University of California, Merced

Capstone Project

Sparse Image and Video Recovery Using Gradient Projection for Linearly Constrained Convex Optimization

Daniel Thompson

A technical report submitted
in partial fulfillment of the requirements for the degree of

Master of Science in Applied Mathematics

May, 2011

UNIVERSITY OF CALIFORNIA, MERCED
Graduate Division

This is to certify that I have examined a copy of a technical report by

Daniel Thompson

and found it satisfactory in all respects, and that any and all revisions required by the examining committee have been made.

Research Advisor: ______________________
Roummel F. Marcia

Reading Committee: ______________________
Arnold D. Kim

Applied Mathematics Graduate Studies Chair: ______________________
Boaz Ilan

____
Date

ABSTRACT OF THE CAPSTONE PROJECT

Sparse Image and Video Recovery Using Gradient Projection for Linearly Constrained Convex Optimization

by

Daniel Thompson
May 2011

University of California, Merced

Abstract

This project concerns the reconstruction of a signal, which corresponds to either an image or a temporally-varying scene. Signal recovery can be accomplished through finding a sparse solution to an ℓ_2-ℓ_1 minimization problem, which can be solved efficiently by gradient projection. In imaging applications, the signal of interest corresponds to nonnegative pixel intensities. Hence, additional nonnegativity constraints are placed upon the ℓ_2-ℓ_1 minimization problem. This results in a more difficult problem to solve, but one with a higher potential for accurate reconstructions. This work focuses on a gradient projection approach for sparse signal recovery that incorporates nonnegativity constraints in the minimization problem. For video recovery we exploit inter-frame correlations to improve upon the naïve approach of solving each frame independently. Numerical results are presented for both an image and video experiment to demonstrate the effectiveness of this approach.

1 Introduction

This research is concerned with the reconstruction of a sparse signal from linear, potentially low-dimensional, observations. Specifically the signal of interest is that of an image. We assume our observation $y \in \mathbb{R}^m$ is a linear projection of the true signal $f^\star \in \mathbb{R}^n$. Then we can model our observation as the linear system $y = Rf^\star + \eta$ where $R \in \mathbb{R}^{m\times n}$ is a linear projection matrix and η is noise. Our goal is to recover $f^\star$ from our observation y. This is not as simple as using the inverse R^{-1} to solve $f^\star = R^{-1}(y - \eta)$. The matrix R is not necessarily square, and in the case that R is square, it is not necessarily invertible. Therefore we must find an estimate to $f^\star$ in some other way than using the inverse. Let us assume for now $\eta = 0$, the noiseless case. Thus we now have $y = Rf^\star$. If y is an element of the column space of R, i.e. there exists a $\widehat{f}$ such that $y = R\widehat{f}$, then $\widehat{f}$ may not be unique. This is the case in which the null space of R is nontrivial. In particular if R is underdetermined with $m < n$, then R necessarily has a nontrivial null space. Since R may be underdetermined, we could find the solution with the lowest nonzero elements. This solution can be found through a different problem formulation that includes a data fitting term balanced with a sparsity enforcing term.

To enforce fidelity to the data in our new problem formulation we can use a least squares term $\|y - Rf^\star\|_2$. We must solve for the signal that minimizes this term due to the noise. Alone this term does not take into account sparsity. Therefore we must include a sparsity enforcing term in our problem formulation. The ℓ^0 norm counts the number of nonzero entries in a vector. Hence we want to minimize $\|f^\star\|_0$ subject to $y = Rf^\star$. However, we must determine the positions of nonzero elements which is a combinatorial problem and is therefore extremely difficult. Therefore we seek a formulation that can be solved more simply. Work by Candès, Romberg and Tao [4] indicates that we can instead minimize $\|f^\star\|_1$ subject to $y = Rf^\star$ to recover the sparsest solution. Minimizing the ℓ^1 norm allows us to use non-combinatorial optimization methods when solving for a reconstruction. By using the ℓ^2 norm and ℓ^1 norm we may now set up a convex ℓ_2-ℓ_1 optimization problem, discussed in Sec. 2. This optimization problem can be solved efficiently by gradient projection, discussed in Sec. 3.

There are multiple methods that can be used for sparse signal recovery for example, homotopy algorithms such as Least Angle Regression (LARS)[6], Iterative shrinking/thresholding (IST) algorithms [7], and matching pursuit [11]. However, such algorithms do not take into account the important property that pixel values of images correspond to light intensities and are naturally nonnegative quantities; thus, their signal estimates may contain negative pixel intensity values. With lower computational cost than some other standard methods, gradient projection is particularly effective for sparse recovery problems and in the case of Gradient Projection for Sparse Reconstruction (GPSR) [8] has been shown to produce estimates faster than state of the art competing methods. We wish to extend this approach to account for nonnegative pixel values. Our goal is to reconstruct an image from a blurred noisy observation by enforcing nonnegativity constraints and show that this leads to an improvement over the conventional gradient projection methods. We extend the developed method to the reconstruction of a sequence of images in a video. In applications such as surveillance and monitoring, the scene changes very little from frame to frame. Therefore inter-frame differences will be very sparse, and ℓ_2-ℓ_1 minimization is particularly well-suited for this type of problem if we express the variables as the frame differences. Just as in the previous method we improve upon current methods by enforcing nonnegativity in our frame estimates

2 The Optimization Problem

Recent development in compressed sensing theory indicates that solving the following ℓ_2-ℓ_1 minimization problem

$$\widehat{f} \equiv \arg\min_{f\in\mathbb{R}^n} \frac{1}{2}\|y - Rf\|_2^2 + \tau\|f\|_1$$

will lead to a highly accurate signal estimate with very high probability [5, 10]. Here $\widehat{f}$ denotes our estimate, y is our noisy observation, R is a linear projection matrix, and $\tau > 0$ is a regularization parameter. The ℓ^2 term is a least-squares term that enforces fidelity to the data while the 1-norm term promotes sparsity in the solution. A larger weight on the ℓ^1 term encourages a greater reduction in this term than in the ℓ^2 term in the minimization process. Thus choosing higher values for τ forces f to be more sparse. Since the signal of interest is an image and therefore corresponds to nonnegative pixel intensities, we impose a nonnegativity constraint on the estimate. Thus we instead solve

$$\begin{aligned} \widehat{f} \equiv\ & \arg\min_{f\in\mathbb{R}^n} \frac{1}{2}\|y - Rf\|_2^2 + \tau\|f\|_1 \\ & \text{subject to} \quad f \geq 0. \end{aligned} \tag{1}$$

Often, the signal $f^\star$ is not sparse in the canonical basis but rather in some other basis, W. Therefore, we write $f^\star = W\theta^\star$, where $\theta^\star$ is sparse. In this research, we use the wavelet basis since most natural images are sparse in this basis. In addition, the discrete wavelet transform W is orthogonal, and matrix-vector multiplication can be performed quickly and efficiently. We now rewrite (1) as

$$\begin{aligned} \widehat{\theta} &\equiv \arg\min_{\theta\in\mathbb{R}^n} \quad \frac{1}{2}\|y - RW\theta\|_2^2 + \tau\|\theta\|_1 \\ & \quad\ \text{subject to} \quad W\theta \geq 0 \\ \widehat{f} &\equiv W\widehat{\theta}. \end{aligned} \tag{2}$$

3 Gradient Projection

One advantage to using gradient projection is that the negative gradient always guarantees a decrease in the objective function with a proper step length. Although using the negative gradient generally has slow convergence, proper step lengths of each iterate can yield improved performance over the classical steepest-descent method [1]. To solve (2) using gradient-based optimization the objective function must be differentiable. The 1-norm term in (2) is not differentiable at $\theta_i = 0$ for all i. We can reformulate the objective function as a differentiable function by decomposing θ into its positive and negative components, $\theta = u - v$ with $u, v \geq 0$ and $u^T v = 0$. This way $\|\theta\|_1 = \mathbb{1}_n^T(u + v)$, where $\mathbb{1}_n \in \mathbb{R}^n$ is the n-vector of ones, and is now differentiable everywhere. Rewriting (2) in terms of these new variables, we get

$$\begin{aligned} (\widehat{u}, \widehat{v}) &\equiv \arg\min_{u,v\in\mathbb{R}^n} \quad \frac{1}{2}\|y - RW(u - v)\|_2^2 + \tau\mathbb{1}_n^T(u + v) \\ & \quad\ \text{subject to} \quad u, v \geq 0, \quad W(u - v) \geq 0 \\ \widehat{f} &\equiv W(\widehat{u} - \widehat{v}). \end{aligned} \tag{3}$$

For the purpose of simplifying notation we let

$$z = \begin{bmatrix} u \\ v \end{bmatrix} \in \mathbb{R}^{2n}, \qquad \widetilde{W} = \begin{bmatrix} W & -W \end{bmatrix} \in \mathbb{R}^{n\times 2n}, \qquad B = \begin{bmatrix} I_{2n} \\ \widetilde{W} \end{bmatrix} \in \mathbb{R}^{3n\times 2n}$$

where I_{2n} is the identity matrix of size $2n$. Then we can rewrite (3) as

$$\begin{aligned} \widehat{z} &\equiv \underset{z\in\mathbb{R}^{2n}}{\arg\min} \quad \phi(z) \equiv \frac{1}{2}\|y - R\widetilde{W}z\|_2^2 + \tau \mathbb{1}_{2n}^T z \\ & \quad\ \text{subject to} \quad Bz \geq 0 \\ \widehat{f} &\equiv \widetilde{W}\widehat{z} \end{aligned} \tag{4}$$

and now $\mathbb{1}_{2n} \in \mathbb{R}^{2n}$. Note, however, that this minimization problem now has twice as many variables as (2). Additionally, new nonnegativity constraints on the variables are introduced.

Now we can apply gradient projection to (4). By choosing an appropriate step length along the negative gradient, i.e., the steepest descent, we can guarantee a decrease in the objective function. However, with a large enough step length this direction may result in an infeasible point. To maintain feasibility we then choose the direction with the smallest possible step length back to the feasible set. Thus we project back onto the feasible set. Then a second step length is computed to minimize the objective function in this new direction. This is known as the two-step gradient projection method [1]. In this approach we move from iterate $z^{(k)}$ to $z^{(k+1)}$ by first defining the vector $z^{(k)} - \alpha^{(k)}\nabla\phi(z^{(k)})$ from the method of steepest descent and projecting it onto the feasible set

$$z_P^{(k)}(\alpha^{(k)}) = P\left(z^{(k)} - \alpha^{(k)}\nabla\phi(z^{(k)})\right). \tag{5}$$

Here, P is the projection operator onto the feasible set $\mathcal{F} \equiv \{z \in \mathbb{R}^{2n} \colon Bz \geq 0\}$, and $\alpha^{(k)} > 0$ is the step length along the negative gradient. Now we perform a line search to find the next iterate. We define $\delta^{(k)} = (z^{(k)} - \alpha^{(k)}\nabla\phi(z^{(k)}))_+ - z^{(k)}$, where $(\cdot)_+$ zeros negative values. Then we find the scalar $\beta^{(k)} \in [0,1]$ that minimizes $\phi(z^{(k)} + \beta^{(k)}\delta^{(k)})$ and set

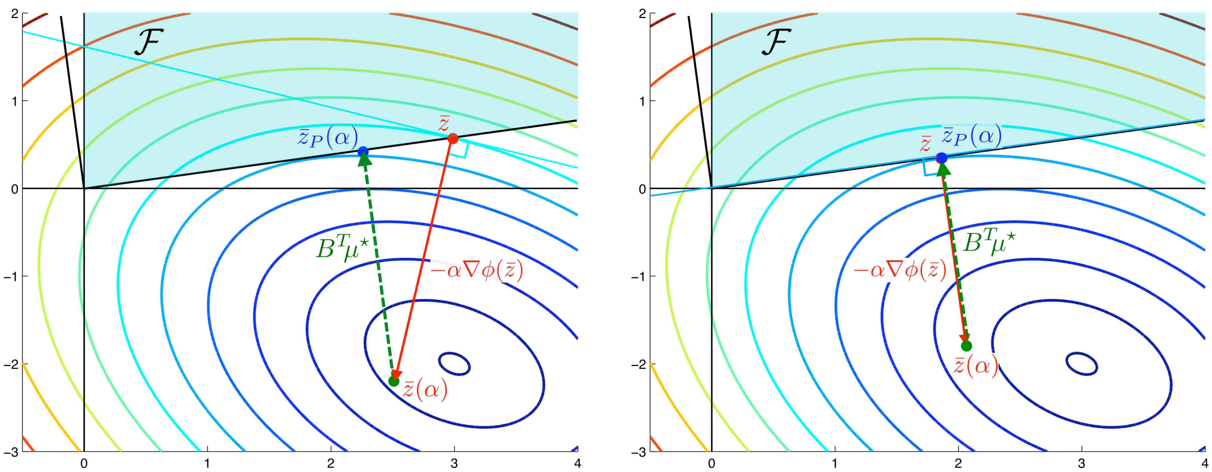
$$z^{(k+1)} = z^{(k)} + \beta^{(k)}\left(z_P^{(k)}(\alpha^{(k)}) - z^{(k)}\right).$$

There are multiple ways to determine $\alpha^{(k)}$. In the case of this research we use the approach described in the Barzilai-Borwein variant of GPSR, known as GPSR-BB, which is among the fastest of GPSR variants tested [8]. We choose this method of determining $\alpha^{(k)}$ because it has been shown to have a lower computational cost and improved performance over the classical steepest descent method [1]. The GPSR-BB algorithm is a gradient projection algorithm that solves the unconstrained problem (2). This approach differs from the basic GPSR algorithm in how the step length $\alpha^{(k)}$ is chosen which is based on work from Barzilai and Borwein [1]. This approach determines the next step length through a quasi-Newton method. We define $\Delta z^{(k)} = z^{(k+1)} - z^{(k)}$ and by Taylor expansion we have

$$\phi(z^{(k)} + \Delta z^{(k)}) \approx \phi(z^{(k)}) + \nabla\phi(z^{(k)})^T\Delta z^{(k)} + \frac{1}{2}\Delta(z^{(k)})^T H_k \Delta z^{(k)}$$

where H_k is an approximation to the Hessian of ϕ at $z^{(k)}$. By taking the gradient of this approximation, setting it equal to zero and assuming H_k is invertible, we can derive

$$\Delta z^{(k+1)} = -H_k^{-1}\Delta\phi(z^{(k)})$$



(a) (b)

Figure 1: A two-dimensional illustration of gradient projection onto a linearly constrained feasible set $\mathcal{F}$. The matrix-vector product $B^T\mu^\star$ is defined in sec 4. (a) relates to Prop. 1a, where $\bar{z}$ is not a stationary point thus there exists a scalar $\bar{\alpha} > 0$ such that for all $\alpha \in (0, \bar{\alpha}]$, $\phi(\bar{z}_P(\alpha)) < \phi(\bar{z})$. (b) relates to Prop. 1b where $\bar{z}$ is a stationary point hence for all step lengths α we project back onto $\bar{z}$, meaning for all $\alpha \geq 0$, $\bar{z}_P(\alpha) = \bar{z}$

which is the same as

$$z^{(k)} = z^{(k)} - H_k^{-1}\nabla\phi(z^{(k)}).$$

Thus we choose α^k so that $\alpha^k I$ approximates H_k^{-1}. Here [1] chooses the approximation to be a multiple of the identity matrix, $H_k = s^k I$, where s^k is a scalar chosen such that this approximation has similar behavior to the true Hessian over the most recent step, i.e., s^k minimizes $\|(z^{(k)} - z^{(k-1)}) - s^k(\phi(z^{(k)}) - \phi(z^{(k-1)}))\|_2^2$. The purpose of this choice of s^k is to provide a two-point approximation to the secant equation underlying quasi-Newton methods. For specifics of how $\alpha^{(k)}$ is computed see [1, 8].

For ease of notation, we drop the superscripts corresponding to the iterates $z^{(k)}$ and denote the current iterate by $\bar{z}$. We define

$$\bar{z}(\alpha) = \bar{z} - \alpha\nabla\phi(\bar{z}) \quad \text{and} \quad \bar{z}_P(\alpha) \equiv P(\bar{z} - \alpha\nabla\phi(\bar{z})).$$

With this the following proposition holds [3]

Proposition 1: Let $\bar{z}$ be a feasible point, i.e., $B\bar{z} \geq 0$.
(a) If $\bar{z}$ is not a stationary point, then there exists a scalar $\bar{\alpha} > 0$ such that

$$\phi(\bar{z}_P(\alpha)) < \phi(\bar{z}), \qquad \text{for all } \alpha \in (0, \bar{\alpha}].$$

(b) The point $\bar{z}$ is stationary if and only if

$$\bar{z}_P(\alpha) = \bar{z} \qquad \text{for all } \alpha \geq 0.$$

Proof:

(a) Since $\bar{z}$ is not a stationary point we are not at a minimum. The next iterate is chosen by moving along the negative gradient, the direction of greatest decrease. Thus we are guaranteed that a sufficiently small step length, $\bar{\alpha}$, will cause a decrease in the objective function. If this step length moves the next iterate out of the feasible set then the projection will move each iterate in the direction of greatest decrease along the border. Thus if $\bar{z}$ is not a stationary point then there exists a $\bar{\alpha}$ such that $\phi(\bar{z}_P(\alpha)) < \phi(\bar{z})$, for all $\alpha \in (0, \bar{\alpha}]$.

(b) See the proof of part (c) of Lemma 3.1 from [3]. The theorem and proof are similar.

We have reached an optimal point when for every step length chosen we project back onto the point we started with. Figure 1 provides an illustrated example of this gradient projection and proposition.

4 Dual Formulation

The projection onto the feasible set $\mathcal{F}$ is the solution to

$$\begin{aligned} \bar{z}_P(\alpha) \quad \equiv \quad & \underset{z \in \mathbb{R}^{2n}}{\arg\min} \quad \pi(z) \equiv \frac{1}{2}\|z - \bar{z}(\alpha)\|_2^2 \\ & \text{subject to} \quad Bz \geq 0. \end{aligned} \tag{6}$$

Since the projection must be solved at each iterate k, (6) must be solved easily and efficiently. Unfortunately solving this minimization problem can be difficult due to the linear constraints. The constraints $u, v \geq 0$ can be satisfied by zeroing negative entries; however, enforcing the constraint $\widetilde{W}z \geq 0$ is not as simple. What we wish to do is set up a related problem in which the optimal solution coincides with that of the projection problem (6). Therefore we propose solving the Lagrange dual problem associated with the primal problem (6). Since (6) is a minimization problem, its dual will be a maximization problem in which the optimal solution provides a lower bound on the minimum value of (6). In the case of this research solving the dual will lead to an easier optimization problem, and the minimum value of the primal problem equals the maximum value of the dual which will help us define the projection $\bar{z}_P(\alpha)$.

To formulate the dual problem we must first form the Lagrangian which involves incorporating the constraints into the function by augmenting the objective function with a weighted sum of the constraints. Therefore, the Lagrangian $\mathscr{L} \colon \mathbb{R}^{2n} \times \mathbb{R}^{3n} \to \mathbb{R}$ associated with (6) is given by

$$\mathscr{L}(z, \mu) = \frac{1}{2}\|z - \bar{z}(\alpha)\|_2^2 - \mu^T Bz,$$

with Lagrange multipliers $\mu \in \mathbb{R}^{3n}$. Next we formulate the Lagrange dual function $g \colon \mathbb{R}^{3n} \to \mathbb{R}$, defined as the minimum value of the Lagrangian over z and given by

$$g(\mu) = \inf_z \mathscr{L}(z, \mu).$$

The dual function yields a lower bound on the optimal value of (6) for any $\mu \geq 0$, i.e., $g(\mu) \leq \pi(z^\star)$. To show this we let $\widetilde{z}$ be a feasible point for (6). Therefore we have $\widetilde{z}$, $B\widetilde{z}$, $\pi(\widetilde{z}) \geq 0$. Now we assume $\mu \geq 0$. Then we have $\mu^T B\widetilde{z} \geq 0$. Since both terms in the Lagrangian are positive we have

$$\mathscr{L}(\widetilde{z}, \mu) = \frac{1}{2}\|\widetilde{z} - \bar{z}(\alpha)\|_2^2 - \mu^T B\widetilde{z} \leq \frac{1}{2}\|\widetilde{z} - \bar{z}(\alpha)\|_2^2.$$

Hence,

$$g(\mu) = \inf_z \mathscr{L}(z, \mu) \le \mathscr{L}(\widetilde{z}, \mu) \le \pi(\widetilde{z}).$$

The best lower bound can be obtained by solving the following Lagrange dual function:

$$\begin{array}{ll} \text{maximize} & g(\mu) \\ \text{subject to} & \mu \ge 0. \end{array} \tag{7}$$

Since $g(\mu) = \inf_z \mathscr{L}(z, \mu)$ we wish to find the point z that would yield this infimum. To do so we calculate $\nabla_z \mathscr{L}(z, \mu) = 0$, from which we find

$$z = \bar{z}(\alpha) + B^T \mu. \tag{8}$$

Substituting (8) in to (7), the dual associated with the primal problem (6) becomes

$$\begin{array}{lll} \mu^\star & \equiv & \underset{\mu \in \mathbb{R}^{3n}}{\text{maximize}} \;\; g(\mu) \;\; = \;\; -\frac{1}{2}\mu^T B B^T \mu - \mu^T B \bar{z}(\alpha) \\ & & \text{subject to} \;\; \mu \ge 0. \end{array} \tag{9}$$

We say strong duality holds when for optimal μ^* and $\bar{z}_p(\alpha)$, $g(\mu^*) = \pi(\bar{z}_p(\alpha))$. This means the optimal value to the dual and primal problems are equal. Therefore if we have the optimal solution to the dual problem we can solve for the optimal argument to the primal problem. Generally strong duality does not hold; however, in this case the objective function $\pi(z)$ is convex and the constraints $Bz \ge 0$ are affine. Therefore the weaker version of Slater's Condition [2] holds and we have strong duality. Thus, the solution to (6) can then be defined from (8) as

$$\bar{z}_P(\alpha) \equiv \bar{z}(\alpha) + B^T \mu^\star$$

by solving (9), which is an easier optimization problem to solve than (6) because the constraints in (9) are simple bound constraints and those in (6) contain linear constraints.

5 Solving the dual

Now we have the projection $\bar{z}_p(\alpha)$ in terms of $\mu^\star$. Therefore we must solve the dual (9). To do this we reformulate the dual as a minimization problem and partition μ as $\mu = [\psi; \zeta]$, where $\psi \in \mathbb{R}^{2n}$ and $\zeta \in \mathbb{R}^n$. Thus (9) can be equivalently written as

$$\begin{array}{ll} \underset{\mu \in \mathbb{R}^{3n}}{\text{minimize}} & h(\mu) \equiv \frac{1}{2}\|\psi + \widetilde{W}^T \zeta + \bar{z}(\alpha)\|_2^2 - \frac{1}{2}\|\bar{z}(\alpha)\|_2^2 \\ \text{subject to} & \psi, \zeta \ge 0. \end{array} \tag{10}$$

We also see that $h(\mu)$ can be expanded to yield

$$h(\mu) = \Big(\frac{1}{2}\|\psi\|_2^2 + \psi^T \bar{z}(\alpha)\Big) + \zeta^T \widetilde{W} \psi + \Big(\|\zeta\|_2^2 + \zeta^T \widetilde{W} \bar{z}(\alpha)\Big).$$

Since we have partitioned μ, we see now that $h(\mu)$ is almost separable with respect to ψ and ζ. Only the $\zeta^T \widetilde{W} \psi$ term couples the variables. This makes component-wise minimization a reasonable approach to solving for μ because many terms would go to zero when differentiating. First we fix ζ and minimize with respect to ψ, and then fix the computed ψ and optimize for ζ. We currently do not have proof that this method converges; however, this process produces a sequence of monotonically decreasing objective function values (see below). Moreover, from numerical experience, only a few iterations of ψ and ζ are needed to produce an accurate estimate of $\bar{z}_p(\alpha)$. We now describe each step more explicitly.

Step 1. Given ζ^{j-1} from the previous iterate, solve

$$\begin{aligned} \psi^j \quad = \quad & \underset{\psi \in \mathbb{R}^{2n}}{\arg\min} \quad \frac{1}{2}\|\psi\|_2^2 + \psi^T \left(\bar{z}(\alpha) + \widetilde{W}^T \zeta^{j-1} \right) \\ & \text{subject to} \quad \psi \geq 0 \end{aligned} \tag{11}$$

which comes from the parts of $h(\mu)$ that do not go to zero when optimizing. By taking the derivative and setting equal to zero we find the solution is given by

$$\psi^j = \left[- \left(\bar{z}(\alpha) + \widetilde{W}^T \zeta^{j-1} \right) \right]_+ \tag{12}$$

where $[x]_+$ represents the positive part of x.
Step 2. Given ψ^j, solve

$$\begin{aligned} \zeta^j \quad = \quad & \underset{\zeta \in \mathbb{R}^{n}}{\arg\min} \quad \|\zeta\|_2^2 + \zeta^T \widetilde{W} \left(\bar{z}(\alpha) + \psi^j \right) \\ & \text{subject to} \quad \zeta \geq 0. \end{aligned} \tag{13}$$

By taking the derivative and setting equal to zero we find the solution is given by

$$\zeta^j = \frac{1}{2} \left[-\widetilde{W} \left(\bar{z}(\alpha) + \psi^j \right) \right]_+ . \tag{14}$$

Note that the sequence $\{h(\upsilon^j, \zeta^j)\}$ produced by this alternating procedure monotonically decreases since

$$h(\upsilon^j, \zeta^j) \;\geq\; h(\upsilon^{j+1}, \zeta^j) \;\geq\; h(\upsilon^{j+1}, \zeta^{j+1}). \tag{15}$$

The result from component-wise minimization must yield a feasible solution, meaning $f = \widetilde{W} z_j \geq 0$ where we define z_j from (8) as $z_j \equiv \bar{z}(\alpha) + B^T \mu^j$. Then

$$\widetilde{W} z_j = \widetilde{W} \left(\bar{z}(\alpha) + B^T \mu^j \right) = \widetilde{W} \bar{z}(\alpha) + \widetilde{W} \upsilon^j + 2\zeta^j$$

Substituting the expression for ζ^j from (14) yields

$$\widetilde{W} z_j = \widetilde{W} \bar{z}(\alpha) + \widetilde{W} \upsilon^j + \left[-\widetilde{W} \left(\bar{z}(\alpha) + \upsilon^j \right) \right]_+ = \left[\widetilde{W} \left(\bar{z}(\alpha) + \upsilon^j \right) \right]_+ \geq 0.$$

Therefore we may terminate at any iteration and have a feasible solution.

6 Numerical Experiments

We want to compare the results of the proposed Linearly Constrained Gradient Projection (LCGP) method to a similar method that does not enforce nonnegativity in the reconstruction. Thus we compare results to those of Gradient Projection for Sparse Reconstruction (GPSR) which is also uses gradient projection to find an estimate but allows for negative pixel values. To show the importance of a nonnegative reconstruction we also compare results with GPSR thresholded (GPSR-T) which zeroes the resulting negative pixel values of GPSR. In addition, we run an experiment to test how well a hybrid between algorithms affects our results. For this we initialize LCGP with the results of GPSR-T, which we call I-LCGP. Within the same time limit these algorithms will estimate the same image from the same observation.

The signal to be reconstructed corresponds to a 256×256 gray-scale image of craters on the planet Mercury [12]. From the true image $f^\star$ we create our own observations. We choose the blur operation R to be the same as that used in Experiment 2 of [8]. After the blur operation we add zero-mean Gaussian noise of variance $\sigma^2 = 25$. This means we add a matrix in which the values range from -5 to 5, the mean is zero, and values are normally distributed. In practice our observations are not always of high precision. Thus we quantize the result to an accuracy of $b = 3$ bits per pixel. Quantization limits the accuracy of our observations by reducing the number of colors used to represent the image, and in this case $b = 3$ bits per pixel is 8 colors. Each algorithm should have ample time to reconstruct the observation, and from experience we choose 10 seconds total for each algorithm. For I-LCGP, we first run GPSR for 5 seconds, threshold the result, initialize LCGP and run it for an additional 5 seconds. We also require an optimal τ value for each method which is determined by running each algorithm for the given time budget and minimizing the root mean square (RMS) defined as RMS $\equiv \|\widehat{f} - f^\star\|/\|f^\star\|$. For I-LCGP the optimal τ value is found first for GPSR over 5 seconds, then using that result we independently find the optimal τ for the initialized LCGP over 5 seconds. Table 1 shows the averaged resulting minimum RMS from each method from 10 different observations, which differed only in the Gaussian noise η. Figure 2 contains the true signal $f^\star$, a degraded observation, and a reconstruction from each method. The figure also includes a magnified region with a contrast-enhancing colormap to highlight the differences between the GPSR-T reconstruction and the LCGP reconstruction.

Ten Averaged Simulations

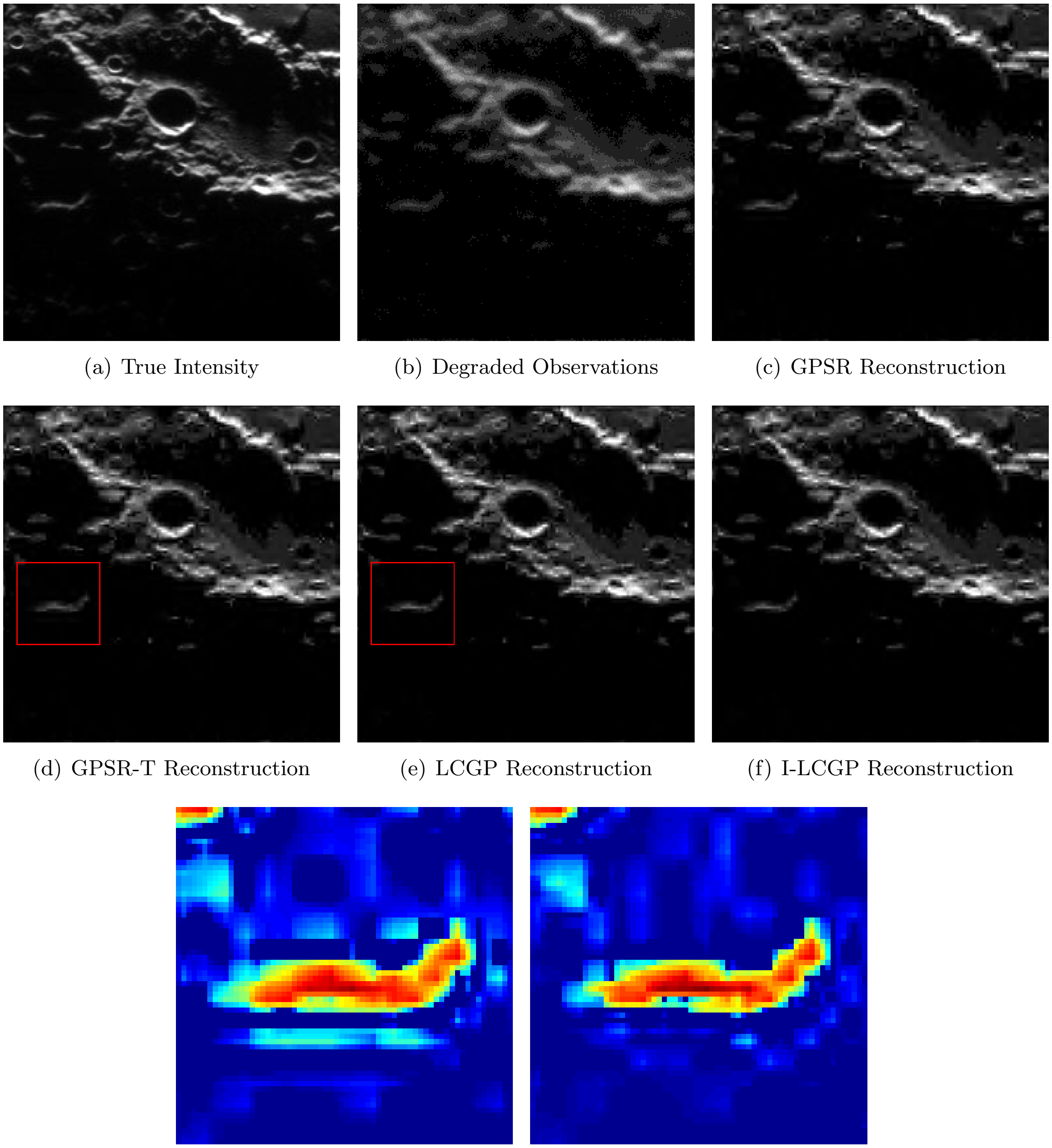
Method	RMS (%)
GPSR	26.08969
GPSR-T	24.86347
I-LCGP	24.00852
LCGP	23.87412

Table 1: Resulting average RMS values over ten simulations for each method considered, where RMS $\equiv \|\widehat{f} - f^\star\|/\|f^\star\|$.

From the table we see that GPSR-T yields an improved reconstruction over GPSR which highlights the importance of a nonnegative reconstruction. We also see that LCGP yields the lowest RMS value which shows that by adding constraints to the optimization problem to yield nonnegative solutions we see a greater increase in performance. The RMS value from the I-LCGP approach is between GPSR-T and LCGP, which shows that it is more effective to run just LCGP over the time budget. The differences between the different methods are subtle but by considering the image location shown in Figures 2(g) and 2(h) we see that the GPSR-T solution contains blocking artifacts near boundaries. The LCGP reconstruction however has fewer blocking artifacts, and captures regions of low intensity more accurately.

7 Video Expansion

Now that we have an established method to reconstruct a single image we wish to extend this method to a video comprised of a series of images. The naive solution to such a problem would be



(a) True Intensity (b) Degraded Observations (c) GPSR Reconstruction

(d) GPSR-T Reconstruction (e) LCGP Reconstruction (f) I-LCGP Reconstruction

(g) Cropped GPSR Solution (h) Cropped LCGP Solution

Figure 2: Results of our numerical experiments. Here we show (a) the true intensity f, (b) the degraded observations y. (c) the reconstruction using GPSR, (d) the reconstruction using GPSR-T. (e) the reconstruction with the proposed LCGP method, (f) the reconstruction using I-LCGP. The images (g) and (h) crop to a particular region (in red square) in the reconstructions (d) and (e) using a different colormap to highlight the differences. Note the blocking and spurious artifacts near boundaries and in regions of near-zero intensity present in the GPSR solution that are less pronounced in the LCGP solution.

to reconstruct each frame individually. In this case we still solve the ℓ_2-ℓ_1 minimization problem

$$\begin{aligned}\widehat{f}_t \quad &\equiv \quad \underset{f_t \in \mathbb{R}^n}{\arg\min} \;\; \frac{1}{2}\|y_t - R_t f_t\|_2^2 + \tau\|f_t\|_1 \\ &\qquad \text{subject to} \quad f_t \geq 0.\end{aligned} \tag{16}$$

Here t denotes the current frame, and just as in Sec. 2 the problem can be rewritten as

$$\begin{aligned}\widehat{\theta}_t \quad &\equiv \quad \underset{\theta_t \in \mathbb{R}^n}{\arg\min} \quad \frac{1}{2}\|y_t - R_t W \theta_t\|_2^2 + \tau\|\theta\|_1 \\ &\qquad \text{subject to} \quad W\theta_t \geq 0 \\ \widehat{f}_t \quad &\equiv \quad W\widehat{\theta}_t.\end{aligned} \tag{17}$$

This leads us to estimating each frame by the previous method. However, we can improve upon our previous approach by making the assumption that the scene changes only small amounts from frame to frame. With this assumption we can take advantage of the correlation between frames. If the changes from frame to frame are small we have that $f_t^\star \approx f_{t+1}^\star$, and consequently, $\theta_t^\star \approx \theta_{t+1}^\star$. Therefore the solution to frame t is often a good approximation to the next frame. Thus we use our solution at frame t as the initialization for frame $t+1$, i.e. $\theta_{t+1}^0 = \widehat{\theta}_t$. This approach can be improved upon further by solving multiple frames simultaneously and again exploiting our assumption.

To solve two frames simultaneously we can set up the optimization problem

$$\begin{aligned}&\underset{\theta_t,\theta_{t+1}}{\text{minimize}} \;\; \frac{1}{2}\left\|\begin{bmatrix} y_t \\ y_{t+1}\end{bmatrix} - \begin{bmatrix} R_t & 0 \\ 0 & R_{t+1}\end{bmatrix}\begin{bmatrix} W & 0 \\ 0 & W\end{bmatrix}\begin{bmatrix}\theta_t \\ \theta_{t+1}\end{bmatrix}\right\|_2^2 + \tau\left\|\begin{bmatrix}\theta_t \\ \theta_{t+1}\end{bmatrix}\right\|_1 \\ &\text{subject to} \quad W\theta_t, W\theta_{t+1} \geq 0.\end{aligned}$$

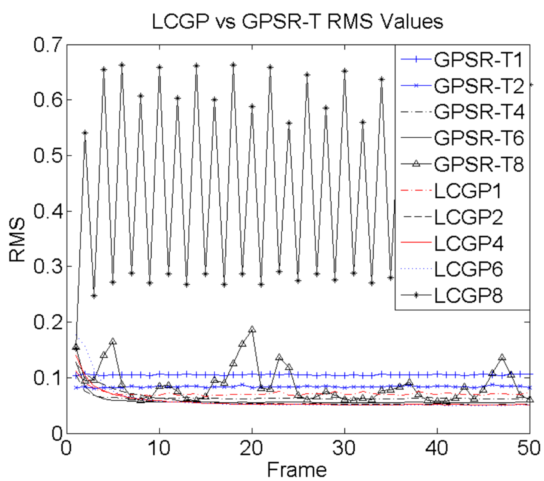
However, this formulation is separable and is equivalent to solving (17). Instead we want to solve some problem that relates the solutions $\theta_t^\star$ and $\theta_{t+1}^\star$ and takes advantage of our assumption that the images change slightly from frame to frame. Since $\theta_t^\star \approx \theta_{t+1}^\star$, the frame difference $\Delta\theta_t^\star = \theta_{t+1}^\star - \theta_t^\star$ must be much more sparse than either frame. Therefore ℓ_2-ℓ_1 minimization would be well-suited here by expressing the variables as the frame differences. Solving for two frames simultaneously we then formulate the problem to solve for $\theta_t^\star$ and the difference $\Delta\theta_t^\star$ instead of $\theta_t^\star$ and $\theta_{t+1}^\star$. Now we solve the coupled optimization problem

$$\begin{aligned}\begin{bmatrix}\widehat{\theta}_t \\ \Delta\widehat{\theta}_t\end{bmatrix} \quad &= \quad \underset{\theta_t,\Delta\theta_t}{\arg\min} \; \frac{1}{2}\left\|\begin{bmatrix} y_t \\ y_{t+1}\end{bmatrix} - \begin{bmatrix} R_t & 0 \\ 0 & R_{t+1}\end{bmatrix}\begin{bmatrix} W & 0 \\ W & W\end{bmatrix}\begin{bmatrix}\theta_t \\ \Delta\theta_t\end{bmatrix}\right\|_2^2 + \tau_1\|\theta_t\|_1 + \tau_2\|\Delta\theta_t\|_1 \\ &\qquad \text{subject to} \quad W\theta_t \geq 0, \; W(\theta_t + \Delta\theta_t) \geq 0\end{aligned} \tag{18}$$

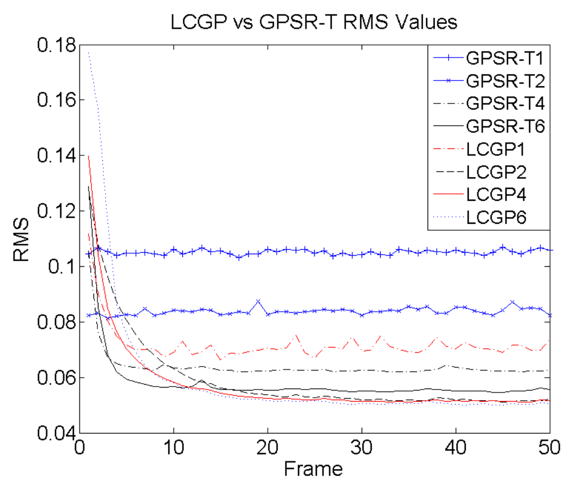
We notice now that there are two τ values. We know $\Delta\theta_t$ will be more sparse than θ_t. Recall that higher τ values promote more sparsity, thus $\tau_2 > \tau_1 > 0$. With our estimates $\widehat{\theta}_t$ and $\Delta\widehat{\theta}_t$ from (18) we initialize $\theta_{t+1}^0 \equiv \widehat{\theta}_t + \Delta\widehat{\theta}_t$ which is a very accurate estimate to the solution $\theta_{t+1}^\star$. Then we solve for $\theta_{t+1}^\star$ and $\Delta\widehat{\theta}_{t+1}^\star = \theta_{t+2}^\star - \theta_{t+1}^\star$. This approach can be extended to solving for more than two frames but, as the number of frames solved for simultaneously increases so does the computational cost. Through the use of gradient projection and the dual, (18) and higher frame formulations can be solved similarly to the single image case. Although not shown here, the multiframe formulations can be solved nontrivially by using the approaches in Sec. 3 through Sec. 5.

8 Numerical Results for Video Expansion

These experiments are similar to those of the single image case in Sec. 6. We compare the results of the proposed LCGP method with those of GPSR and GPSR-T. Each algorithm reconstructs the same observations with the same time constraints. The video considered is from the Apollo 11 moonwalk [9]. We choose this video because there is no camera movement, and the little movement in the scene is from astronauts walking in the distance. For this experiment we take a 60 frame segment from the video but estimate only the first 50. We choose to do this because when reconstructing frame 50 the multiframe algorithms require further frames to work with. The frames to be reconstructed consist of a 128×256 cropped section in gray scale. The blurring operation R used to create our observation is the same as in the experiments of Sec. 6. Although in the problem formulation of the previous section each observation has a different blurring operation R_t, in this experiment all observations use the same blurring operator R. Next, zero mean Gaussian noise is added with variance $\sigma^2 = 16$. Thus the noise matrix we add has zero mean, is normally distributed and contains values that range from -4 to 4. Each algorithm is allotted a maximum time $t = 3$ seconds to reconstruct a group of frames. We choose this time limit because using a multiframe method on a single group of frames will not yield a very accurate reconstruction, as seen by the initial values of Figure 3 (b). However, with the initialization step future frame reconstructions become more accurate than earlier reconstructions. In these experiments we use each algorithm to reconstruct 1, 2, 4, 6, and 8 frames simultaneously. From experience we notice that these methods on average stabilize around frame 20, thus we choose the optimal τ values to be those that yield the minimum average RMS over frames 20 to 50.



(a) LCGP vs GPSR-T RMS values, 1,2,4,6,8 frame methods

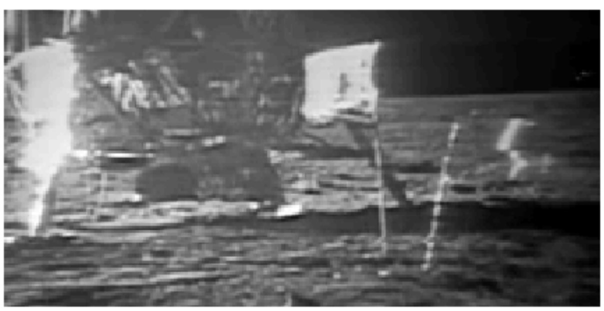


(b) LCGP vs GPSR-T RMS values, 1,2,4,6 frame methods

Figure 3: Results of our numerical experiments for multiframe methods. Here we show (a) RMS evolution at each time frame for 1,2,4,6 and 8-frame GPSR-T and LCGP methods, (b) RMS evolution at each time frame for 1,2,4 and 6-frame GPSR-T and LCGP methods.

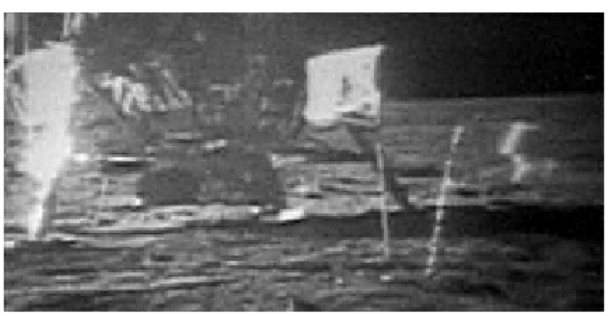
Figure 3 shows us the resulting RMS graphs of each method, and we see that the proposed LCGP methods which enforce nonnegativity constraints are an improvement over the GPSR methods. Not shown are the GPSR results, only the GPSR-T results. This is because GPSR-T is an improvement over GPSR and always yields lower RMS values. Figure 3 (a) includes the 8-frame methods and Figure 3 (b) does not. With the 3 second time limit the 8-frame methods cannot perform enough iterations to produce an accurate reconstruction, and the result is a disproportionate increase in the RMS seen in Figure 3 (a). To get a closer look at the other methods we include Figure 3 (b) which does not include the 8-frame methods. The earlier frame reconstructions of the multiframe methods also lack sufficient iterations and the result is the high initial RMS seen in the graph. We note that LCGP performs fewer iterations than GPSR within a given time limit and thus breaks down faster than GPSR as seen in the 8-frame case in Figure 3 (a). This effect is also noticed in the case of LCGP6 (6 frame method for LCGP) in which the rate of decline falls between that of LCGP2 and LCGP4. From the graph of Figure 3 (b) we see that with the exception of GPSR2-T, all the multiframe methods yield a significant improvement over both single frame methods. Generally the LCGP methods have higher initial RMS values than the GPSR-T methods due to LCGP having a higher computational cost and fewer overall iterations. However, by frame 20 all methods have stabilized and we see that LCGP2, LCGP4, and LCGP6 overtake all other methods and have the lowest RMS values.

Figure 4 compares the reconstruction of the 50th frame of GPSR6-T and LCGP4. We choose 6-frame for GPSR and 4-frame for LCGP to compare because they have the lowest RMS values for the 50th frame of the GPSR-T and LCGP algorithms. Similarly to the single image case the differences between each reconstruction are not immediately noticeable but by focusing on a particular region we can see that the GPSR reconstruction contains more noise than the LCGP reconstruction, particularly in regions of low intensity.



(a) True Intensity

(b) Degraded Observations



(c) GPSR6-T Reconstruction

(d) LCGP4 Reconstruction

Figure 4: Reconstructions of our numerical experiments for multiframe methods. Here we show (a) the true intensity f, (b) the degraded observations y, (c) the reconstruction using GPSR6-T, (d) the reconstruction with the proposed LCGP4 method. We notice that GPSR6-T contains noise that is less pronounced in LCGP4 and can be most easily seen in regions of near-zero intensity.

9 Conclusion

In this research we present a method for nonnegative, sparse reconstruction of an image through solving a constrained ℓ_2-ℓ_1 minimization problem, and extend this method to a video. Incorporating the nonnegativity constraints into the problem, then solving through the use of gradient projection and the dual has proven to yield accurate reconstructions. Our numerical results in Sec. 6 show that LCGP reconstructions are an improvement over GPSR, another gradient projection method that does not take into account the nonnegative pixel values of images. With the assumption that scenes change only small amounts from frame to frame, the developed extension takes advantage of the correlation between frames to improve upon naively solving each frame individually, despite additional computational costs. Our numerical results in Sec. 8 show that by solving multiple frames simultaneously and expressing the variables as frame differences we have a significant improvement over the single frame solution. In addition, by incorporating nonnegativity constraints we can further improve upon the multiframe reconstructions. This research highlights the improvements that can be achieved by attaining a nonnegative recosntruction of an image.[8]

10 Acknowledgements

I would like to thank Professor Roummel Marcia for his guidance and supervision of this research project and Arnold Kim for serving as an additional faculty member overseeing this research project. I would also like to thank Zachary Harmany and Professor Rebecca Willett from Duke University for the opportunity to work with them on this research. This work was supported by NSF Grants DMS-08-11062/DMS-09-65711. Portions of this work were presented at the IEEE International Conference on Image Processing, Hong Kong, China, September 2010, and at the IEEE International Conference on Acoustic, Speech, and Signal Processing, Prague, Czech Republic, May 2011.

References

[1] J. Barzilai and J. Borwein, Two-point step size gradient methods, IMA Journal of Numerical Analysis, 8 (1988), pp. 141–148.

[2] S. Boyd and L. Vandenberghe, Convex Optimization, Cambridge University Press, 1st ed., 2004.

[3] P. H. Calamai and J. J. Moré, Projected gradient methods for linearly constrained problems, Mathematical Programming, 39 (1987), pp. 93–116.

[4] E. J. Candès, J. Romberg, and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, IEEE Transactions on Information Theory, 52 (2006), pp. 489–500.

[5] E. J. Candès and T. Tao, The dantzig selector: Statistical estimation when p is much larger than n, IEEE Trans. Inform. Theory, 52 (2006), p. 40364048.

[6] B. Efron, T. Hastie, I. Johnstone, and R. Tibshirani, Least angle regression, The Annals of Statistics, 32 (2004), pp. 407–499.

[7] M. A. T. Figueiredo and R. D. Nowak, An EM algorithm for wavelet-based image restoration, IEEE Transactions on Image Processing, 12 (2003), pp. 906–916.

[8] M. A. T. Figueiredo, R. D. Nowak, and S. J. Wright, Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems, IEEE J. Sel. Top. Sign. Proces.: Special Issue on Convex Optimization Methods for Signal Processing, 1 (2007), pp. 586–597.

[9] Goddard Space Flight Center NASA, Pia12272: Apollo 11. `http://www.archive.org/details/PartiallyRestoredVideoBuzzAldrinCarryingExperimentPackages`, 2009,.

[10] J. Haupt and R. Nowak, Signal reconstruction from noisy random projections, IEEE Trans. Inform. Theory, 52 (2006), p. 40364048.

[11] S. Mallat and Z. Zhang, Matching pursuits with time-frequency dictionaries, IEEE Transactions on Signal Processing, 41 (1993), pp. 3397 – 3415.

[12] NASA/Johns Hopkins University Applied Physics Laboratory/Carnegie Institution of Washington, Pia12272: Seeing double? `http://photojournal.jpl.nasa.gov/tiff/PIA12272.tif`, 2009.