# SPARSE RECONSTRUCTION BY SEPARABLE APPROXIMATION

*Stephen J. Wright* [(a)], *Robert D. Nowak* [(b)]

[(a)] Computer Sciences Department
[(b)] Dept. of Electrical and Computer Eng.
University of Wisconsin-Madison
Madison, WI, USA

*Mário A. T. Figueiredo*

Instituto de Telecomunicações
Dept. of Electrical and Computer Eng.
Instituto Superior Técnico
Lisboa, Portugal

**ABSTRACT**

Finding sparse approximate solutions to large underdetermined linear systems of equations is a common problem in signal/image processing and statistics. Basis pursuit, the least absolute shrinkage and selection operator (LASSO), wavelet-based deconvolution and reconstruction, and compressed sensing (CS) are a few well-known areas in which problems of this type appear. One standard approach is to minimize an objective function that includes a quadratic ($\ell_2$) error term added to a sparsity-inducing (usually $\ell_1$) regularizer. We present an algorithmic framework for the more general problem of minimizing the sum of a smooth convex function and a nonsmooth, possibly nonconvex, sparsity-inducing function. We propose iterative methods in which each step is an optimization subproblem involving a separable quadratic term (diagonal Hessian) plus the original sparsity-inducing term. Our approach is suitable for cases in which this subproblem can be solved much more rapidly than the original problem. In addition to solving the standard $\ell_2 - \ell_1$ case, our approach handles other problems, *e.g.*, $\ell_p$ regularizers with $p \neq 1$, or group-separable (GS) regularizers. Experiments with CS problems show that our approach provides state-of-the-art speed for the standard $\ell_2 - \ell_1$ problem, and is also efficient on problems with GS regularizers.

***Index Terms***— sparse approximation, compressed sensing, optimization, reconstruction.

## 1. INTRODUCTION

### 1.1. Problem Formulation

There is growing interest in finding fast algorithms for solving the convex unconstrained optimization problem

$$\min_{\mathbf{x}\in\mathbb{R}^n} \ \frac{1}{2}\|\mathbf{y}-\mathbf{A}\mathbf{x}\|_2^2 + \tau\|\mathbf{x}\|_1, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^k$, $\mathbf{A} \in \mathbb{R}^{k\times n}$ (usually $k < n$) and $\tau \in \mathbb{R}^+$. Problems of the form (1) can be used to identify a sparse approximate solution to the underdetermined system $\mathbf{y} = \mathbf{A}\mathbf{x}$, and have become familiar over the past three decades, particularly in signal processing. Several algorithms have been proposed for solving (1) and its variants; see [15] for a recent overview of the work in this domain.

In this paper we propose algorithms for solving the following generalization of the problem (1):

$$\min_{\mathbf{x}} \ \phi(\mathbf{x}) := f(\mathbf{x}) + \tau\, c(\mathbf{x}), \tag{2}$$

where $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth and convex function, and $c : \mathbb{R}^n \to \mathbb{R}$, usually called the *regularizer* or *regularization term*, is finite for all $\mathbf{x} \in \mathbb{R}^n$, but not necessarily smooth nor convex. We assume also for much of the discussion that $c$ is *separable*, that is,

$$c(\mathbf{x}) = \sum_{i=1}^{n} c_i(x_i). \tag{3}$$

We also consider *group* (or *block*) *separability*, characterized by

$$c(\mathbf{x}) = \sum_{i=1}^{m} c_i(\mathbf{x}_{[i]}), \tag{4}$$

where $\mathbf{x}_{[1]}, \mathbf{x}_{[2]}, \ldots, \mathbf{x}_{[m]}$ are $m$ disjoint sub-vectors of $\mathbf{x}$. We are especially interested in cases in which $\nabla f(\mathbf{x})$ is inexpensive to compute, relative to the cost of computing/storing the Hessian of $f$.

This paper presents an approach to solving problems of the form (2) that has two desirable properties: a) it is computationally competitive with the state-of-the-art algorithms designed for the standard $\ell_2-\ell_1$ problem (1); b) it is versatile enough to handle a broad class of generalizations of (1), such as problems in which the $\ell_1$ regularizer is replaced with an $\ell_p$-norm or with a group-separable regularizer.

### 1.2. Proposed Approach

Our approach generates a sequence of iterates $\mathbf{x}^k$, $k = 1, 2, \ldots$ by solving separable subproblems of the following form:

$$\mathbf{x}^{k+1} \in \arg\min_{\mathbf{z}} \ (\mathbf{z}-\mathbf{x}^k)^T\nabla f(\mathbf{x}^k) + \frac{\alpha_k}{2}\|\mathbf{z}-\mathbf{x}^k\|_2^2 + \tau\, c(\mathbf{z}), \tag{5}$$

where $\alpha_k \in \mathbb{R}^+$. We refer to this approach as SpaRSA (for **Spa**rse **R**econstruction by **S**eparable **A**pproximation).

Different variants of the approach are distinguished by different choices of $\alpha_k$. We focus on variants based on the formula proposed by Barzilai and Borwein (BB) [1] in the context of smooth nonlinear minimization; see also [8, 16]. BB methods have also been applied to constrained problems [2], especially bound-constrained quadratic programs [7, 15, 22]. To our knowledge, BB methods have not been previously used for problems involving nonsmooth terms, though this usage is a natural extension of the basic idea. We also consider monotone variants, in which $\alpha_k$ is increased as necessary to force a decrease in the objective function at every step.

### 1.3. Related Work

SpaRSA is closely related to *iterative shrinkage/thresholding* (IST) (a.k.a. *iterative denoising*, *thresholded Landweber*, *forward-backward splitting*) algorithms [6, 9, 11, 13, 14, 17]. The form of the subproblem is the same, but IST methods use a more conservative

choice of $\alpha_k$. In fact, it can be argued that SpaRSA is a speeded-up IST with better performance resulting from variation of $\alpha_k$.

SpaRSA is also related to the GPSR (*gradient projection for sparse reconstruction*) method recently presented by the authors of this manuscript [15]. While matching the speed of GPSR on the $\ell_2 - \ell_1$ case, SpaRSA can be generalized beyond that case.

## 2. THE PROPOSED APPROACH

### 2.1. The SpaRSA Framework

The SpaRSA framework for problem (2) is as follows.

**Algorithm** *SpaRSA*
1. choose factor $\eta > 1$ and constants $\alpha_{\min}$, $\alpha_{\max}$ ($0 < \alpha_{\min} < \alpha_{\max}$);
2. set iteration counter $k \leftarrow 0$;
3. choose initial guess $\mathbf{x}^0$;
4. **repeat**
5. choose $\alpha_k \in [\alpha_{\min}, \alpha_{\max}]$;
6. **repeat**
7. $\mathbf{x}^{k+1} \leftarrow$ solution of sub-problem (5);
8. $\alpha_k \leftarrow \eta\, \alpha_k$;
9. **until** $\mathbf{x}^{k+1}$ satisfies an acceptance criterion
10. $k \leftarrow k + 1$;
11. **until** stopping criterion is satisfied.

The several variants of SpaRSA are defined by two key steps of the algorithm: the choice of $\alpha_k$ (line 5) and the acceptance criterion (line 9). It is worth noting here that IST algorithms belong to the SpaRSA class. If $c$ is convex, if the acceptance criterion accepts any $\mathbf{x}^{k+1}$, and if we use a constant $\alpha_k$ satisfying the conditions given, *e.g.*, in [6], we have a convergent IST algorithm. SpaRSA allows less conservative choices of $\alpha_k$, often leading to faster convergence.

### 2.2. Solving the Subproblems

By dropping irrelevant additive terms independent of $\mathbf{z}$, the subproblem (5) at line 7 of the algorithm can be rewritten as

$$\mathbf{x}^{k+1} \in \arg\min_{\mathbf{z}} \ \frac{1}{2} \left\| \mathbf{z} - \mathbf{u}^k \right\|_2^2 + \frac{\tau}{\alpha_k}\, c(\mathbf{z}), \tag{6}$$

where $\mathbf{u}^k = \mathbf{x}^k - \nabla f(\mathbf{x}^k)/\alpha_k$. Since the term $\|\mathbf{z} - \mathbf{u}^k\|_2^2$ is a strictly convex function of $\mathbf{z}$, (6) has a unique solution when $c$ is convex. (For nonconvex $c$, there may exist several local minimizers.) In signal processing terms, (6) is called a denoising problem [13].

If $c$ has the separable form (3), the subproblem (6) is also separable and can be written as

$$x_i^{k+1} \in \arg\min_{z} \ \frac{(z - u_i^k)^2}{2} + \frac{\tau}{\alpha_k}\, c_i(z), \quad i = 1, 2, \ldots, n. \tag{7}$$

Separability is key to the efficiency of SpaRSA and IST algorithms. For some choices of $c_i$, the minimization in (7) has a unique closed form solution. When $c(\mathbf{z}) = \|\mathbf{z}\|_1$ (thus $c_i(z) = |z|$), we have

$$\arg\min_{z} \ \frac{(z - u_i^k)^2}{2} + \frac{\tau\,|z|}{\alpha_k} \ = \text{soft}\left(u_i^k, \frac{\tau}{\alpha_k}\right), \tag{8}$$

where $\text{soft}(u, a) \equiv \text{sign}(u) \max\{|u| - a, 0\}$ is the well-known soft-threshold function.

Another notable case is the so-called $\ell_0$ quasi-norm $c(\mathbf{z}) = \|\mathbf{z}\|_0 = \sum_i 1_{x_i \neq 0}$. In this case, we have

$$\arg\min_{z} \ \frac{(z - u_i^k)^2}{2} + \frac{\tau}{\alpha_k}\, 1_{x_i \neq 0} = \text{hard}\left(u_i^k, \sqrt{\frac{2\,\tau}{\alpha_k}}\right), \tag{9}$$

where $\text{hard}(u, a) \equiv u\, 1_{|u| > a}$ is the hard-threshold function.

When $c_i(z) = |z|^p$, *i.e.*, $c(\mathbf{z}) = \|\mathbf{z}\|_p^p$, the closed form solution of (7) is known for $p = 1$ (see (8)), $p = 4/3$, $p = 3/2$, and $p = 2$. See [5, 6], for further details and theory about problems (6) and (7).

### 2.3. Choosing $\alpha_k$: The Barzilai-Borwein Method.

In the most basic variant of the Barzilai-Borwein (BB) approach, we choose $\alpha_k$ such that $\alpha_k\, \mathbf{I}$ mimics the true Hessian $\nabla^2 f(\mathbf{x})$ over the most recent step. Defining

$$\mathbf{s}^k = \mathbf{x}^k - \mathbf{x}^{k-1}, \ \text{ and } \ \mathbf{r}^k = \nabla f(\mathbf{x}^k) - \nabla f(\mathbf{x}^{k-1}),$$

we require that $\alpha_k\, \mathbf{s}^k \approx \mathbf{r}^k$ in the least-squares sense, leading to

$$\alpha_k = \arg\min_{\alpha} \|\alpha\, \mathbf{s}^k - \mathbf{r}^k\|_2^2 = (\mathbf{s}^k)^T \mathbf{r}^k / [(\mathbf{s}^k)^T \mathbf{s}^k]. \tag{10}$$

When $f(\mathbf{x}) = (1/2)\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$, the previous expression becomes $\alpha_k = \|\mathbf{A}\, \mathbf{s}^k\|_2^2 / \|\mathbf{s}^k\|_2^2$. These formulas can be safeguarded appropriately to ensure that $\alpha_k$ remains in the range $[\alpha_{\min}, \alpha_{\max}]$.

### 2.4. The Acceptance Criterion

In the simplest variant of the SpaRSA scheme, the acceptance criterion is trivial: accept whatever $\mathbf{z}$ solves the subproblem (5) as the new iterate $\mathbf{x}^{k+1}$, even if it yields an increase in the objective function $\phi$. We consider also a variant in which $\alpha_k$ is viewed as a damping parameter in the subproblem (6), which is increased until the solution of this subproblem yields a decrease in $\phi$. In this scheme, the acceptance criterion may be $\phi(\mathbf{x}^{k+1}) < \phi(\mathbf{x}^k)$, or we may enforce a more stringent variant that requires the margin of decrease to be at least some (positive constant) multiple of the decrease promised by the subproblem (5). The initial choice of $\alpha_k$ can be given by (10), or by modifying the value $\alpha_{k-1}$ from the previous iteration. We call the former variant of the algorithm SpaRSA-monotone.

The existence of a value of $\alpha_k$ sufficiently large to ensure a decrease in the objective at each iteration can be inferred from the connection between (6) and the following trust-region subproblem:

$$\min_{\mathbf{z}} \ \nabla f(\mathbf{x}^k)^T (\mathbf{z} - \mathbf{x}^k) + \tau c(\mathbf{z}) \ \text{ subject to } \ \|\mathbf{z} - \mathbf{x}^k\|_2 \leq \Delta_k.$$

It also follows from the known fact, which underlies the monotonicity of IST algorithms [14], that there is a constant $\bar{\alpha} > 0$ such that descent is assured whenever $\alpha_k \geq \bar{\alpha}$.

### 2.5. Warm Starting and Continuation

The SpaRSA approach benefits from a good starting point $\mathbf{x}^0$, which suggests that we can use the solution of (2), for a given value of $\tau$, to initialize SpaRSA in solving (2) for a nearby value of $\tau$. The second run will typically be significantly faster than the first one. An important application of warm-starting is *continuation*, as recently suggested in [17]. The speed of SpaRSA algorithms may degrade considerably for smaller values of the regularization parameter $\tau$. However, if we use SpaRSA to solve (2) for a larger value of $\tau$, then decrease $\tau$ in steps toward its desired value, running SpaRSA with warm-start for each successive value of $\tau$, we are often able to identify the solution much more efficiently than if we just ran SpaRSA once for the desired value of $\tau$ from a "cold start."

## 3. GROUP-SEPARABLE REGULARIZERS

In this section we consider group-separable (GS) regularizers of the form (4). In this case, the minimization (6), instead of decoupling into a set of one-dimensional minimizations (7), decouples into a set of $m$ independent multi-dimensional minimizations, of the form

$$\min_{\mathbf{w}\in\mathbb{R}^l} \ \frac{1}{2}\|\mathbf{w}-\mathbf{b}\|_2^2 + \beta\,\Phi(\mathbf{w}), \tag{11}$$

where $l$ is the dimension of $\mathbf{x}_{[i]}$, $\mathbf{b}=\mathbf{u}_{[i]}^k$, $\Phi=c_i$, and $\beta=\tau/\alpha_k$.

GS regularizers are desirable when there exists a group structure in $\mathbf{x}$, which arises naturally in many applications.

- In brain imaging, the voxels associated with different functional regions (e.g., motor or visual cortices) may be grouped together in order to identify a sparse set of regional events. In [3, 4], an EM algorithm (equivalent to IST) was proposed for solving problems of this type.
- A GS-$\ell_2$ penalty ($\Phi(\mathbf{w})=c_i(\mathbf{w})=\|\mathbf{w}\|_2$) was proposed for source localization in sensor arrays [20]; second-order cone programming was used to solve the optimization problem.
- In gene expression analysis, some genes are organized in functional groups. This has motivated an approach called CAP (composite absolute penalty) [25], which has the form (4), and uses a greedy optimization scheme [26].

GS regularizers have also been proposed for ANOVA regression models [19, 21, 24], and Newton-type optimization methods have been proposed in that context. An interior-point method for the GS-$\ell_\infty$ case ($\Phi(\mathbf{w})=c_i(\mathbf{w})=\|\mathbf{w}\|_\infty$) was proposed in [23]. The SpaRSA framework is versatile enough to handle the GS regularizes arising all in the applications described above.

As in [5, 6], convex analysis can be used to obtain the solution of (11). If $\Phi$ is a norm, it is proper, convex (maybe not strictly so), and homogenous. Since the quadratic term in (11) is proper and strictly convex, this problem has a unique solution, which can be written explicitly as follows:

$$w=\mathbf{b}-P_{\beta C_\Phi}(\mathbf{b}), \tag{12}$$

where $P_B$ denotes the orthogonal projector onto set $B$, and $C_\Phi$ is a 1-ball in the dual norm $\Phi^\star$, that is, $C_\Phi=\{\mathbf{w}\in\mathbb{R}^l:\Phi^\star(\mathbf{w})\le 1\}$.

For $\Phi(\mathbf{w})=\|\mathbf{w}\|_2$, the dual norm is also $\Phi^\star(\mathbf{w})=\|\mathbf{w}\|_2$, thus $\beta C_{\|\cdot\|_2}=\{\mathbf{w}\in\mathbb{R}^l:\|\mathbf{w}\|_2\le\beta\}$. Clearly, if $\|\mathbf{b}\|_2\le\beta$, then $P_{\beta C_{\|\cdot\|_2}}(\mathbf{b})=\mathbf{b}$, thus $\mathbf{b}-P_{\beta C_{\|\cdot\|_2}}(\mathbf{b})=0$. If $\|\mathbf{b}\|_2>\beta$, then $P_{\beta C_{\|\cdot\|_2}}(\mathbf{b})=\beta\,\mathbf{b}/\|\mathbf{b}\|_2$. These two cases are written compactly as

$$w=\frac{\mathbf{b}}{\|\mathbf{b}\|_2}\,\max\{\|\mathbf{b}\|_2-\beta,0\}\,. \tag{13}$$

Naturally, if $l=1$, (13) reduces to the scalar soft-threshold (8).

For $\Phi(\mathbf{w})=\|\mathbf{w}\|_\infty$, the dual norm is $\Phi^\star(\mathbf{w})=\|\mathbf{w}\|_1$, thus $\beta C_{\|\cdot\|_\infty}=\{\mathbf{w}\in\mathbb{R}^n:\|\mathbf{w}\|_1\le\beta\}$. In this case, the solution of (11) is the residual of the orthogonal projection of $\mathbf{b}$ onto the $\ell_1$ $\beta$-ball. This projection (thus also the residual) can be computed with $O(l\log l)$ cost, as recently shown in [3, 4, 10].

## 4. EXPERIMENTS

### 4.1. Speed Comparisons for the $\ell_2-\ell_1$ Problem

The purpose of our first experiment is to compare SpaRSA with the state-of-the-art algorithms IST and GPSR (see Subsection 1.3), and the *l1_ls* method [18], in a typical CS scenario (as in [15, 18]): $f(\mathbf{x})=\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2$, with $\mathbf{A}$ a $2^{10}\times 2^{12}$ random matrix; $\mathbf{y}$ is generated as $\mathbf{y}=\mathbf{A}\mathbf{x}_{\text{true}}+\mathbf{e}$, where $\mathbf{e}$ is a Gaussian white vector with variance $10^{-4}$, and $\mathbf{x}_{\text{true}}$ is a vector with 160 randomly placed $\pm1$ spikes and zeros elsewhere. We use the $\ell_1$ regularizer $c(\mathbf{x})=\|\mathbf{x}\|_1$, and $\tau=0.1\,\|\mathbf{A}^T\mathbf{y}\|_\infty$, as in [15, 18]. In this (and all other) experiments, $\alpha_{\text{max}}=1/\alpha_{\text{min}}=10^{30}$ and $\eta=2$ (for SpaRSA-monotone). To perform the comparison, independently of the adopted stopping rule, we first run *l1_ls* and then the other algorithms until each reaches the same value of the objective function reached by *l1_ls*. Table 1 reports the CPU times required by SpaRSA, two variants of GPSR, *l1_ls*, and IST, as well as the final mean squared error (MSE) of the reconstructions with respect to $\mathbf{x}_{\text{true}}$. These results show that, for this $\ell_2-\ell_1$ problem, SpaRSA is slightly faster than GPSR and clearly faster than *l1_ls* and IST, while achieving a similar value of MSE.

**Table 1**. CPU times (average over 10 runs) of several algorithms on the CS experiment described in the text.

| Algorithm | CPU time (secs.) | MSE |
|---|---|---|
| SpaRSA | 0.44 | 2.42e-3 |
| SpaRSA-monotone | 0.45 | 2.49e-3 |
| GPSR-BB | 0.55 | 2.81e-3 |
| GPSR-Basic | 0.69 | 2.59e-3 |
| *l1_ls* | 6.56 | 2.51e-3 |
| IST | 2.76 | 2.51e-3 |

An indirect comparison with other codes can be made via [18, Table 1], which shows that *l1_ls* outperforms the method from [12] (6.9 vs 11.3 secs.), as well as *$\ell_1$-magic* by about two orders of magnitude and `pdco` from *SparseLab* by about one order of magnitude.

The second experiment assesses how the computational cost of SpaRSA grows with the size of matrix $\mathbf{A}$, using a setup similar to the one in [15, 18]. Assuming that the computational cost is $O(n^\gamma)$, we obtain empirical estimates of $\gamma$. SpaRSA and SpaRSA-monotone have empirical exponents of .88 and .87, respectively, similar to the values .86 and .87 of GPSR and GPSR-Basic. IST has a similar exponent .89, but a worse constant. For *l1_ls*, we found $\gamma=1.21$, in agreement with the value 1.2 reported in [18].

### 4.2. Group-Separable Regularizers

Here we illustrate the use of SpaRSA with the GS regularizers considered in Section 3. In our example, $\mathbf{x}_{\text{true}}$ is a $2^{12}$-dimensional vector, divided into $m=64$ groups of length $l_i=64$. As above, $\mathbf{A}$ a $2^{10}\times 2^{12}$ random matrix and $\mathbf{y}$ is generated as $\mathbf{y}=\mathbf{A}\mathbf{x}_{\text{true}}+\mathbf{e}$, where $\mathbf{e}$ is Gaussian white noise with variance $10^{-4}$. To generate $\mathbf{x}_{\text{true}}$, we randomly choose 8 groups and fill them with zero-mean Gaussian random samples of unit variance; all other groups are filled with zeros. Finally we run SpaRSA, with $f(\mathbf{x})=\|\mathbf{A}\mathbf{x}-\mathbf{y}\|_2^2$ and $c(\mathbf{x})$ as given by (4), where $c_i(\mathbf{x}_{[i]})=\|\mathbf{x}_{[i]}\|_2$. The value of $\tau$ is hand-tuned for optimal performance. Fig. 1 shows the result obtained by SpaRSA, based on the GS-$\ell_2$ regularizer, which successfully recoverers the group structure of $\mathbf{x}_{\text{true}}$, as well as the result obtained with the classical $\ell_1$ regularizer, for the best choice of $\tau$.

In the second experiment, we consider a similar scenario, with a single difference. Each active group, instead of being filled with Gaussian random samples, is filled with ones. This case is clearly more adequate for a GS-$\ell_\infty$ regularizer, as illustrated in Fig. 2, which achieves an almost perfect reconstruction, with an MSE 2 orders of magnitude smaller than what is obtained with a GS-$\ell_2$ regularizer.

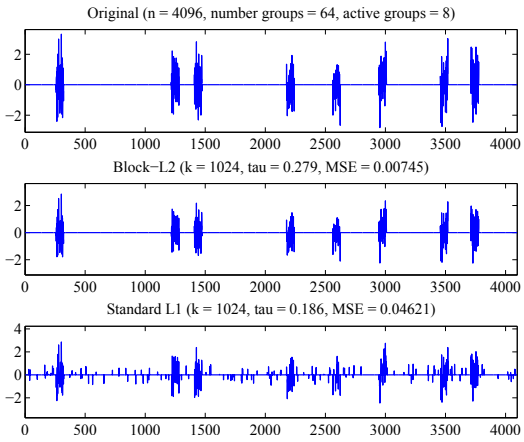


**Fig. 1**. Comparison of GS-$\ell_2$ regularizer with conventional $\ell_1$ regularizer. Exploiting known group structure provides a dramatic gain.

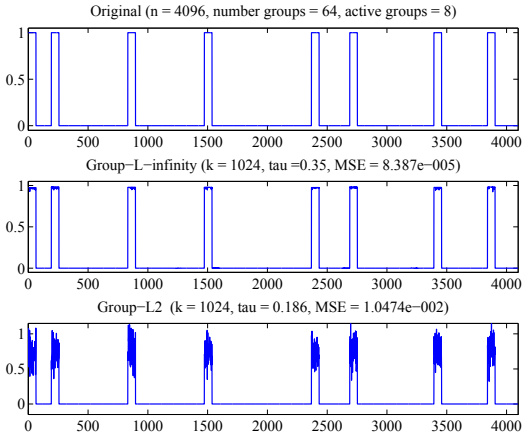


**Fig. 2**. Comparison of GS-$\ell_2$ and GS-$\ell_\infty$ regularizers. Signals with uniform behavior within groups benefit from the GS-$\ell_\infty$ regularizer.

## 5. CONCLUDING REMARKS

In this paper, we have introduced the SpaRSA algorithmic framework for solving large-scale optimization problems involving the sum of a smooth error term and a possibly nonsmooth regularizer. We give experimental evidence that SpaRSA matches the speed of the state-of-the-art method when applied to the $\ell_2 - \ell_1$ problem, and show that SpaRSA can be generalized to other regularizers such as those with group-separable structure. Ongoing work includes a more thorough experimental evaluation involving wider classes of regularizers, and theoretical analysis of the convergence properties.

## 6. REFERENCES


[1] J. Barzilai, J. Borwein, "Two point step size gradient methods," *IMA Journal of Numerical Analysis,* vol. 8, pp. 141–148, 1988.

[2] E. Birgin, J. Martinez, M. Raydan, "Nonmonotone spectral projected gradient methods on convex sets," *SIAM Journal on Optimization,* vol. 10, pp. 1196-1211, 2000.

[3] A. Bolstad, B. Van Veen, R. Nowak, "Space-time sparsity regularization for the magnetoencephalography inverse problem", *Proc. of the IEEE Itern. Conference on Biomedical Imaging*, Arlington, VA, 2007.

[4] A. Bolstad, B. Van Veen, R. Nowak, R. Wakai, "An expectation-maximization algorithm for space-time sparsity regularization of the MEG inverse problem", *Proc. of the Itern. Conference on Biomagnetism*, Vancouver, BC, Canada, 2006.

[5] C. Chaux, P. Combettes, J.-C. Pesquet, V. Wajs, "A variational formulation for frame-based inverse problems," *Inverse Problems*, vol. 23, pp. 1495-1518, 2007.

[6] P. Combettes, V. Wajs, "Signal recovery by proximal forward-backward splitting," *SIAM Journal on Multiscale Modeling & Simulation*, vol. 4, pp. 1168–1200, 2005.

[7] Y.-H. Dai, R. Fletcher. "Projected Barzilai-Borwein methods for large-scale box-constrained quadratic programming," *Numerische Mathematik*, vol. 100, pp. 21–47, 2005.

[8] Y.-H. Dai, W. Hager, K. Schittkowski, H. Zhang, "The cyclic Barzilai-Borwein method for unconstrained optimization," *IMA Journal of Numerical Optimization*, vol. 26, pp. 604-627, 2006.

[9] I. Daubechies, M. Defriese, C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint", *Comm. on Pure and Applied Math.*, vol. LVII, pp. 1413-1457, 2004.

[10] I. Daubechies, M. Fornasier, I. Loris, "Accelerated projected gradient method for linear inverse problems with sparsity constraints." Preprint available at `www.dsp.ece.rice.edu/cs`, 2007.

[11] M. Elad, "Why simple shrinkage is still relevant for redundant representations?", *IEEE Trans. Inf. Th.*, vol. 52, pp. 5559-5569, 2006.

[12] D. Donoho and Y. Tsaig. "Fast solution of L1-norm minimization problems when the solution may be sparse," Technical Report, Institute for Computational and Math. Eng., Stanford University, 2006.

[13] M. Figueiredo, R. Nowak, "An EM algorithm for wavelet-based image restoration," *IEEE Trans. Image Proc.*, vol. 12, pp. 906–916, 2003.

[14] M. Figueiredo, R. Nowak, "A bound optimization approach to wavelet-based image deconvolution," *IEEE Intern. Conf. on Image Processing – ICIP'05*, Genoa, Italy, 2005.

[15] M. Figueiredo, R. Nowak, S. Wright, "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems", *IEEE Journal of Selected Topics in Signal Processing*, 2007 (to appear). Available at `www.lx.it.pt/~mtf/GPSR`

[16] L. Grippo, M. Sciandrone, "Nonmonotone globalization techniques for the Barzilai-Borwein method," *Computational Optimization and Applications,* vol. 32, pp. 143-169, 2002.

[17] T. Hale, W. Yin, Y. Zhang, "A fixed-point continuation method for $\ell_1$-regularized minimization with applications to compressed sensing." TR07-07, Dept. of Comput. and Appl. Math., Rice University, 2007.

[18] S. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinvesky. "A method for large-scale $\ell_1$-regularized least squares problems with applications in signal processing and statistics," Tech. Report, Dept. of Electrical Engineering, Stanford University, 2007.

[19] Y. Kim, J. Kim, Y. Kim, "Blockwise sparse regression", *Statistica Sinica*, vol. 16, 2006.

[20] D. Malioutov, M. Cetin, A. Willsky, "Sparse signal reconstruction perspective for source localization with sensor arrays", *IEEE Trans. Signal Processing*, vol. 53, pp. 3010-3022, 2005.

[21] L. Meier, S. van de Geer, P. Buhlmann, "The group LASSO for logistic regression", Research Report 131, ETH-Zurich, 2006.

[22] T. Serafini, G. Zanghirati, L. Zanni. "Gradient projection methods for large quadratic programs and applications in training support vector machines," *Optim. Meth. and Software*, vol. 20, pp. 353–378, 2004.

[23] B. Turlach, W. N. Venables, and S. J. Wright. "Simultaneous variable selection", *Technometrics*, vol. 27, pp. 349-363, 2005.

[24] M. Yuan, Y. Lin, "Model selection and estimation in regression with grouped variables", *J. Royal Stat. Soc. B*, vol. 68, pp. 49-67, 2006.

[25] P. Zhao, G. Rocha, B. Yu, "Grouped and hierarchical model selection through composite absolute penalties", TR 703, Statistics Department, University of California - Berkeley, 2007.

[26] P. Zhao, B. Yu, "Boosted Lasso", Technical Report, Statistics Department, University of California - Berkeley, 2004.