A NONCONVEX ADMM ALGORITHM FOR GROUP SPARSITY WITH SPARSE GROUPS

Rick Chartrand and Brendt Wohlberg

Los Alamos National Laboratory
Los Alamos, NM 87545, USA

ABSTRACT

We present an efficient algorithm for computing sparse representations whose nonzero coefficients can be divided into groups, few of which are nonzero. In addition to this group sparsity, we further impose that the nonzero groups themselves be sparse. We use a nonconvex optimization approach for this purpose, and use an efficient ADMM algorithm to solve the nonconvex problem. The efficiency comes from using a novel shrinkage operator, one that minimizes nonconvex penalty functions for enforcing sparsity and group sparsity simultaneously. Our numerical experiments show that combining sparsity and group sparsity improves signal reconstruction accuracy compared with either property alone. We also find that using nonconvex optimization significantly improves results in comparison with convex optimization.

Index Terms— Sparse representations, group sparsity, shrinkage, nonconvex optimization, alternating direction method of multipliers

1. INTRODUCTION

Sparse representations have developed into an important tool for signal processing and classification over the past decade. These methods are based on the assumption that signals of interest have a sparse representation, that is, a representation as a linear combination with only a few nonzero entries, of the columns in a predefined (or learned) dictionary matrix. It is often the case, however, that additional structure can be expected in the support of the nonzero elements in such a representation. A *group*-sparse vector can be divided into groups of components such that few groups contain nonzero values, but groups that do are not necessarily sparse. A closely related notion, sometimes called *joint* sparsity, is that of a set of sparse vectors having the union of their supports be sparse. Putting such vectors into a matrix as columns, the matrix will have few nonzero rows, while nonzero rows need not be sparse. These structured sparse representations are typically obtained by replacing the problem

$$\min_{\mathbf{x}} \alpha\|\mathbf{x}\|_1 + \tfrac{1}{2}\|\Phi\mathbf{x} - \mathbf{y}\|_2^2 \tag{1}$$

This research was supported by the U.S. Department of Energy through the LANL/LDRD Program.

with

$$\min_{\mathbf{x}} \alpha \sum_{i=1}^{M} \|\mathbf{x}[i]\|_2 + \tfrac{1}{2}\|\Phi\mathbf{x} - \mathbf{y}\|_2^2, \tag{2}$$

where $\mathbf{x}[i]$ is the ith group of $\mathbf{x}$. A number of theoretical results regarding this type of decomposition have been derived [1–3], with evidence that it can provide improved performance over simple ℓ^1 regularization, particularly in classification tasks [4–7].

In some cases, however, further refinement can be beneficial. While nonzero components may be clustered into groups, the nonzero groups may also be sparse. In such a setting, we still wish to enforce sparsity, while simultaneously encouraging group sparsity. This goal can be achieved by the problem

$$\min_{\mathbf{x}} \alpha\|\mathbf{x}\|_1 + \beta \sum_{i=1}^{M} \|\mathbf{x}[i]\|_2 + \tfrac{1}{2}\|\Phi\mathbf{x} - \mathbf{y}\|_2^2. \tag{3}$$

This optimization problem has been applied in low-dimensional nonlinear signal modeling [8], and has been shown to provide improved performance in some classification [9] and source separation problems [10].

All of these papers used convex optimization to enforce the sparse-and-group-sparse model. Motivated by many previous results in compressive sensing showing that improved performance can be obtained by using nonconvex optimization instead [11–14], in this paper we enforce both sparsity and group sparsity using nonconvex regularization. The former is a slight modification of a regularization derived in [13, 14], which is designed to be minimized using a computationally efficient shrinkage operator. In this work, we use a shrinkage which jointly solves a nonconvex optimization problem for enforcing our model. We present in Sec. 2 the ADMM algorithm that will exploit our shrinkage-based approach, after which our new shrinkage and corresponding penalty function are discussed in Section 3. We show the value of our approach via numerical experiments, presented in Sec. 4.

2. ADMM ALGORITHM

In this section we discuss the alternating direction, method of multipliers (ADMM) approach [15–18], which uses variable splitting to decompose our problem into easily solvable

subproblems. For notational simplicity we present the joint sparsity variant of (3), in which the groups are rows of a coefficient matrix in the *multiple measurement vector* context, though it is easily extended to the group sparsity problem:

$$\min_X \alpha\|X\|_1 + \beta\sum_{i=1}^{M}\|\mathbf{X}^i\|_2 + \tfrac{1}{2}\|\Phi X - Y\|_F^2. \quad (4)$$

The N columns of Y each contain a signal in $\mathbb{R}^L$, for which sparse representations are sought using the $L \times M$ dictionary Φ via coefficient vectors stored as columns in the $M \times N$ matrix X. Here $\|\cdot\|_1$ is the entrywise ℓ^1 norm, $\|\cdot\|_F$ is the Frobenius (entrywise ℓ^2) norm, and the $\mathbf{X}^i$ are rows of X. The first term promotes sparsity, while the second promotes group sparsity in the sense of few rows having nonzero entries.

We introduce an auxiliary variable W for the splitting:

$$\min_{W,X} \alpha\|W\|_1 + \beta\sum_{i=1}^{M}\|\mathbf{W}^i\|_2 + \tfrac{1}{2}\|W - X\|_F^2 + \tfrac{1}{2}\|\Phi X - Y\|_F^2. \quad (5)$$

We can regard W as a proxy for X, and the new third term a relaxation of the equality constraint $W = X$. We proceed by alternately fixing one variable and solving for the other (*i.e.*, alternating directions), and iterating. With fixed W, the X subproblem is quadratic, becoming a simple linear equation:

$$(I + \Phi^T\Phi)X = W + \Phi^T Y. \quad (6)$$

Note that the system matrix remains fixed throughout. With fixed X, we have the following subproblem for W:

$$\min_W \alpha\|W\|_1 + \beta\sum_{i=1}^{M}\|\mathbf{W}^i\|_2 + \tfrac{1}{2}\|W - X\|_F^2. \quad (7)$$

It turns out that this can be solved very easily, by means of *shrinkages*.

Definition 1. *Define shrinkage mappings S_1 and $\mathcal{S}_1$ from $\mathbb{R}^N \times \mathbb{R}_+$ to $\mathbb{R}^N$ by*

$$S_1(\mathbf{x}, \alpha)_i = \frac{x_i}{|x_i|}\max\{0, |x_i| - \alpha\}, \quad (8)$$

$$\mathcal{S}_1(\mathbf{x}, \alpha) = \frac{\mathbf{x}}{\|\mathbf{x}\|_2}\max\{0, \|\mathbf{x}\|_2 - \alpha\}; \quad (9)$$

where both expressions are taken to be zero when the second factor is zero.

The shrinkage (8) is known as *soft thresholding*. Its occurrence in many algorithms related to sparsity is due to it being the proximal mapping for the ℓ^1 norm:

$$\arg\min_{\mathbf{w}} \alpha\|\mathbf{w}\|_1 + \tfrac{1}{2}\|\mathbf{w} - \mathbf{x}\|_2^2 = S_1(\mathbf{x}, \alpha). \quad (10)$$

Proposition 2. *The solution to* (7) *is given row-wise by*

$$\mathbf{W}^i = \mathcal{S}_1(S_1(\mathbf{X}^i, \alpha), \beta). \quad (11)$$

It has previously been noted [8, 10] that (7) has an explicit solution in terms of shrinkages, but the particular expression of (11) is, to the best of our knowledge, new. Prop. 2 is a special case of Thm. 8 below.

The last ingredient is to enforce the equality of W and X at convergence, using the method of multipliers. We introduce a dual variable (or Lagrange multiplier) Λ, which we update at each iteration by adding the residual $X - W$:

$$\min_{W,X} \alpha\|W\|_1 + \beta\sum_{i=1}^{M}\|\mathbf{W}^i\|_2 + \tfrac{1}{2}\|W - X - \Lambda\|_F^2 + \tfrac{1}{2}\|\Phi X - Y\|_F^2. \quad (12)$$

3. SHRINKAGE FOR NONCONVEX MINIMIZATION

We now generalize the approach of the previous section, to bring the benefits of nonconvex optimization to our problem.

Definition 3. *Let $p \in \mathbb{R}$. Define shrinkage mappings S_p and $\mathcal{S}_p$ from $\mathbb{R}^N \times \mathbb{R}_+$ to $\mathbb{R}^N$ by*

$$S_p(\mathbf{x}, \alpha)_i = \frac{x_i}{|x_i|}\max\{0, |x_i| - \alpha^{2-p}|x_i|^{p-1}\}, \quad (13)$$

$$\mathcal{S}_p(\mathbf{x}, \alpha) = \frac{\mathbf{x}}{\|\mathbf{x}\|_2}\max\{0, \|\mathbf{x}\|_2 - \alpha^{2-p}\|\mathbf{x}\|_2^{p-1}\}; \quad (14)$$

where both expressions are taken to be zero when the second factor is zero.

We now seek to generalize (10) to the general case, with the aim of applying it with $p < 1$:

Proposition 4. *Let $p \in \mathbb{R}, \alpha > 0$. Then there is a real-valued function G such that for any $\mathbf{x} \in \mathbb{R}^N$,*

$$\arg\min_{\mathbf{w}} \alpha G(\mathbf{w}) + \tfrac{1}{2}\|\mathbf{w} - \mathbf{x}\|_2^2 = S_p(\mathbf{x}, \alpha), \quad (15)$$

with $G(\mathbf{w}) = \sum_{i=1}^{N} g(w_i)$ for some scalar function g.

The proof is essentially the same as in [14], with only minor changes required due to the different power of α in our shrinkage (13). We emphasize that for $p \leq 1$, the minimizer given by Prop. 4 is unique and global, nonconvexity of G notwithstanding, as $\alpha G(\mathbf{w}) + \|\mathbf{w}\|_2^2/2$ is strictly convex.

Except for special values of p, we are unable to write G explicitly. See Fig. 1 for numerically-computed plots; the function $g(w)$ grows like $|w|^p/p + C$ for large $|w|$ and some C (or $\log|w| + C$ for $p = 0$). However, the point is not to be able to compute G efficiently, it is to solve efficiently optimization problems having G as a penalty function. Our ADMM approach in Sec. 2 will be able to use the shrinkage property of Prop. 4 for this purpose.

We can also prove indirectly several properties of g (and hence G). The proof of the following Lemma follows as in [14, Prop. 3]:

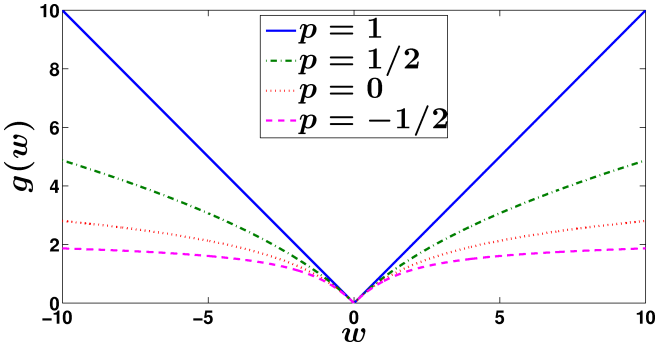

Fig. 1. Plots of the function g of Prop. 4, using $\alpha = 1$. The smaller the value of p, the slower the growth of g, with g being bounded above when $p < 0$.

Lemma 5. *The function g of Prop. 4 is radial, radially increasing, continuous, differentiable except at 0 with $\partial g(0) = [-1, 1]$, concave on $(-\infty, 0)$ and $(0, \infty)$, and satisfies the triangle inequality.*

Note that the subdifferential of the nonconvex function g is in the sense of [19, Def. 8.3.(a)]. We also need the following property:

Lemma 6. *The function G of Prop. 4 is* subdifferentially regular.

See [19, Def. 7.2.5] for a definition of subdifferential regularity. A rough description is G "locally" majorizes its supporting hyperplanes, doing so in an infinitesimal sense at points of differentiability, while at 0 majorizing lines with slopes in $(-1, 1)$ on a neighborhood of 0. The importance of Lemma 6 for us is given by [19, Thm. 10.1]: although G is nonconvex, first-order optimality conditions for our optimization problems involving G will not only be necessary for a local minimizer, they will be sufficient as well.

We now establish the corresponding property for $\mathcal{S}_p$, which is known for $p = 1$ [20, Lemma 3.3]:

Proposition 7. *Let g be as in Prop. 4. Then*

$$\arg\min_{\mathbf{w}} \alpha g(\|\mathbf{w}\|_2) + \tfrac{1}{2}\|\mathbf{w} - \mathbf{x}\|_2^2 = \mathcal{S}_p(\mathbf{x}, \alpha). \tag{16}$$

Proof. Since $\alpha g(\|\mathbf{w}\|_2)$ is radial and increasing in $\|\mathbf{w}\|_2$, the minimizer will have the form $\mathbf{w} = t\mathbf{x}$ for some $t \in [0, 1]$. Our optimization problem becomes

$$\min_{t\in[0,1]} \alpha g(t\|\mathbf{x}\|_2) + \tfrac{1}{2}\|t\mathbf{x} - \mathbf{x}\|_2^2. \tag{17}$$

Since $\|t\mathbf{x} - \mathbf{x}\|_2 = (t-1)\|\mathbf{x}\|_2 = t\|\mathbf{x}\|_2 - \|\mathbf{x}\|_2$, letting $s = t\|\mathbf{x}\|_2 = \|\mathbf{w}\|_2$ we obtain

$$\min_{s\in[0,\|\mathbf{x}\|_2]} \alpha g(s) + \tfrac{1}{2}(s - \|\mathbf{x}\|_2)^2. \tag{18}$$

By Prop. 4 with $\|\mathbf{x}\|_2$ in place of $\mathbf{x}$, the solution is $s = \max\{0, \|\mathbf{x}\|_2 - \alpha^{2-p}\|\mathbf{x}\|_2^{p-1}\}$. Since $\mathbf{w} = s\mathbf{x}/\|\mathbf{x}\|_2$, the result follows. □

The following theorem generalizes Prop. 2 to the case of general p:

Theorem 8. *Let X be an $M \times N$ matrix, and let $\alpha, \beta > 0$. There are functions $G_{\alpha,p}, g_{\beta,q}$ such that a local minimizer of the optimization problem*

$$\min_W \alpha G_{\alpha,p}(W) + \beta \sum_{i=1}^{M} g_{\beta,q}(\|\mathbf{W}^i\|_2) + \tfrac{1}{2}\|W - X\|_F^2 \tag{19}$$

is given row-wise by

$$\mathbf{W}^i = \mathcal{S}_q(S_p(\mathbf{X}^i, \alpha), \beta). \tag{20}$$

Proof. The optimization problem (19) is row-separable, so we consider the case of $\mathbf{x} \in \mathbb{R}^N$, let $\mathbf{w} = \mathcal{S}_q(S_p(\mathbf{x}, \alpha), \beta)$, and show that this is a local minimizer of

$$\min_{\mathbf{w}} \alpha G_{\alpha,p}(\mathbf{w}) + \beta g_{\beta,q}(\|\mathbf{w}\|_2) + \tfrac{1}{2}\|\mathbf{w} - \mathbf{x}\|_2^2. \tag{21}$$

Denote $\mathbf{v} = S_p(\mathbf{x}, \alpha)$. We define $g_{\beta,q}$ as in Prop. 7 with β in place of α and q in place of p.

We consider first the case that $\mathbf{w} = \mathbf{0}$, which is equivalent to $\|\mathbf{v}\|_2 \le \beta$. We define $G_{\alpha,p}$ as in Prop. 4; then $\mathbf{v}$ is the (unique, global) minimizer of the left-side of (15). By Lemma 6, it suffices to show that the first-order optimality conditions hold for (21), in this case

$$\mathbf{0} \in \alpha\partial G_{\alpha,p}(\mathbf{0}) + \beta g_{\beta,q}(0)\partial\|\cdot\|_2(\mathbf{0}) - \mathbf{x}. \tag{22}$$

This is equivalent to the existence of $\mathbf{y} \in [-1, 1]^N$ and $\mathbf{z}$ with $\|\mathbf{z}\|_2 \le 1$ such that $\mathbf{x} = \alpha\mathbf{y} + \beta\mathbf{z}$. Now by first-order optimality and the definition of $\mathbf{v}$ and $G_{\alpha,p}$,

$$0 \in \alpha\partial G_{\alpha,p}(\mathbf{v}) + \mathbf{v} - \mathbf{x}. \tag{23}$$

Recall that $\partial g_{\alpha,p}(0) = [-1, 1]$; by concavity, $|g'_{\alpha,p}(t)| \le 1$ for $t \neq 0$ also. Thus there is $\mathbf{y} \in [-1, 1]^N$ such that $\mathbf{x} = \alpha\mathbf{y} + \mathbf{v}$; since $\|\mathbf{v}\|_2 \le \beta$ by assumption, the proof is complete for the case of $\mathbf{w} = \mathbf{0}$.

Now assume $\mathbf{w} \neq \mathbf{0}$, so $\|\mathbf{v}\|_2 > \beta$. By definition of $\mathcal{S}_q$, we have that $\mathbf{w} = t\mathbf{v}$ for some $t > 0$, namely $t = 1 - \beta^{2-q}\|\mathbf{v}\|_2^{q-2}$. Define $G_{\alpha,p}$ as in Prop. 4 with $t\alpha$ in place of α. We need to show that

$$\mathbf{0} \in \alpha\partial G_{\alpha,p}(\mathbf{w}) + \beta g'_{\beta,q}(\|\mathbf{w}\|_2)\frac{\mathbf{w}}{\|\mathbf{w}\|_2} + \mathbf{w} - \mathbf{x}. \tag{24}$$

Since $\mathbf{w} = \mathcal{S}_q(\mathbf{v}, \beta)$, by Prop. 7 we have that $\mathbf{w}$ minimizes $\beta g_{\beta,q}(\|\mathbf{w}\|_2) + \|\mathbf{w} - \mathbf{v}\|_2^2$, so that

$$\beta g'_{\beta,q}(\|\mathbf{w}\|_2)\frac{\mathbf{w}}{\|\mathbf{w}\|_2} + \mathbf{w} - \mathbf{v} = \mathbf{0}. \tag{25}$$

By substituting (25) into (24), it remains only to show that $\mathbf{0} \in \alpha\partial G_{\alpha,p}(\mathbf{w}) + \mathbf{v} - \mathbf{x}$. By applying Prop. 4 to our choice of G with $t\mathbf{x}$ in place of $\mathbf{x}$, we obtain that

$$\min_{\mathbf{u}} t\alpha G_{\alpha,p}(\mathbf{u}) + \tfrac{1}{2}\|\mathbf{u} - t\mathbf{x}\|_2^2 \tag{26}$$

is solved by

$$\begin{aligned}\mathbf{u} &= S_p(t\mathbf{x}, t\alpha) = \frac{t\mathbf{x}}{|t\mathbf{x}|}\max\{\mathbf{0}, |t\mathbf{x}| - (t\alpha)^{2-p}|t\mathbf{x}|^{p-1}\} \\ &= tS_p(\mathbf{x}, \alpha) = t\mathbf{v} = \mathbf{w},\end{aligned} \tag{27}$$

where the vector operations are to be understood componentwise. First-order optimality therefore tells us that

$$\mathbf{0} \in t\alpha\partial G_{\alpha,p}(\mathbf{w}) + t\mathbf{v} - t\mathbf{x}. \tag{28}$$

Dividing through by t completes the proof. □

Now we are ready to state our proposed algorithm. The following is our generalization of (12):

$$\begin{aligned}\min_{W,X} \alpha G_{\alpha,p}(W) + \beta\sum_{i=1}^{M} g_{\beta,q}(\|\mathbf{W}^i\|_2) \\ + \tfrac{1}{2}\|W - X - \Lambda\|_F^2 + \tfrac{1}{2}\|\Phi X - Y\|_F^2.\end{aligned} \tag{29}$$

We obtain our algorithm by solving the X and W subproblems and updating Λ at each iteration:

Input: signals Y, dictionary Φ, parameters α, β
Precompute: factorization of $I + \Phi^T\Phi$
Initialize: $W_0 = \Lambda_0 = \mathbf{0}$
for *number of iterations* **do**
  $(I + \Phi^T\Phi)X_n = W_{n-1} - \Lambda_{n-1} + \Phi^T Y$
  $\mathbf{W}_n^i = \mathcal{S}_q(S_p(\mathbf{X}_n^i, \alpha), \beta)$ for each i
  $\Lambda_n = \Lambda_{n-1} + X_n - W_n$
end
Output: Sparse, group-sparse coefficient vectors X

Algorithm 1: ADMM algorithm for sparsity with group sparsity

4. NUMERICAL RESULTS

We construct a synthetic test problem as follows. A random dictionary Φ of size $L \times M$ is constructed from i.i.d., standard normal values. J of the M columns in the dictionary are randomly selected, with a uniform distribution, to be allowed to be associated with non-zero coefficients in the randomly generated coefficient $M \times N$ matrix $\hat{X}$. For each column in $\hat{X}$, K of the allowed J entries are randomly selected, with a uniform distribution, to have nonzero values, and these nonzero values are assigned from a standard normal distribution. A reference signal is defined as $\hat{Y} = \Phi\hat{X}$, and a test signal Y is constructed by adding Gaussian white noise of standard deviation σ. For all the results reported here, $L = 512, M = 2048, K = 8, J = 64, N = 64$, and $\sigma = 5.0$.

We compare a number of variants of (29) in reconstructing $\hat{X}$ given Y, using different combinations of values of p and q, as well as omitting either the sparsity penalty or the group-sparse penalty (or equivalently, setting α or β to be zero in Alg. 1). When parameters α or β are not explicitly fixed to be zero, we compute the optimal parameter values for each problem by performing a search for those values that minimize the error in reconstructing $\hat{X}$. A new random $\hat{X}$, $\hat{Y}$, and Y are then generated, and the various optimizations are compared using the optimal parameters from the previous stage. The results of these experiments are presented in Table 1. Note that the best performance for a convex problem ($p = 1, q = 1$) is obtained when both α and β are nonzero, and that the corresponding problems that are nonconvex in both regularization terms exhibit significantly improved performance (and the best performance is also achieved when both α and β are nonzero).

p	q	SNR (dB)
1	N/A	5.39
N/A	1	-0.12
1	1	6.33
1/2	1	7.70
1	1/2	6.60
1/2	1/2	8.88
−1/2	N/A	6.61
−1/2	1	8.52
N/A	−1/2	5.01
1	−1/2	8.54
−1/2	−1/2	9.76

Table 1. SNR of recovered coefficients for choices of p and q ("N/A" means that the corresponding penalty term was omitted).

5. CONCLUSIONS

We have developed an ADMM algorithm for finding group-sparse representations having sparse groups using nonconvex regularization. With regard to prior work, a sparse-group, group sparse approach was used in [8, 10], but using convex optimization. An ADMM approach for group sparsity was developed in [18]; this algorithm can be considered a special case of the algorithm presented here with $p = 1$, $q = 1$, and $\alpha = 0$. A nonconvex approach for group sparsity, using a difference-of-convex-functions algorithm, has previously been proposed [21], but without enforcing sparsity of nonzero groups.

In the numerical experiments presented here we have shown that minimization of the nonconvex functional with terms for both sparse groups and group-sparsity can provide significantly better performance than either the convex functional with both terms, or a nonconvex functional with only one of these terms.

6. REFERENCES

[1] L. Jacob, G. Obozinski, and J. Vert, "Group Lasso with overlap and graph Lasso," in *International Conference on Machine Learning*, 2009, pp. 433–440.

[2] J. Huang and T. Zhang, "The benefit of group sparsity," *Ann. Statist.*, vol. 38, pp. 1978–2004, 2010.

[3] M. Stojnic, F. Parvaresh, and B. Hassibi, "On the reconstruction of block-sparse signals with an optimal number of measurements," *IEEE Trans. Signal Process.*, vol. 57, pp. 3075–3085, 2009.

[4] A. Majumdar and R. K. Ward, "Classification via group sparsity promoting regularization," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2009, pp. 861–864.

[5] E. Elhamifar and R. Vidal, "Robust classification using structured sparse representation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1873–1879.

[6] J. Gao, Q. Shi, and T. S. Caetano, "Dimensionality reduction via compressive sensing," *Pattern Recognit. Lett.*, vol. 33, pp. 1163–1170, 2012.

[7] Y. Liu, F. Wu, and Y. Zhuang, "Group sparse representation for image categorization and semantic video retrieval," *Sci. China Inform. Sci.*, vol. 54, pp. 2051–2063, 2011.

[8] B. Wohlberg, R. Chartrand, and J. Theiler, "Local principal component pursuit for nonlinear datasets," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2012, pp. 3925–3928.

[9] M. Vincent and N. R. Hansen, "Sparse group lasso and high dimensional multinomial classification," 2012, arXiv:1205.1245.

[10] P. Sprechmann, I. Ramirez, G. Sapiro, and Y. C. Eldar, "C-HiLasso: A collaborative hierarchical sparse modeling framework," *IEEE Trans. Signal Process.*, vol. 59, pp. 4183–4198, 2011.

[11] R. Chartrand, "Exact reconstructions of sparse signals via nonconvex minimization," *IEEE Signal Process. Lett.*, vol. 14, pp. 707–710, 2007.

[12] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 035020, pp. 1–14, 2008.

[13] R. Chartrand, "Fast algorithms for nonconvex compressive sensing: MRI reconstruction from very few data," in *IEEE International Symposium on Biomedical Imaging*, 2009.

[14] ——, "Nonconvex splitting for regularized low-rank + sparse decomposition," *IEEE Trans. Signal Process.*, vol. 60, pp. 5810–5819, 2012.

[15] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operator-Splitting Methods in Nonlinear Mechanics*. Philadelphia, Pennsylvania: SIAM, 1989.

[16] T. Goldstein and S. Osher, "The split Bregman method for L1 regularized problems," *SIAM J. Imaging Sci.*, vol. 2, pp. 323–343, 2009.

[17] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, pp. 1–122, 2010.

[18] W. Deng, W. Yin, and Y. Zhang, "Group sparse optimization by alternating direction method," Rice University Computational and Applied Mathematics, Tech. Rep. TR11-06, 2011.

[19] R. T. Rockafellar and R. J.-B. Wets, *Variational Analysis*. Berlin: Springer-Verlag, 1998.

[20] J. Yang, W. Yin, Y. Zhang, and Y. Wang, "A fast algorithm for edge-preserving variational multichannel image restoration," *SIAM J. Imaging Sci.*, vol. 2, pp. 569–592, 2009.

[21] S. Xiang, X. Shen, and J. Ye, "Efficient sparse group feature selection via nonconvex optimization," 2012, arXiv:1205.5075.