NON-CONVEX GROUP SPARSITY: APPLICATION TO COLOR IMAGING
Angshul Majumdar and Rabab K. Ward
Department of Electrical and Computer Engineering, University of British Columbia

ABSTRACT

This work investigates a group-sparse solution to the under-determined system of linear equations $b = Ax$ where the unknown $x$ is formed of a group of vectors $x_i$'s. A group-sparse solution has only a few $x_i$ vectors as non-zeroes while the rest are zeroes. To seek a group-sparse solution generally a convex optimization problem is solved. Such an optimization criterion is unsuitable when the system is highly under-determined or when some of the vector $x_i$'s are themselves sparse. For such cases, we propose an alternate non-convex optimization problem. Simulation results show that the proposed method yields significantly improved results (2 orders of magnitude) over the standard method. We also apply the proposed group-sparse optimization in a novel fashion to the problem of color imaging. The new method shows an improvement of more than 1dB over the standard method.

Index Terms— group sparsity, color imaging, compressed sensing.

1. INTRODUCTION

Consider an under-determined system of linear equations

$$y_{m 	imes 1} = A_{m 	imes n} x_{n 	imes 1}, m < n$$

where $x = [x_1, x_2, ..., x_C]^T$ and $y = [y_1, y_2, ..., y_m]^T$.

The problem is to solve (1) so that the solution is sparse in groups, i.e. only a few $x_i$ vectors have non-zero values while the other $x_i$'s are zeroes. Problems of this nature arise in several areas of signal processing and machine learning, such as DNA microarrays [1], sparse channel estimation [2], grouped linear regression [3], multi-response linear regression [4], simultaneous variable selection [5], grouped logistic regression [6] etc.

To recover the group-sparse solution to (1) the following optimization problem is usually sought:

$$\min \| x \|_{2,1} \quad \text{subject to} \ y = Ax$$

(2)

where $\| x \|_{2,1} = \sum_{i=1}^{C} \| x_i \|_2$.

The geometrical justification for employing such a mixed norm $\| \cdot \|_{2,1}$ for recovering group-sparse solutions is given in [3]. The inner $l_2$-norm of $\| x \|_{2,1}$ enforces selection of all the coefficients within a group while the outer $l_{1}$-norm promotes sparsity in the number of selected groups.

While solving (2) it is assumed that all the coefficients of a group $x_i$ are non-zeroes; however this might not always be the case. The groups may be sparse internally, i.e. some of the coefficients within $x_i$'s may be zeroes. If that is the case, the inner $l_2$-norm of $\| x \|_{2,1}$ (which favors a dense solution) will not yield the desired solution. Therefore a norm that enforces sparsity within the groups is required. Sparsity within the group can be enforced by employing $l_p$ inner norm where $0 < p \leq 1$.

The outer $l_1$-norm of $\| x \|_{1,1}$ behaves similarly to the $l_1$-norm of the standard sparse estimation problem [7, 8], but instead of promoting sparsity over the coefficients, it promotes sparsity on the selection of groups ($x_i$'s). From the signal processing literature it is known that a fractional $l_{r}$-norm (i.e. $0 < r < 1$) enforces better sparsity than the $l_1$-norm [9]. In other words a fractional norm can solve (1) with significantly fewer number of equations than required by the $l_1$-norm. Thus to ensure sparser selection of groups for solving (1), it is imperative to use a fractional outer norm.

Keeping the above discussion in mind, we propose the following novel optimization problem for obtaining a group sparse solution to (1):

$$\min \| x \|_{m,p} \quad \text{subject to} \ y = Ax$$

(3)

where $\| x \|_{m,p} = \left( \sum_{i=1}^{C} \| x_i \|_{m}^p \right)^{1/p}$, $0 < m \leq 2$ and $0 < p \leq 1$.

This is a non-convex optimization problem, for which FOCUSS (FOCal Under-determined System Solver) [10] is suitable as a solver. One can also employ gradient based methods for solving (3), but it was recently proved that the FOCUSS has faster convergence rate (super-linear) than gradient based methods (linear). Therefore, we will base our optimization algorithm on the FOCUSS.

The rest of the paper will be divided into several sections. In the next section a brief review of FOCUSS is provided and the algorithm for solving (3) is derived. In Section 3, we discuss how the problem of compressive color imaging can be framed as a group-sparsity problem. Section 4 presents the experimental results, and Section 5 concludes the work.
2. MAIN ALGORITHM

2.1. Brief Review of FOCUSS

FOCUSS is especially suited for solving problems of the following type:

\[ \min E(x) \text{ subject to } y = Ax \]  

where \( E(x) \) is a diversity measure.

The Lagrangian for (4) is

\[ L(x, \lambda) = E(x) + \lambda^T (Ax - b) \]  

Following the theory of Lagrangian, the stationary point for (5) needs to be obtained by solving

\[ \nabla_x L(\hat{x}, \hat{\lambda}) = 0 \]  

\[ \nabla_\lambda L(\hat{x}, \hat{\lambda}) = 0 \]

(6a)

(6b)

Solving (6) requires the gradient of \( E(x) \) with respect to \( x \), please note that so as to preserve generality the explicit functional form of \( E(x) \) has not been stated. In the diversity measures of interest to the signal processing community, the gradient of \( E(x) \) can be expressed as,

\[ \nabla_x E(x) = \alpha(x) \Pi(x) x \]  

where \( \alpha(x) \) is a scalar, and \( \Pi(x) \) is a diagonal matrix.

Solving (6) yields and applying (7) yields

\[ \hat{x} = -\frac{1}{\alpha(\hat{x})} \Pi(\hat{x})^{-1} A^T \hat{\lambda} \]  

(8a)

\[ \hat{\lambda} = -\alpha(x)(\alpha \Pi(\hat{x})^{-1} A^T)^{-1} y \]  

(8b)

Substituting (8b) into (8a) yields

\[ \hat{x} = \Pi(\hat{x})^{-1} A^T (\alpha \Pi(\hat{x})^{-1} A^T)^{-1} y \]  

(9)

The expression (9) is an implicit function of \( x \). Therefore it needs to be solved iteratively in the following manner:

\[ x(t + 1) = \Pi(x(t))^{-1} A^T (\alpha \Pi(x(t))^{-1} A^T)^{-1} y \]  

(10)

We have made the notation as similar as possible to the original work [10]. There is a basic problem however with this FOCUSS approach. The theory of linear Lagrangian used to derive the FOCUSS algorithm is valid for convex problems, i.e. for convex diversity measures. However, FOCUSS is almost always used for solving non-convex diversity measures such as \( \ell_p \)-norm minimization [9]. There is no theoretical guarantee that the FOCUSS will converge to the desired solution for such non-convex problems, but practical experiments show that this method (with a little modification) almost always provides exceptionally good results. We have therefore adopted this method to solve our non-convex optimization problem (3).

2.2. Group-Sparse Diversity Measure

The diversity measure for the group-sparse optimization problem (3) is

\[ \| x \|_{m,p} = (\sum_{i=1}^C \| x_i \|_m^p)^{1/p} \]

The minimizer for the above is the same as the minimizer of

\[ \| x \|_{m,p}^m = (\sum_{i=1}^C \| x_i \|_m^p) \]

For mathematical convenience we will use the latter form as the divergence measure.

The gradient is

\[ \nabla_x = [\frac{\partial}{\partial x_{1,1}}, \ldots, \frac{\partial}{\partial x_{n,n}}, \ldots, \frac{\partial}{\partial x_{C,1}}, \ldots, \frac{\partial}{\partial x_{C,n}}]^T \]  

(11)

For each of the terms in (11), the derivative is calculated to be

\[ \frac{\partial}{\partial x_{j,j}} \| x \|_{m,p}^m = p \| x_i \|_{m}^{p-m} x_{j,j} \]  

(12)

Comparing (12) with (7), we have

\[ \alpha(x) = p \]

\[ \Pi(x) = diag(\| x_i \|_{m}^{p-m} | x_{j,j} |^{m-2}) \]  

(13)

In principle we can apply (10) repeatedly to solve our problem (3). However, certain modifications to (13) are required for practical purposes. We will discuss the practical algorithmic steps in the next subsection.

2.3. Algorithm for Group-Sparse Optimization

The optimization problem (3) is non-convex. Therefore, there is a chance that the optimization algorithm converges to local minima. To reduce the chances of converging to a local minimum, a simple modification has been proposed in [9]. Following this suggestion, we apply a damping factor to (13), i.e.,

\[ \Pi(x) = diag(\| x_i \|_{m}^{p-m} | x_{j,j} |^{m-2} + \delta) \]  

(14)

where \( \delta \) is the damping factor, whose value is reduced at each iteration.

The damping factor also sets bound to the values of the diagonal matrix and does not allow its inverse to become excessively large. The convergence (to a local minimum) of such a damped FOCUSS class of algorithms has been proven in [11], where it is shown that such algorithms have super-linear convergence.

Computing (10) at each step is computationally expensive since it involves finding explicit matrix inverses. To avoid this expense we propose to solve (10) by the conjugate gradient (CG) method, in the following steps:

\[ R = diag((\| x_i \|_{m}^{p-m} | x_{j,j} |^{m-2} + \delta)^{-1/2}) \]

(15a)

\[ \Phi = AR \]  

(15b)

\[ z(t + 1) = \min \{ y - \Phi z \} \]  

(15c)

\[ x(t + 1) = Rz(t + 1) \]  

(15d)
This solution precludes computing explicit inverses. Moreover it does not require A to be specified explicitly; it can also be applied as fast operators.

It is not necessary to solve (15c) perfectly, as only a few steps of CG is sufficient for practical implementation.

3. COMPRESSIVE COLOR IMAGING

As a practical application to the problem of group-sparse optimization we pose a novel problem of reconstructing color images acquired by Compressed Sensing techniques. Color images when acquired comprise three color channels – Red (R), Green (G) and Blue (B). Each channel is known to be sparse in the DCT or Wavelet domain. The synthesis and analysis equations for each color channel are expressed as,

\[ K = \Phi^T x_K, K \in \{ R, G, B \} \]

\[ x_K = \Phi K \]

where \( K \) is the originally color channel in the pixel domain and \( x_K \) is the sparse transform coefficient. \( \Phi \) is the sparsifying transform e.g. DCT or Wavelet.

If compressive sampling is performed on each color channel by a suitable matrix (obeying RIP [12]), then instead of acquiring the R, G and B values, compressive sampling obtains their projections in the following manner:

\[ y_K = P_K K = P_K \Phi^T x_K, K \in \{ R, G, B \} \]

where \( y_K \) represent the compressive samples and \( P_K \) is the projection matrix for each channel. The projection matrix may be the same for all channels, but for the sake of generality we assume that it is different for each channel.

In compact matrix-vector notation, (17) can be expressed as

\[
\begin{bmatrix}
    y_R \\
    y_G \\
    y_B
\end{bmatrix} =
\begin{bmatrix}
    P_R \Phi^T & 0 & 0 \\
    0 & P_G \Phi^T & 0 \\
    0 & 0 & P_B \Phi^T
\end{bmatrix}
\begin{bmatrix}
    x_R \\
    x_G \\
    x_B
\end{bmatrix}
\]

or \( y = Ax \)

The equations in (18) can be re-arranged so that ‘x’ has the following group sparse structure:

\[ x = [x_{1,R}, x_{1,G}, x_{1,B}, \ldots, x_{C,R}, x_{C,G}, x_{C,B}] \]

Here C is the total number of transform coefficients. The coefficients are indexed by \( j=1\ldots C \). Each index corresponds to a group of triplets \( (x_{j,R}, x_{j,G}, x_{j,B}) \).

It is well known that the R, G and B color channels are usually highly correlated. Since the coefficients \( x_R, x_G, x_B \) are the orthogonal transforms of the color channels, the transform coefficients in these cases will also be correlated with each other. Ordinarily it is assumed that if at an index \( j \), \( x_R \) is high, it is likely that the corresponding coefficient values of \( x_G \) and \( x_B \) will also be high. But this is not always true as there are many naturally occurring scenes like the sky (blue), cherry/apples (red) and grass (green) where only one of the color channels is active. Therefore it is incorrect to assume that all the three coefficients will ALWAYS have similar values ( e.g. all have high values or all have zero values) In such cases, the problem formulation should allow a solution where, at some indices, only one of the color coefficient values will be non-zero and the others zeroes. This leads to a group sparsity problem where some of the groups are themselves sparse. Therefore instead of employing (2) one should apply our proposed optimization problem (3) to solve the reconstruction problem.

4. EXPERIMENTAL EVALUATION

4.1. Results on Synthetic Data

In these experiments, the sparse vector was of length 150, which was divided randomly into 25 groups. Out of the 25 groups, 3 groups were active, i.e., had non-zero coefficients. The reconstruction accuracy is measured in terms of Normalized Mean Squared Error (NMSE). Each experimental configuration was repeated 10,000 times, and the tabulated results are the average over all the runs.

Table 1 shows how the reconstruction error varies with the value of the inner norm (m), keeping the outer norm (p) at a constant value 0.4. The number of random projections has been fixed to 50 for this set of experiments. Table 2 shows how the reconstruction error varies when the outer norm (p) is changed and the inner norm (m) is kept at a constant value 0.8. Table 3 shows how the reconstruction error varies when the sparsity is fixed (50%) but the outer norm (p) is changed with the number of random of projections.

<table>
<thead>
<tr>
<th>Value of m</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>80</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>4.5X10^-6</td>
<td>4.3X10^-6</td>
<td>3.2X10^-5</td>
<td>5.0X10^-6</td>
<td>4.9X10^-4</td>
</tr>
<tr>
<td>0.8</td>
<td>5.9X10^-6</td>
<td>4.3X10^-6</td>
<td>3.5X10^-5</td>
<td>5.0X10^-6</td>
<td>4.5X10^-4</td>
</tr>
<tr>
<td>1</td>
<td>7.1X10^-6</td>
<td>5.0X10^-6</td>
<td>5.0X10^-5</td>
<td>3.0X10^-5</td>
<td>1.3X10^-4</td>
</tr>
<tr>
<td>2</td>
<td>1.9X10^-4</td>
<td>1.6X10^-4</td>
<td>1.1X10^-4</td>
<td>1.2X10^-4</td>
<td>1.4X10^-5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Value of p</th>
<th>10</th>
<th>25</th>
<th>50</th>
<th>80</th>
<th>100</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4</td>
<td>5.9X10^-6</td>
<td>4.3X10^-6</td>
<td>3.5X10^-5</td>
<td>5.0X10^-6</td>
<td>4.5X10^-4</td>
</tr>
<tr>
<td>0.6</td>
<td>1.9X10^-4</td>
<td>1.9X10^-4</td>
<td>1.4X10^-5</td>
<td>2.2X10^-5</td>
<td>3.9X10^-4</td>
</tr>
<tr>
<td>1</td>
<td>5.4X10^-4</td>
<td>8.5X10^-5</td>
<td>0.0025</td>
<td>0.0174</td>
<td>0.1074</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Projections</th>
<th>p=0.4</th>
<th>p=0.6</th>
<th>p=1</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>3.5X10^-6</td>
<td>1.4X10^-5</td>
<td>0.0248</td>
</tr>
<tr>
<td>100</td>
<td>2.2X10^-5</td>
<td>7.7X10^-6</td>
<td>3.8X10^-4</td>
</tr>
<tr>
<td>150</td>
<td>1.5X10^-5</td>
<td>5.2X10^-5</td>
<td>1.8X10^-4</td>
</tr>
</tbody>
</table>
Tables 1 and 2 show that when the coefficients within each group are sparse, fractional norms give significantly better results by two orders of magnitude, compared with the standard $l_2,p$-norm. This improvement in performance can be seen in less sparse (80% full) groups as well. Only when all the coefficients are non-zeroes does the standard solution yield better results by an order or magnitude.

Table 3 shows how the reconstruction result varies with the number of random projections for a fixed sparsity level. Significant improvements of several orders of magnitude are achieved using a fractional (non-convex) outer norm compared to the standard unity (convex) norm.

### 4.2. Results on Real Color Images

Color imaging experiments were carried out on the images shown in Fig. 1. Discrete Cosine Transform (DCT) was used as the sparsifying transform. A random projection matrix was formed by normalizing the columns of an i.i.d. Gaussian matrix for each of the three channels (R, G and B). The number of projections was varied as a fraction of the length of each color channel.

![Fig. 1. Barbara, Lena and Peppers (Left to Right)](image)

Table 4 shows the result of color image reconstruction (in terms of PSNR). The number of random projections was kept at 30% of the total number of pixels for each color channel. Two values of the inner ($m = 1$ and 2) and outer ($p = 0.6$ and 1) norm were considered for each image.

<table>
<thead>
<tr>
<th>Image</th>
<th>$m=1,p=0.6$</th>
<th>$m=1,p=1$</th>
<th>$m=2,p=0.6$</th>
<th>$m=2,p=1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Barbara</td>
<td>34.15</td>
<td>33.66</td>
<td>33.51</td>
<td>33.10</td>
</tr>
<tr>
<td>Lena</td>
<td>33.40</td>
<td>32.72</td>
<td>32.76</td>
<td>32.11</td>
</tr>
<tr>
<td>Peppers</td>
<td>35.19</td>
<td>34.83</td>
<td>34.47</td>
<td>34.04</td>
</tr>
</tbody>
</table>

Table 4 shows that significant improvement can be achieved by deviating from the standard convex optimization problem ($\parallel \cdot \parallel_{l,1}$ minimization) employed for group-sparsity. An improvement in PSNR values of more than 1dB can be achieved by solving the non-convex optimization problem ($\parallel \cdot \parallel_{l,0.6}$ minimization).

### 5. CONCLUSION

This paper explores the problem of obtaining a group-sparse solution to an under-determined system of linear equations. Unlike the standard model of group sparsity, which frames the problem as one of convex optimization, we propose a very flexible model that can frame (and solve) group-sparsity as either a convex or non-convex optimization problem. This flexibility permits the solution of a wider variety of group-sparse optimization problems (different levels of sparsity and under-determination).

Experiments on synthetic data show that the proposed method (non-convex optimization) yields reconstruction results which are about two orders of magnitude better than the standard (convex optimization) method.

The group-sparse optimization problem is applied for the first time to the problem of acquiring color images via compressive sampling. Results show significant improvement in PSNR values (more than 1 dB) using our non-convex optimization solution compared with the results obtained using the standard convex solution of group-sparsity.

### 6. REFERENCES


