbol{\Phi}^T \mathbf{y}_t, \boldsymbol{\Sigma}_x = (\sigma^{-2} \boldsymbol{\Phi}^T \boldsymbol{\Phi} + \mathbf{\Gamma}^{-1})^{-1}. \quad (4)$$

Type II method alternatively maximizes the marginal likelihood and gives corresponding cost function in the $\boldsymbol{\gamma}$ space:

$$p(\mathbf{y}_t | \boldsymbol{\gamma}, \sigma^2) = \int p(\mathbf{y}_t | \mathbf{x}_t; \sigma^2) p(\mathbf{x}_t; \boldsymbol{\gamma}) d\mathbf{x}_t \quad (5)$$

$$\bar{L}(\boldsymbol{\gamma}) = \mathbf{y}_t^T \boldsymbol{\Sigma}_y^{-1} \mathbf{y}_t + \log |\boldsymbol{\Sigma}_y|. \quad (6)$$

The cost function $\bar{L}(\boldsymbol{\gamma})$ is then optimized by iterating expectation maximization (EM) steps among $\boldsymbol{\mu}_t$ and $\boldsymbol{\gamma}$ [4].

2.2. A New Expectation Maximization Formulation For Slow Dynamic Process

The work in [10] maximizes the likelihood on signal innovation $\Delta \mathbf{x}$ while it essentially still optimizes Eq.(5) having \mathbf{x}_{t-1} as the prior mean for $p(\mathbf{x}_t; \boldsymbol{\gamma})$. Instead, in order to specify a slow varying signal $\mathbf{x}_t \in \mathbb{R}^M$ at time stamp t , we propose to use the following dynamic model involving an improper hyperprior probability density $p(\boldsymbol{\gamma})$:

$$\mathbf{x}_t = \mathbf{x}_{t-1} + \Delta \mathbf{x}, \mathbf{y}_t = \mathbf{y}_{t-1} + \Delta \mathbf{y}, \Delta \mathbf{y} = \boldsymbol{\Phi} \Delta \mathbf{x} + \mathbf{n}_t \quad (7)$$

$$p(\boldsymbol{\gamma}) \propto \prod_i e^{-\frac{x_{t-1,i}^2}{2\gamma_i}}, \quad p(\Delta \mathbf{x} | \boldsymbol{\gamma}) = \prod_i (2\pi\gamma_i)^{-\frac{1}{2}} e^{-\frac{\Delta x_i^2}{2\gamma_i}} \quad (8)$$

where $\Delta \mathbf{x}$ represents the signal innovation vector between the time stamps t and $t-1$, \mathbf{n}_t follows $\mathcal{N}(0, \sigma^2)$. By imposing an improper hyperprior probability density $p(\boldsymbol{\gamma})$ as in Eq.(8), we propose to optimize:

$$p(\Delta \mathbf{y} | \boldsymbol{\gamma}, \sigma^2) = \int p(\Delta \mathbf{y} | \Delta \mathbf{x}) p(\Delta \mathbf{x} | \boldsymbol{\gamma}) p(\boldsymbol{\gamma}) d\Delta \mathbf{x}. \quad (9)$$

By taking log of Eq.(9) and omitting the irrelevant terms, we yield the new cost function:

$$\hat{L}(\boldsymbol{\gamma}) = \Delta \mathbf{y}^T \boldsymbol{\Sigma}_{\Delta \mathbf{y}}^{-1} \Delta \mathbf{y} + \log |\boldsymbol{\Sigma}_{\Delta \mathbf{y}}| + \mathbf{x}_{t-1}^T \mathbf{\Gamma}^{-1} \mathbf{x}_{t-1}, \quad (10)$$

where $\boldsymbol{\Sigma}_{\Delta \mathbf{y}} = \sigma^2 \mathbf{I} + \boldsymbol{\Phi} \mathbf{\Gamma} \boldsymbol{\Phi}^T$ for which the optimal parameters γ_i could not be obtained in closed form. The EM formulation hence alternatively optimizes the conditional expectation $E_{\Delta \mathbf{x} | \mathbf{y}, \boldsymbol{\gamma}, \sigma^2} \log[p(\Delta \mathbf{y} | \Delta \mathbf{x}) \cdot p(\Delta \mathbf{x} | \boldsymbol{\gamma}) \cdot p(\boldsymbol{\gamma})]$, which through differentiation with regard to $\boldsymbol{\gamma}$ yields the optimal γ_i as:

$$\gamma_i = E_{\Delta \mathbf{x} | \mathbf{y}, \boldsymbol{\gamma}, \sigma^2} (\Delta x_i^2 + x_{t-1,i}^2) = (\Delta \mu_i)^2 + \Sigma_{\Delta x, ii} + x_{t-1,i}^2, \quad (11)$$

where $\Sigma_{\Delta x} = (\sigma^{-2} \boldsymbol{\Phi}^T \boldsymbol{\Phi} + \mathbf{\Gamma}^{-1})^{-1}$. The expectation of $\Delta \mathbf{x}$ is derived as $\Delta \boldsymbol{\mu} = \sigma^{-2} \boldsymbol{\Sigma}_{\Delta x} \boldsymbol{\Phi}^T \Delta \mathbf{y}$. The effect of $p(\boldsymbol{\gamma})$ is as follows. When $|x_{t-1,i}|$ is large, the corresponding γ_i is discouraged from having a value near zero, and therefore the

associated larger variance of Δx_i means $|\Delta x_i|$ is less likely to be near zero. This is consistent with the intuition that existing large entries $|x_{t-1,i}|$ are likely to have a wider dynamic range than small or zero-valued ones. Larger $|x_{t-1,i}|$ also naturally requires larger changes wherever support pruning is necessary to occur. In contrast, when $|x_{t-1,i}|$ becomes small, the hyperprior on γ_i becomes increasingly flat, allowing the inherent sparsity-inducing mechanism of SBL to dominate. This implies that the associated Δx_i is likely to be near zero unless it is particularly useful in representing $\Delta \mathbf{y}$. When $x_{t-1,i} = 0$, the hyperprior becomes perfectly flat, and Eq.(11) reduces to the standard SBL. Note that without $p(\boldsymbol{\gamma})$, SBL applied to the innovations even favors sparse variations on nonzero entries in \mathbf{x}_{t-1} , making gradual signal drifts difficult to model.

3. DYNAMIC REWEIGHTED ALGORITHM DR ℓ_2

The optimization of Eq.(9) differs from Eq.(5) in involving \mathbf{x}_{t-1} as the prior mean for $p(\mathbf{x}_t; \boldsymbol{\gamma})$ and the presumed improper hyperprior for $p(\boldsymbol{\gamma})$. To investigate the mechanism behind the intuitive $\boldsymbol{\gamma}$ updating rule proposed in Sec.2.2, we construct an upper-bound on the cost function Eq.(10) in the source space $\Delta \mathbf{x}$, from which we retrieve the identical update rule, independently from the statistical standpoint. Since $\min_{\Delta \mathbf{x}} \frac{1}{\sigma^2} \|\Delta \mathbf{y} - \boldsymbol{\Phi} \Delta \mathbf{x}\|_2^2 + \Delta \mathbf{x}^T \mathbf{\Gamma}^{-1} \Delta \mathbf{x} = \Delta \mathbf{y}^T \boldsymbol{\Sigma}_{\Delta \mathbf{y}}^{-1} \Delta \mathbf{y}$ holds, we majorize $\hat{L}(\boldsymbol{\gamma})$ into $\mathcal{L}(\boldsymbol{\gamma}, \Delta \mathbf{x})$:

$$\mathcal{L}(\boldsymbol{\gamma}, \Delta \mathbf{x}) = \frac{1}{\sigma^2} \|\Delta \mathbf{y} - \boldsymbol{\Phi} \Delta \mathbf{x}\|_2^2 + \Delta \mathbf{x}^T \mathbf{\Gamma}^{-1} \Delta \mathbf{x} + \mathbf{x}_{t-1}^T \mathbf{\Gamma}^{-1} \mathbf{x}_{t-1} + \log |\boldsymbol{\Sigma}_{\Delta \mathbf{y}}|, \quad (12)$$

on which we proceed to minimize over $\boldsymbol{\gamma}$:

$$\begin{aligned} L(\Delta \mathbf{x}) &\triangleq \min_{\boldsymbol{\gamma} \geq \mathbf{0}} \mathcal{L}(\boldsymbol{\gamma}, \Delta \mathbf{x}) \\ &= \frac{1}{\sigma^2} \|\Delta \mathbf{y} - \boldsymbol{\Phi} \Delta \mathbf{x}\|_2^2 + \min_{\boldsymbol{\gamma} \geq \mathbf{0}} (\Delta \mathbf{x}^T \mathbf{\Gamma}^{-1} \Delta \mathbf{x} \\ &\quad + \mathbf{x}_{t-1}^T \mathbf{\Gamma}^{-1} \mathbf{x}_{t-1} + \log |\boldsymbol{\Sigma}_{\Delta \mathbf{y}}|) \\ &= \frac{1}{\sigma^2} \|\Delta \mathbf{y} - \boldsymbol{\Phi} \Delta \mathbf{x}\|_2^2 + g_{dr}(\Delta \mathbf{x}). \end{aligned} \quad (13)$$

By using $h^*(\mathbf{z})$ to represent the concave conjugate [11] of $\log |\boldsymbol{\Sigma}_{\Delta \mathbf{x}}^{-1}|$ with respect to vector $\text{diag}(\mathbf{\Gamma}^{-1})$, $g_{dr}(\Delta \mathbf{x})$ is upper bounded in the $\Delta \mathbf{x}$ space:

$$\begin{aligned} g_{dr}(\Delta \mathbf{x}) &\leq \mathbf{x}_{t-1}^T \mathbf{\Gamma}^{-1} \mathbf{x}_{t-1} + \Delta \mathbf{x}^T \mathbf{\Gamma}^{-1} \Delta \mathbf{x} + \log |\boldsymbol{\Sigma}_{\Delta \mathbf{y}}| \\ &= \mathbf{x}_{t-1}^T \mathbf{\Gamma}^{-1} \mathbf{x}_{t-1} + \Delta \mathbf{x}^T \mathbf{\Gamma}^{-1} \Delta \mathbf{x} + \log |\mathbf{\Gamma}| + \log |\boldsymbol{\Sigma}_{\Delta \mathbf{x}}^{-1}| + \log \sigma^{2N} \\ &\leq \mathbf{x}_{t-1}^T \mathbf{\Gamma}^{-1} \mathbf{x}_{t-1} + \Delta \mathbf{x}^T \mathbf{\Gamma}^{-1} \Delta \mathbf{x} + \log |\mathbf{\Gamma}| \\ &\quad + \mathbf{z}^T \boldsymbol{\gamma}^{-1} - h^*(\mathbf{z}) + 2N \log \sigma \\ &= 2N \log \sigma - h^*(\mathbf{z}) + \sum_i \left(\frac{x_{t-1,i}^2 + \Delta x_i^2}{\gamma_i} + \frac{z_i}{\gamma_i} + \log \gamma_i \right), \end{aligned} \quad (14)$$

We define the sum term in Eq.(14) as $f_{\ell_2}(\Delta \mathbf{x}, \boldsymbol{\gamma}, \mathbf{z})$, and perform coordinate descent method over:

$$\min_{\Delta \mathbf{x}; \boldsymbol{\gamma}, \mathbf{z} \geq \mathbf{0}} \frac{1}{\sigma^2} \|\Delta \mathbf{y} - \boldsymbol{\Phi} \Delta \mathbf{x}\|_2^2 - h^*(\mathbf{z}) + f_{\ell_2}(\Delta \mathbf{x}, \boldsymbol{\gamma}, \mathbf{z}), \quad (15)$$

Algorithm 1 Dynamic Reweighted ℓ_2 algorithm (DR ℓ_2)

Input: Sensing matrix Φ , past estimation \mathbf{x}_{t-1} measurements \mathbf{y}_t and \mathbf{y}_{t-1}

Initialization: $k = 0$, $\mathbf{x}_t = \mathbf{0}$, $\Delta \mathbf{x} = \mathbf{0}$, $\gamma_i = 1, i \leq M$;

while halting criterion false **do**

$k = k + 1$; $\diamond k$ is the iteration index

$\Delta \mathbf{x}^k = \sigma^{-2} \Sigma_x \Phi^T (\mathbf{y}_t - \mathbf{y}_{t-1})$;

$\mathbf{x}_t^k = \mathbf{x}_{t-1} + \Delta \mathbf{x}^k$;

$\gamma_i = x_{t-1,i}^2 + \Sigma_{\Delta x,ii} + \Delta x_i^2, \gamma = [\gamma_1, \gamma_1 \dots \gamma_M]^T$;

$\Sigma_{\Delta x} = (\sigma^{-2} \Phi^T \Phi + \Gamma^{-1})^{-1}$;

end while

return \mathbf{x}_t ;

in which $\gamma_i = x_{t-1,i}^2 + \Delta x_i^2 + z_i$ is the minimizer having fixed z_i and Δx_i , while the optimal $\mathbf{x}_{t,op}$ in each iteration becomes:

$$\mathbf{x}_{t,op} = \mathbf{x}_{t-1} + \sigma^{-2} \Sigma_x \Phi^T (\mathbf{y}_t - \mathbf{y}_{t-1}), \quad (16)$$

and $z_i = \frac{\partial \log |\Sigma_{\Delta x}|}{\partial \gamma_i} = \Sigma_{\Delta x,ii}$ is the optimal value due to the concavity property of $\log |\Sigma_{\Delta x}^{-1}|$ [11]. If we replace the notation $\Sigma_{\Delta x,ii}$ in Eq.(11) with the z_i obtained here, it is clear that the coordinate descent updates derived here exactly match the EM steps proposed in Sec.2.2. Ignoring irrelevant terms, the effective regularization with respect to $\Delta \mathbf{x}$ behind the EM formulation becomes:

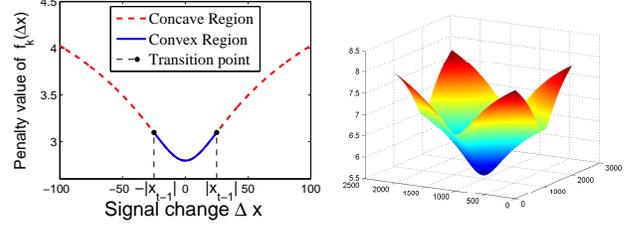
$$\min_{\Delta \mathbf{x}; \mathbf{z}, \gamma \geq 0} \frac{1}{\sigma^2} \|\Delta \mathbf{y} - \Phi \Delta \mathbf{x}\|_2^2 + f_{\ell_2}(\Delta \mathbf{x}, \gamma, \mathbf{z}), \quad (17)$$

$$f_{\ell_2}(\Delta \mathbf{x}, \gamma, \mathbf{z}) = \sum_i \left(\frac{x_{t-1,i}^2 + \Delta x_i^2}{\gamma_i} + \frac{z_i}{\gamma_i} + \log \gamma_i \right). \quad (18)$$

We now concentrate our attention to the penalty $f_{\ell_2}(\Delta \mathbf{x}, \gamma, \mathbf{z})$ satisfying:

$$\begin{aligned} \sum_i \log(x_{t-1,i}^2 + \Delta x_i^2) &\leq \sum_i \log(x_{t-1,i}^2 + \Delta x_i^2 + z_i) \\ &\leq \sum_i \left(\frac{x_{t-1,i}^2 + \Delta x_i^2}{\gamma_i} + \frac{z_i}{\gamma_i} + \log \gamma_i - 1 \right) \\ &= f_{\ell_2}(\Delta \mathbf{x}, \gamma, \mathbf{z}) - M, \end{aligned} \quad (19)$$

for which the second equality becomes tight iff $\gamma_i = x_{t-1,i}^2 + \Delta x_i^2 + z_i$ holds. It is now clear that the minimizer of the object function Eq.(15) is essentially iterating in $\Delta \mathbf{x}$ space to minimize the strict upper-bounding hyperplanes of $\sum_i \log(x_{t-1,i}^2 + \Delta x_i^2 + z_i)$. The non-negative duality term z_i forming the first inequality in Eq.(19) recalls the tuning term ϵ used in [12]. In fact, [12] and [13] have shown that tuning factors in similar forms potentially assist in avoiding convergence to local minima and also stabilize the reweighted reconstruction. In the noiseless case, $z_i = 0$ upon the convergence to local optima, and the kernel of the penalty in $\Delta \mathbf{x}$ space is effectively the upper-bound of $f_k(\Delta \mathbf{x}) = \sum_i \log(x_{t-1,i}^2 + \Delta x_i^2)$.



(a) 1D example of $f_k(\Delta \mathbf{x})$.

(b) 2D example of $f_k(\Delta \mathbf{x})$.

Fig. 1. Visualization of Penalty $f_k(\Delta \mathbf{x})$

4. THE CONVEX-CONCAVE PENALTY

A Log-sum penalty is a concave non-decreasing function of the coefficient magnitudes, therefore risking convergence to local minima. On the other hand, the concave penalty favors sparse solutions while its convex counterparts guarantee global optima. We show here that the penalty $f_k(\Delta \mathbf{x}) = \sum_i \log(x_{t-1,i}^2 + \Delta x_i^2)$ beneficially exhibits both of the desired properties that stem from its ‘convex-concave’ nature. Fig.1(a) and Fig.1(b) visualize the 1D ($M = 1$) and 2D ($M = 2$) cases of the penalty $f_k(\Delta \mathbf{x})$ respectively. It is \mathbf{x}_{t-1} that determines the transition point of $f_k(\Delta \mathbf{x})$ with regard to the signal innovation $\Delta \mathbf{x}$. As a result, in the blue solid curve region where the signal variations are smaller than \mathbf{x}_{t-1} , $f_k(\Delta \mathbf{x})$ leads to a convex penalty that encourages a non-sparse solution of $\Delta \mathbf{x}$. In contrast, larger innovations than \mathbf{x}_{t-1} inducing the concave penalty have the sparsity-enforcing effect as the red dashed region shows. Small non-sparse variations hence are driven to take place on the past support ($0 < |\Delta x_i| < |x_{t-1,i}|$) as the blue region indicates, rather than triggering support changes. Support additions ($|\Delta x_i| > |x_{t-1,i}| = 0$) and removals ($|\Delta x_i| = |x_{t-1,i}| > 0$) are therefore promoted to be sparse as the red dashed concave curve indicates. In this way, DR ℓ_2 takes advantage of the previous estimations to transition from a convex penalty region to a concave penalty to avoid false support innovations.

In order to obtain an elementary perspective of what $g_{dr}(\Delta \mathbf{x})$ is suggesting, we remove the i^{th} column ϕ_i from Φ and define $\tilde{\Phi} \in \mathbb{R}^{N \times (M-1)}$ as the dictionary consisting of the remaining columns from Φ , while $\tilde{\Gamma} \in \mathbb{R}^{(M-1) \times (M-1)}$ is a diagonal matrix excluding the i^{th} row and i^{th} column in Γ . Substituting $\mathbf{G} = \sigma^2 \mathbf{I} + \tilde{\Phi} \tilde{\Gamma} \tilde{\Phi}^T$, $c = [\phi_i^T \mathbf{G}^{-1} \phi_i]^{-1}$, $r(\Delta \mathbf{x}) = \mathbf{x}_{t-1}^T \Gamma^{-1} \mathbf{x}_{t-1} + \Delta \mathbf{x}^T \Gamma^{-1} \Delta \mathbf{x}$, we yield:

$$\begin{aligned} g_{dr}(\Delta \mathbf{x}) &\leq \mathbf{x}_{t-1}^T \Gamma^{-1} \mathbf{x}_{t-1} + \Delta \mathbf{x}^T \Gamma^{-1} \Delta \mathbf{x} + \log |\Sigma_{\Delta y}| \\ &= r(\Delta \mathbf{x}) + \log \left| \sigma^2 \mathbf{I} + \tilde{\Phi} \tilde{\Gamma} \tilde{\Phi}^T + \gamma_i \phi_i \phi_i^T \right| \\ &= r(\Delta \mathbf{x}) + \log |\mathbf{G}| + \log \left| 1 + \gamma_i \phi_i^T \mathbf{G}^{-1} \phi_i \right| \\ &= r(\Delta \mathbf{x}) + \log |\mathbf{G}| + \log |c^{-1}| + \log |\gamma_i + c| \\ &= \frac{x_{t-1,i}^2 + \Delta x_i^2}{\gamma_i} + \log |\gamma_i + c| + p, \end{aligned} \quad (20)$$

where $p = \log |\mathbf{G}| + \log |c^{-1}| + \sum_{j \neq i} \frac{x_{t-1,j}^2 + \Delta x_j^2}{\gamma_j}$ is the term independent from $x_{t-1,i}$ and Δx_i . The entrywise optimization cost with regard to Δx_i becomes $f_{ic}(\Delta x_i) = \frac{x_{t-1,i}^2 + \Delta x_i^2}{\gamma_i} + \log |\gamma_i + c|$. Notice $c = (\phi_i^T \mathbf{G}^{-1} \phi_i)^{-1}$ actually indicates how closely the i^{th} column ϕ_i is correlated to the remaining atoms $\tilde{\Phi}$. In an extreme scenario that $\sigma^2 = 0$ and the i^{th} atom is orthogonal to any other columns in $\tilde{\Phi}$, c vanishes and the entrywise $f_{ic}(\Delta x_i)$ is equivalent to the 1D case of $f_k(\Delta \mathbf{x})$ with an identical update rule of $\text{DR}\ell_2$. In this way, the estimation of each Δx_i will benefit from the same characteristic described earlier, namely either to trigger sparse support change or merely non-sparse small variation given $x_{t-1,i}$. If c is large due to any significant correlation with the remaining columns, the transition point drifts away from $x_{t-1,i}$. Conclusively, the more a dictionary is behaving like an orthogonal basis, the more aggressively $\text{DR}\ell_2$ can distinguish relevant columns with lower noise.

5. SIMULATION RESULTS

We apply $\text{DR}\ell_2$ to reconstruct the synthetic dynamic signal generated from $\mathcal{N}(0, 1)$ having length $M = 100$ and sparsity $K = 7$ (number of nonzero entries in \mathbf{x}_{t-1}). In both of the experiments, the signal to noise ratio (SNR) is 17dB. In order to evaluate robustness against support changes, we demonstrate the algorithm performance where the elapsed time is only one time stamp. However, tracking multiple time stamps gives similar results that are not shown due to limited space. The sampling matrix Φ remains a constant Gaussian matrix throughout the experiments. In both of the experiments, we randomly remove 2 existing supports from the \mathbf{x}_{t-1} while ΔK support additions are randomly generated from an amplitude distribution $\mathcal{N}(0, 1)$. In addition, we impose random amplitude variations produced from Gaussian distribution $\mathcal{N}(0, 0.5)$ to be added to the remaining support in order to model the slow dynamic process. In this way, the innovation vector $\Delta \mathbf{x}$ is of sparsity $K_{inn} = \Delta K + (K - 2)$ which is a challenging case. The reconstruction performance with regard to the compression ratio N/M is also presented, where N is the number of measurements. We conduct 500 trials for each experiment. The MSE (Mean Square Error) adopts the convention $\|\mathbf{x}_t - \tilde{\mathbf{x}}_t\|_2^2 / \|\mathbf{x}_t\|_2^2$, where \mathbf{x}_t indicates the true signal and $\tilde{\mathbf{x}}_t$ represents the reconstructed signal. We use RSS to denote the rate of successful support reconstruction, computed by $K_r / [500(K - 2 + \Delta K)]$ where K_r is the number of correct supports obtained from $\tilde{\mathbf{x}}_t$ in 500 trials.

The algorithms compared include state-of-the-art dynamic and reweighted algorithms: Modified CS (ModCS) [9], Regularized modified CS (RegMod) [8], Block SBL-EM (BSBL) [7] [14], the Reweighted l2 SBL ($\text{SBL}\ell_2$) [13] and Reweighted l1 ($\text{Re}\ell_1$) [12]. Notice that the previous estimate \mathbf{x}_{t-1} is input to all the algorithms that require it.

In the first experiment, $N = 25$ is fixed and the number

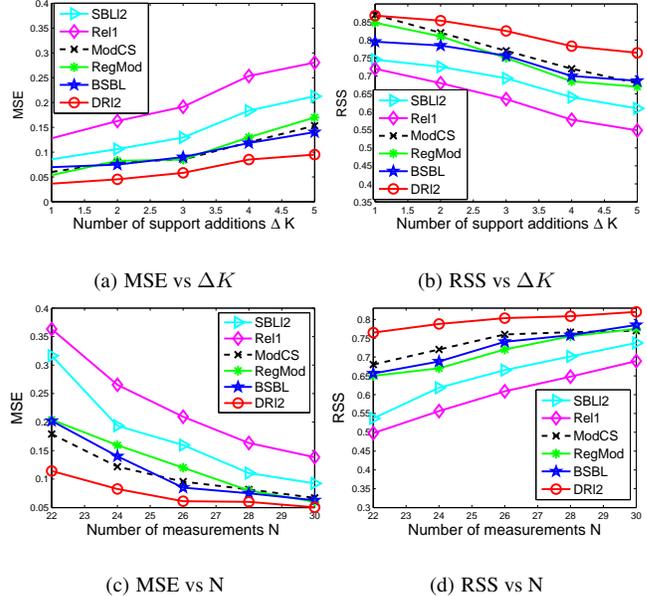


Fig. 2. Performance evaluated by mean square error (MSE) and the rate of successful support detection (RSS)

of support additions ΔK varies from 1 through 5 to investigate robustness against unexpected support innovations. As Fig.2(a) and Fig.2(b) illustrate, $\text{DR}\ell_2$ always offers the best reconstruction against signal variations. In contrast, although benefiting from prior knowledge obtained from x_{t-1} as $\text{DR}\ell_2$ does, ModCS and RegMod are much vulnerable to support changes since both of the algorithms heavily count on the old estimates without adaptation to adjust the bias. BSBL-EM fails due to the strong block sparsity assumed that produces estimation artifacts caused by the noise. $\text{SBL}\ell_2$ exhibits the advantage of the non-separable penalty, but is completely independent from past information and hence performs poorly with small N/M .

The second experiment evaluates algorithm performance at low sampling rates. N ranges from 22 to 30, indicating the compression ratios from 0.22 to 0.3. As Fig.2(c) and Fig.2(d) illustrate, $\text{DR}\ell_2$ not only remains robust against noise and the varying support, but also leverages the reconstruction accuracy at lower sampling rates.

6. CONCLUSION

In this paper, we propose the novel dynamic CS algorithm $\text{DR}\ell_2$. We reveal the ‘convex-concave’ feature inherent in the penalty, which benefits the reconstructions. $\text{DR}\ell_2$ is not only robust against significant signal variations even in noisy cases, but also achieves more accurate reconstruction at lower sampling rates in comparison with state-of-the-art CS algorithms.

7. REFERENCES

- [1] E. J. Candès, “Compressive sampling,” in *Proceedings of the International Congress of Mathematicians*, Madrid, Spain, 2006.
- [2] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, “Sparse solutions to linear inverse problems with multiple measurement vectors,” *Signal Processing, IEEE Transactions on*, vol. 53, no. 7, 2005.
- [3] W. Chen and I. Wassell, “Energy-efficient signal acquisition in wireless sensor networks: a compressive sensing framework,” *Wireless Sensor Systems, IET*, vol. 2, no. 1, 2012.
- [4] M. E. Tipping, “Sparse bayesian learning and the relevance vector machine,” *J. Mach. Learn. Res.*, vol. 1, Sep. 2001.
- [5] D. Wipf and B. Rao, “Sparse bayesian learning for basis selection,” *Signal Processing, IEEE Transactions on*, vol. 52, no. 8, 2004.
- [6] D. Wipf, B. Rao, and S. Nagarajan, “Latent variable bayesian models for promoting sparsity,” *Information Theory, IEEE Transactions on*, vol. 57, no. 9, 2011.
- [7] Z. Zhang and B. Rao, “Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation,” *Signal Processing, IEEE Transactions on*, vol. 61, no. 8, 2013.
- [8] W. Lu and N. Vaswani, “Exact reconstruction conditions for regularized modified basis pursuit,” *Signal Processing, IEEE Transactions on*, vol. 60, no. 5, 2012.
- [9] N. Vaswani and W. Lu, “Modified-cs: Modifying compressive sensing for problems with partially known support,” *Signal Processing, IEEE Transactions on*, vol. 58, no. 9, pp. 4595–4607, 2010.
- [10] E. Karsenas, K. Leung, and W. Dai, “Tracking dynamic sparse signals using hierarchical bayesian kalman filters,” in *Acoustics, Speech, and Signal Processing (ICASSP), 2013 IEEE International Conference on*, May 2013.
- [11] S. Boyd and L. Vandenberghe, *Convex Optimization*. New York, NY, USA: Cambridge University Press, 2004.
- [12] E. J. Candès, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted ℓ_1 minimization,” *Journal of Fourier Analysis and Applications*, 14(5):877-905, 2007.
- [13] D. Wipf and S. Nagarajan, “Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, Apr 2010.
- [14] Z. Zhang and B. Rao, “Sparse signal recovery with temporally correlated source vectors using sparse bayesian learning,” *Selected Topics in Signal Processing, IEEE Journal of*, vol. 5, no. 5, Sept 2011.