Trust-Region Methods for Nonconvex Sparse Recovery Optimization

Lasith Adhikari and Roummel F. Marcia
Applied Mathematics,
University of California, Merced,
Merced, CA 95343 USA

Jennifer B. Erway and Robert J. Plemmons
Departments of Mathematics and Computer Science,
Wake Forest University,
Winston-Salem, NC 27109 USA

***Abstract*—We solve the ℓ_2-ℓ_p sparse recovery problem by transforming the objective function into an unconstrained differentiable function and apply a limited-memory trust-region method. Unlike gradient projection-type methods, which uses only the current gradient, our approach uses gradients from previous iterations to obtain a more accurate Hessian approximation. Preliminary numerical experiments with simulated compressive sensing 1D data are provided to illustrate that our proposed approach eliminates spurious solutions more effectively while improving the computational time to converge in comparison to standard approaches.**

I. Introduction

Sparse signal recovery methods form a ubiquitous component of modern compressed sensing technology. Compressed sensing is essentially based on the fact that most signals and images contain only a small amount of crucial information. By intelligently discarding or ignoring unnecessary data by software or physical sensors, one can reduce the data set size while retaining sufficient information to faithfully reconstruct the original signal or image. Basically, one can design an efficient sensing or sampling scheme by capturing the useful information in a sparse signal and compressing it into a small amount of data. This leads to the need for solving a highly underdetermined inverse problem by appropriate optimization methods for recovering the signal. Much of the early mathematical work on sparse signal recovery can be found in papers such as those involving Candes, Gehm, Fowler, Marcia, Romberg, Tao, Willett, Zhang, etc., see e.g., [1], [2], [3], [4], [5], to name just a few. Much of this work was based on the use of the ℓ_1 norm in solving the resulting constrained optimization problem. Our interest here is in applying algorithms based on trust-region methods to sparse signal recovery using an ℓ_p regularization norm, $0 < p < 1$, where the objective function is nonconvex, thus extending our work in [6].

Currently, there is much interest in both image reconstruction methods from applying random projections for sparsifying image data, and in developing methods for signal recovery from compressively sensed data using physical cameras or sensors. In the first case, the random projection problem is a linear inverse problem stated as, $y = Af$, where $f \in \mathbb{R}^n$ is the true signal to recover, $A \in \mathbb{R}^{k\times n}$ is the projection matrix, with independent and identically distributed (i.i.d.) random entries, and $y \in \mathbb{R}^k$ is the observed signal. Sometimes rows of A can be orthogonalized and normalized, see e.g., [7], [8], [9]. A theorem by Candès, Rumberg and Tao [1] states that if $k \geq cm \log n$, where c is a constant and m is the number of nonzeros in f, then f is the unique solution to the convex optimization problem,

$$\min \|f\|_1 \quad \text{subject to } \|Af - y\|_2^2 < \epsilon, \tag{1}$$

for an overwhelming percentage of row vector sets for A with cardinality k. A standard way to approach solving (1) is to solve the ℓ_1 regularized convex optimization problem

$$\underset{f\in\mathbb{R}^n}{\text{minimize}} \ \frac{1}{2}\|Af - y\|_2^2 + \tau\|f\|_1. \tag{2}$$

In this paper we replace the ℓ_1 norm by the ℓ_p norm, $0 < p < 1$, which has been shown to have advantages in recovering sparse signals in compressive sensing, see, e.g., [10].

In the compressive imaging literature, there have been efforts on applying random projection methods based on physical sensors with coded apertures, see e.g., Marcia and Willett [4]. Also, the double-disperser coded-aperture snapshot imager (DD-CASSI) [2], and the single-disperser coded-aperture snapshot imager (SD-CASSI) [5], [11], can snapshot a hyerspectral image 3D dataset, and both systems can be formulated as inverse problems for recovering sparse signals or images. In addition, Zhang, Prasad and Plemmons [12] have processed compressive sensing data obtained with the use of spatial light modulator (SLM) data encoders for 4D spectro-polarimetric data. Such data will be used in our future tests.

The paper is organized as follows. In Section II, we present the sparse recovery problem defined by the nondifferentiable p-norm regularization, and formulate the original problem as a differentiable unconstrained optimization problem. In Section III, we describe a quasi-Newton trust-region method to solve problem. Numerical results on simulated 1D sparse signal recovery data are provided to illustrate the effectiveness of the proposed method are in Section IV. Finally, there are conclusions and directions of future research in Section V.

II. Problem Formulation

This paper concerns solving the sparse recovery problem

$$\underset{f\in\mathbb{R}^n}{\text{minimize}} \ \frac{1}{2}\|Af - y\|_2^2 + \tau\|f\|_p^p, \tag{3}$$

where $A \in \mathbb{R}^{\tilde{m}\times\tilde{n}}$, $f \in \mathbb{R}^{\tilde{n}}$, $y \in \mathbb{R}^{\tilde{m}}$, $\tilde{m} \ll \tilde{n}$, $0 < p < 1$, and $\tau > 0$ is a constant regularization parameter. By letting $f = u - v$, where $u, v \geq 0$, we write (3) as the constrained but differentiable optimization problem

$$\begin{array}{ll} \underset{u,v\in\mathbb{R}^{\tilde{n}}}{\text{minimize}} & \frac{1}{2}\|A(u-v)-y\|_2^2 + \tau\sum_{j=1}^{\tilde{n}}((u)_j+(v)_j)^p \\ \text{subject to} & u, v \geq 0, \end{array} \quad (4)$$

where $(\cdot)_j$ denotes the jth element of the vector $(\cdot)$. We transform (4) into an unconstrained optimization problem by the change of variables

$$(u)_j = \log(1+e^{(\tilde{u})_j}) \quad \text{and} \quad (v)_j = \log(1+e^{(\tilde{v})_j}),$$

where $(\tilde{u})_j, (\tilde{v})_j \in \mathbb{R}$ for $1 \leq j \leq \tilde{n}$ (see [13], [14], [6]). With these definitions, u and v are guaranteed to be non-negative. Thus, (4) is equivalent to the following minimization problem:

$$\begin{aligned} \min_{\tilde{u},\tilde{v}\in\mathbb{R}^{\tilde{n}}} \Phi(\tilde{u},\tilde{v}) \triangleq \frac{1}{2}\sum_{i=1}^{\tilde{m}}\left[\left\{\sum_{j=1}^{\tilde{n}}(A)_{i,j}\log\left(\frac{1+e^{(\tilde{u})_j}}{1+e^{(\tilde{v})_j}}\right)\right\}-(y)_i\right]^2 \\ + \tau\sum_{j=1}^{\tilde{n}}\left(\log(1+e^{(\tilde{u})_j})+\log(1+e^{(\tilde{v})_j})\right)^p, \end{aligned} \quad (5)$$

where $(\cdot)_{i,j}$ indicates the (i,j)th element of the matrix $(\cdot)$. To compute the gradient of $\Phi(\tilde{u},\tilde{v})$, we define the following: let $\tilde{w}_1, \tilde{w}_2 \in \mathbb{R}^{\tilde{n}}$ with

$$\begin{aligned} (\tilde{w}_1)_i &= \log((1+e^{(\tilde{u})_i})/(1+e^{(\tilde{v})_i})), \\ (\tilde{w}_2)_i &= \left(\log(1+e^{(\tilde{u})_i})+\log(1+e^{(\tilde{v})_i})\right)^{p-1}, \end{aligned}$$

and let

$$\Phi_1(\tilde{u},\tilde{v}) = A^TA\tilde{w}_1 \quad \text{and} \quad \Phi_2(\tilde{u},\tilde{v}) = \tau p\tilde{w}_2.$$

Furthermore, let $D_{\tilde{u}}, D_{\tilde{v}} \in \mathbb{R}^{\tilde{n}\times\tilde{n}}$ be diagonal matrices with

$$(D_{\tilde{u}})_{i,i} = \frac{e^{(\tilde{u})_i}}{1+e^{(\tilde{u})_i}} \quad \text{and} \quad (D_{\tilde{v}})_{i,i} = \frac{e^{(\tilde{v})_i}}{1+e^{(\tilde{v})_i}},$$

Then the gradient of the function $\Phi(\tilde{u},\tilde{v})$, i.e., $\nabla\Phi = [\nabla_{\tilde{u}}\Phi; \nabla_{\tilde{v}}\Phi]$ is given by

$$\begin{aligned} \nabla_{\tilde{u}}\Phi &= D_{\tilde{u}}\left[\ \Phi_1(\tilde{u},\tilde{v}) - A^Ty + \Phi_2(\tilde{u},\tilde{v})\right], \\ \nabla_{\tilde{v}}\Phi &= D_{\tilde{v}}\left[-\Phi_1(\tilde{u},\tilde{v}) + A^Ty + \Phi_2(\tilde{u},\tilde{v})\right]. \end{aligned}$$

We propose solving (5) using a limited-memory quasi-Newton trust-region optimization approach, which we describe in the next section.

Related work. In previous work [6], we developed quasi-Newton trust-region methods for sparsity recovery. Quasi-Newton methods have been previously shown to be effective for sparsity recovery problems (see e.g., [15], [16], [17], [18]). Trust-region methods have also been implemented for sparse reconstruction (see e.g., [19], [20]). Our approach is novel in the transformation of the sparse recovery problem to a differentiable unconstrained minimization problem and in the use of eigenvalues for efficiently solving the trust-region subproblem. Furthermore, we incorporate non-convex ℓ_p norms to bridge the the sparsity promoting ℓ_1 norm and the ℓ_0 counting semi-norm. Non-convex regularizers have been previously investigated in [21], [22], [23], [24], [25].

Notation. Throughout this paper, we denote the identity matrix by I, with its dimension dependent on the context.

III. Methodology

In this section, we begin by describing quasi-Newton matrices and then demonstrate their use in a trust-region method to solve (5).

A. Limited-Memory Quasi-Newton Matrices

Given a continuously-differentiable function $\Phi(x)$ as in (5), i.e., $x = [\tilde{u},\tilde{v}]^T \in \mathbb{R}^n$, and a sequence of iterates $\{x_k\}$, quasi-Newton matrices can be used to approximate $\nabla^2\Phi(x)$. Traditionally, a quasi-Newton matrix is generated by updating an initial matrix B_0 using a sequence of pairs $\{(s_k, y_k)\}$, where

$$s_k \triangleq x_{k+1} - x_k \quad \text{and} \quad y_k \triangleq \nabla\Phi(x_{k+1}) - \nabla\Phi(x_k).$$

One of the most widely-used update formulae is the Broyden-Fletcher-Goldfab-Shanno (BFGS) update given by

$$B_{k+1} \triangleq B_k - \frac{1}{s_k^TB_ks_k}B_ks_ks_k^TB_k + \frac{1}{y_k^Ts_k}y_ky_k^T, \quad (6)$$

provided $y_k^Ts_k \neq 0$ and B_0 is a symmetric positive-definite matrix. For simplicity, B_0 is often taken to be a scalar multiple of the identity matrix, i.e., $B_0 = \gamma I$ for some $\gamma > 0$. In large-scale optimization, so-called *limited-memory* BFGS (L-BFGS) matrices can be formed that make use of only the m most-recently computed pairs $\{(s_k, y_k)\}$. In practice, m is small, e.g., $m < 10$ (see [26]); the primary advantage of limited-memory quasi-Newton matrices is to place an upper bound on memory and storage requirements. Provided B_0 is a symmetric positive-definite matrix and $y_k^Ts_k > 0$, the sequence of matrices $\{B_k\}$ obtained using the BFGS update is symmetric positive definite. Often in practice, the condition $y_k^Ts_k > 0$ can be enforced by either skipping updates when $y_k^Ts_k \leq 0$ or using a *Wolfe line search* to help generate the sequence $\{x_k\}$ (for more details, see, e.g., [27]). For the proposed trust-region method, $\nabla^2\Phi(x)$ is approximated by an L-BFGS matrix.

It is well-known that L-BFGS matrices exhibit a *compact formulation*, i.e., a L-BFGS matrix B_{k+1} can be written as

$$B_{k+1} = B_0 + \begin{bmatrix} \\ \Psi_k \\ \\ \end{bmatrix}\begin{bmatrix} \ M_k \ \end{bmatrix}\begin{bmatrix} \quad \Psi_k^T \quad \end{bmatrix}$$

for some $\Psi_k \in \mathbb{R}^{n\times 2(k+1)}$ and $M_k \in \mathbb{R}^{2(k+1)\times 2(k+1)}$. In fact, Byrd et al. [26] show that for L-BFGS matrices,

$$\Psi_k = \begin{bmatrix} B_0S_k & Y_k \end{bmatrix} \text{ and } M_k = -\begin{bmatrix} S_k^TB_0S_k & L_k \\ L_k^T & -D_k \end{bmatrix}^{-1},$$

with

$$\begin{aligned} S_k &\triangleq \begin{bmatrix} s_0 & s_1 & s_2 & \cdots & s_k \end{bmatrix} \in \mathbb{R}^{n\times(k+1)}, \\ Y_k &\triangleq \begin{bmatrix} y_0 & y_1 & y_2 & \cdots & y_k \end{bmatrix} \in \mathbb{R}^{n\times(k+1)}, \end{aligned}$$

and L_k is the strictly lower triangular part, U_k is the strictly upper triangular part, and D_k is the diagonal part of

$$S_k^T Y_k = L_k + D_k + U_k.$$

It is important to note that $S_k^T Y_k$ is a small matrix in the limited-memory setting, i.e., $S_k^T Y_k \in \mathbb{R}^{(k+1)\times(k+1)}$ with $k \ll n$. The compact representation for an L-BFGS matrix will be used in the trust-region method for solving (5).

B. Trust-Region Methods

Trust-region methods are one of two important classes of methods for unconstrained optimization (see, e.g., [27], [28]). Basic trust-region methods generate a sequence of iterates $\{x_k\}$ by the relation $x_{k+1} = x_k + p_k$, where p_k is an approximate solution to the *trust-region subproblem* given by

$$\begin{array}{rcll} p_k & = & \underset{p\in\mathbb{R}^n}{\arg\min} & q_k(p) \triangleq g_k^T p + \frac{1}{2} p^T B_k p \qquad (7) \\ & & \text{subject to} & \|p\|_2 \leq \delta_k, \end{array}$$

where $g_k \triangleq \nabla\Phi(x_k)$, B_k is an approximation to $\nabla^2\Phi(x_k)$, and δ_k is a given positive constant. At the end of each trust-region iteration, the *trust-region radius* δ_k is used to update δ_{k+1}; depending on how well the quadratic model predicted actual decreases in the function $\Psi(x)$ from x_k to x_{k+1}, the trust-region radius is possibly increased or decreased for the next iteration. To solve (5), B_k is taken to be an L-BFGS matrix.

Generally speaking, computing an approximate solution to the trust-region subproblem is the computational bottleneck for most trust-region methods. While trust-region subproblems can be defined using any norm, there is an important advantage in using the Euclidean two-norm: There are well-known optimality conditions for a global solution to (7) (see [29], [30]). These optimality conditions allow one to monitor how close iterates for solving the subproblem are to a global solution; they have also inspired algorithms that solve (7) by explicitly trying to satisfy the optimality conditions. There are two important advantages in taking B_k to be an L-BFGS matrix: (i) the trust-region subproblem is convex and (ii) B_k has structure that can be exploited to solve (7) efficiently.

C. Solving the Trust-Region Subproblem

To solve the trust-region subproblem when B_k is an L-BFGS matrix, we use the method described in [6]. This method solves each trust-region subproblem to high accuracy using the optimality conditions for a global solution to (7). Tailored to the case when B_k is positive definite, the optimality conditions in [29] and [30] are given in the following theorem:

Theorem 1. *Let δ be a positive constant. A vector p^* is the unique global solution of the trust-region subproblem (7) if and only if $\|p^*\|_2 \leq \delta$ and there exists a unique $\sigma^* \geq 0$ such that*

$$(B + \sigma^* I)p^* = -g \quad \textit{and} \quad \sigma^*(\delta - \|p^*\|_2) = 0. \qquad (8)$$

We begin by transforming the optimality equations (8) using the spectral decomposition of B_k. As shown in [6], the spectral decomposition of B_k can be obtained efficiently using the compact formulation of B_k. We briefly review this here; for more details, see [6]. If $B_k = B_0 + \Psi M \Psi^T$ and $\Psi = QR$ is the "thin" QR factorization of Ψ, then

$$B_k = \gamma I + QRMR^T Q^T,$$

when $B_0 = \gamma I$ and $\gamma > 0$. Since RMR^T is a small $2k \times 2k$ matrix, its spectral decomposition $V\hat{\Lambda}V^T$ can be quickly computed. Then, letting $P \triangleq [\ QV \quad (QV)^\perp] \in \mathbb{R}^{n\times n}$ such that $P^T P = PP^T = I$, the spectral decomposition of B_k is given by

$$B_k = P\Lambda P^T, \text{ where } \Lambda \triangleq \begin{bmatrix} \Lambda_1 & 0 \\ 0 & \Lambda_2 \end{bmatrix} = \begin{bmatrix} \hat{\Lambda} + \gamma I & 0 \\ 0 & \gamma I \end{bmatrix}, \quad (9)$$

where $\Lambda_1 \in \mathbb{R}^{2k\times 2k}$,, and $\Lambda_2 = \gamma I_{n-2k}$. Using the spectral decomposition of B_k, the optimality equations (8) become

$$\begin{aligned} (\Lambda + \sigma^* I)v^* &= -\tilde{g} && (10) \\ \sigma^*(\|v^*\|_2 - \delta) &= 0 && (11) \\ \|v^*\|_2 &\leq \delta, && (12) \end{aligned}$$

for some scalar $\sigma^* \geq 0$ and $v^* = P^T p^*$, where p^* is the global solution to (7).

For practical considerations, it is important to order the eigenvalues found in $\hat{\Lambda}$ (and their associated eigenvectors in QU) in increasing order. In [6], it is shown how to compute the solution v^* to (10) and p^* the solution to (7) *by formula* once σ^* is known. In the worst-case scenario, σ^* can be efficiently found using Newton's method applied to a special function such that starting from $\sigma^{(0)} = 0$ the Newton iterates $\{\sigma^{(i)}\}$ converge monotonically and quadratically to σ^*; in the other scenario, $\sigma^* = 0$ is optimal. (For details regarding the Newton iteration to find σ^* and then computing p^* see [6].)

Algorithm 1 details the quasi-Newton trust-region method. For details on the subproblem solver, see [6, Algorithm 1].

ALGORITHM 1: TrustSpa-ℓ_p

Define parameters: m, $0 < \tau_1 < 0.5$, $0 < \varepsilon$;
Initialize $x_0 \in \mathbb{R}^n$ and compute $g_0 = \nabla\Phi(x_0)$;
Let $k = 0$;
while not converged
  if $\|g_k\|_2 \leq \varepsilon$ **then done**
  Find p_k that solves (7) using [6, Algorithm 1];
  Compute $\rho_k = (\Psi(x_k + p_k) - \Psi(x_k))/q_k(p_k)$;
  Compute g_{k+1} and update B_{k+1};
  if $\rho_k \geq \tau_1$ **then**
    $x_{k+1} = x_k + p_k$;
  else
    $x_{k+1} = x_k$;
  end if
  Compute trust-region radius δ_{k+1};
  $k \leftarrow k + 1$;
end while

IV. Numerical Results

We demonstrate the effectiveness of the proposed method, called TrustSpa-ℓ_p, by reconstructing a sparse signal (in the canonical basis) of size 4,096 with 160 randomly assigned nonzeros with maximum amplitude ± 1 (see Fig. 1(a)). The system matrix (A) is randomly generated with samples from a standard Gaussian distribution, which linearly projects the true signal (f^*) to the low-dimensional observations y (see Fig. 1(b)). These observations are corrupted by 5% of Gaussian noise, where the noise level $(\%) = 100 \cdot \|Af^* - y\|_2 / \|y\|_2$.

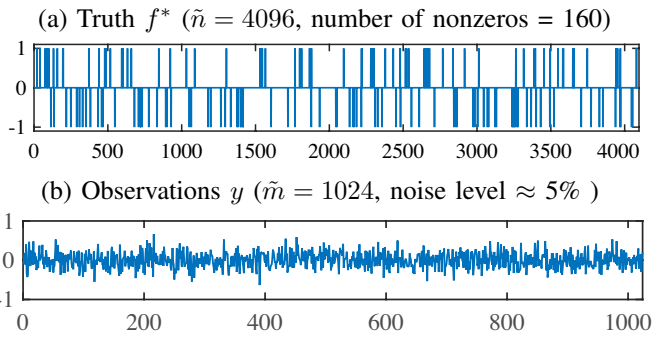

Fig. 1. Experimental setup: (a) True signal f^* of size 4,096 with 160 $\pm$ spikes, (b) low-dimensional observations y with 5% Gaussian noise. Noise level $(\%) = 100 \cdot \|Af^* - y\|_2 / \|y\|_2$.

We implemented the proposed TrustSpa-ℓ_p method in MATLAB R2015a using a PC with Intel Core i7 2.8GHz processor with 16GB memory. The results are compared with the widely-used Gradient Projection for Sparse Reconstruction (GPSR) method [31] without the debiasing option and the more recent TrustSpa-ℓ_1 method [6]. In these experiments, all the methods are initialized at the same starting point, i.e., zero and terminate if the relative objective values do not significantly change, i.e, $|\Phi(x_{k+1}) - \Phi(x_k)|/|\Phi(x_k)| \leq 10^{-8}$. The regularization parameter τ in (3) is tuned for the minimum mean-squared error (MSE = $\frac{1}{n}\|\hat{f} - f^*\|_2^2$, where $\hat{f}$ is an estimate of f^*).

Analysis. Using the compressive measurements y (see Fig. 1(b)), a single-trial optimal reconstruction using the proposed TrustSpa-ℓ_p for $p = 0.7$ is given by Fig. 2(a). The TrustSpa-ℓ_p reconstruction, $\hat{f}_{\text{TS-}\ell_p}$, has MSE 1.403×10^{-5}, while the TrustSpa-ℓ_1, $\hat{f}_{\text{TS-}\ell_1}$ (see Fig. 2(c) for a zoomed region of $\hat{f}_{\text{TS-}\ell_1}$) has MSE 9.347×10^{-5}, and the GPSR reconstruction (see [6] for the reconstruction) has MSE 1.624×10^{-4}. Specifically, notice that the $\hat{f}_{\text{TS-}\ell_1}$ has more spurious artifacts (shown in black in Fig. 2(c)) than the artifacts in $\hat{f}_{\text{TS-}\ell_p}$ (see black spikes in Fig. 2(b)). Quantitatively, $\hat{f}_{\text{TS-}\ell_p}$ has only 263 nonzero components greater than 10^{-6} in absolute value, while $\hat{f}_{\text{TS-}\ell_1}$ has 579 nonzeros.

In addition, we analyzed the discrepancy between the true and approximated amplitudes from each method using magnitude error plots (see Fig. 3). In particular, the magnitude of error of the $\hat{f}_{\text{TS-}\ell_p}$ is much closer to zero (see the red color spikes in Fig. 3(a) and (b)) than the magnitude of error of the GPSR reconstruction $\hat{f}_{\text{GPSR}}$ (see blue spikes in Fig. 3(a)) and

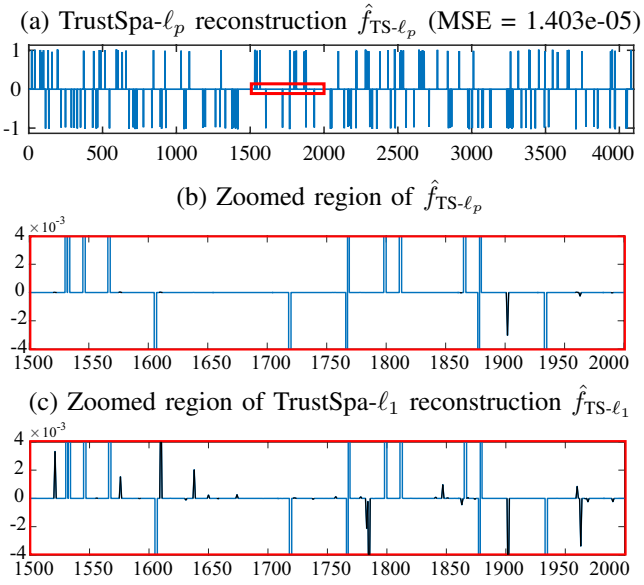

Fig. 2. (a) TrustSpa-ℓ_p reconstruction, $\hat{f}_{\text{TS-}\ell_p}$, with $p = 0.7$, (b) a zoomed region of $\hat{f}_{\text{TS-}\ell_p}$, (c) the corresponding zoomed region of TrustSpa-ℓ_1 reconstruction, $\hat{f}_{\text{TS-}\ell_1}$. MSE = $(1/n)\|\hat{f} - f^*\|_2^2$. Note the presence of artifacts (represented in black spikes) in the TrustSpa-ℓ_1 reconstruction that are rarely present in the TrustSpa-ℓ_p reconstruction.

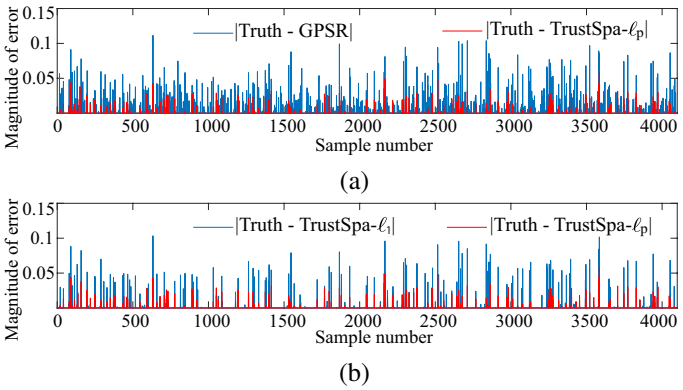

Fig. 3. Magnitude of error between the true signal f^* and reconstructions: (a) magnitude of error of the GPSR reconstruction $\hat{f}_{\text{GPSR}}$ vs. magnitude of error of $\hat{f}_{\text{TS-}\ell_p}$, (b) magnitude of error of $\hat{f}_{\text{TS-}\ell_1}$ vs. magnitude of error of $\hat{f}_{\text{TS-}\ell_p}$. Note the lower error for the proposed TrustSpa-ℓ_p when $p = 0.7$ (represented in red), whose values are more closer to zero.

the $\hat{f}_{\text{TS-}\ell_1}$ (see blue spikes in Fig. 3(b)). These results indicate that the TrustSpa-ℓ_p method better approximates the truth in terms of signal amplitude.

Furthermore, ten-trial average MSE values and average computational times for each method are reported in Table I, showing that the above single trial claims are robust with different Gaussian noise realizations.

Discussion. In our experience with this 5% Gaussian noise corrupted dataset, there is no any significant improvement in MSE value for p-values less than 0.7. Therefore, we have used the TrustSpa-ℓ_p reconstruction with $p = 0.7$ as a representative for the TrustSpa-ℓ_p method.

TABLE I
RECONSTRUCTION MSE AND COMPUTATIONAL TIME FOR THE RESULTS AVERAGED OVER TEN-TRIALS. MSE = $(1/n)\|\hat{f} - f^*\|_2^2$.

	Ten-trial average	
Method	MSE	Time (seconds)
GPSR	1.758×10^{-4}	4.45
TrustSpa-ℓ_1	9.827×10^{-5}	3.52
TrustSpa-ℓ_p	1.791×10^{-5}	1.57

V. CONCLUSIONS AND FUTURE RESEARCH

In this paper, we proposed a quasi-Newton trust-region method for solving a non-convex penalized sparsity recovery problem. We formulate the minimization problem as a smooth unconstrained optimization problem using a change of variables. We bridge the commonly-used convex ℓ_1 norm with the ℓ_0 quasi-norm using a non-convex ℓ_p norm. Numerical results show that the proposed TrustSpa-ℓ_p approach eliminates spurious solutions more effectively than using an ℓ_1-regularization term. Furthermore, the proposed method converges faster over the 10-trial average.

Future work includes applications of our method to sparse recovery in higher dimensional data. In particular, we plan extensions of the tests from 1D in the current paper (sparse signal recovery) to 2D images, 3D hyperspectral datacubes, and 4D spectro-polarimetric data. Some real data developed for our work in [4], [5], [7], [8], [9], [12], [32] will be used.

ACKNOWLEDGMENT

This work was supported by National Science Foundation Grant CMMI-1333326 and CMMI-1334042. Lasith Adhikari's research is supported by the UC Merced Graduate Student Opportunity Fellowship Program. Research by Plemmons was supported by the U.S. Air Force Office of Scientific Research (AFOSR) under Grant no. FA9550-15-1-0286.

REFERENCES

[1] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, p. 489, 2006.

[2] M. Gehm, R. John, D. Brady, R. Willett, and T. Schulz, "Single-shot compressive spectral imaging with a dual-disperser architecture," *Optics Express*, vol. 15, no. 21, pp. 14 013–14 027, 2007.

[3] J. Fowler, "Compressive-projection principal component analysis," *Image Processing, IEEE Transactions on*, vol. 18, no. 10, pp. 2230–2242, 2009.

[4] R. Marcia and R. Willett, "Compressive coded aperture superresolution image reconstruction," in *IEEE International Conference on Acoustics, Speech and Signal Processing*. IEEE, 2008, pp. 833–836.

[5] Q. Zhang, R. Plemmons, D. Kittle, D. Brady, and S. Prasad, "Joint segmentation and reconstruction of hyperspectral data with compressed measurements," *Applied Optics*, vol. 50, no. 22, pp. 4417–4435, 2011.

[6] L. Adhikari, J. Erway, S. Lockhart, and R. F. Marcia, "Trust-region methods for sparsity relaxation," 2016, Submitted to *2016 IEEE International Conference on Image Processing*.

[7] J. Zhang, J. Erway, X. Hu, Q. Zhang, and R. Plemmons, "Randomized SVD methods in hyperspectral imaging," *J. Electrical and Computer Engineering, in press*, 2012.

[8] Q. Zhang, V. Pauca, and R. Plemmons, "Randomized methods in lossless compression of hyperspectral data," *Journal of Applied Remote Sensing*, vol. 7, no. 1, p. 074598, 2013.

[9] Q. Zhang and R. Plemmons, "Image reconstruction from double random projection," *IEEE Transactions on Image Processing*, vol. 23, no. 6, pp. 2501–2513, 2014.

[10] M. Jing, X. Zhou, and C. Qi, "Quasi-newton iterative projection algorithm for sparse recovery," *Neurocomputing*, vol. 144, no. 5, pp. 169–173, 2014.

[11] A. Wagadarikar, R. John, R. Willett, and D. Brady, "Single disperser design for coded aperture snapshot spectral imaging," *Applied optics*, vol. 47, no. 10, pp. B44–B51, 2008.

[12] Q. Zhang, S. Prasad, and R. Plemmons, "Shape, pose, and material recovery of solar-illuminated surfaces from compressive spectral-polarimetric image data," in *Proc. AMOS Tech. Conf., Maui, HI*, 2013.

[13] A. Banerjee, S. Merugu, I. S. Dhillon, and J. Ghosh, "Clustering with Bregman divergences," *The Journal of Machine Learning Research*, vol. 6, pp. 1705–1749, 2005.

[14] A. Oh and R. Willett, "Regularized non-Gaussian image denoising," *ArXiv Preprint 1508.02971*, 2015.

[15] J. Yu, S. Vishwanathan, S. Günter, and N. N. Schraudolph, "A quasi-Newton approach to nonsmooth convex optimization problems in machine learning," *The Journal of Machine Learning Research*, vol. 11, pp. 1145–1200, 2010.

[16] J. Lee, Y. Sun, and M. Saunders, "Proximal Newton-type methods for convex optimization," in *Advances in Neural Information Processing Systems*, 2012, pp. 836–844.

[17] S. Becker and J. Fadili, "A quasi-newton proximal splitting method," in *Advances in Neural Information Processing Systems*, 2012, pp. 2618–2626.

[18] G. Zhou, X. Zhao, and W. Dai, "Low rank matrix completion: A smoothed l0-search," in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2012, pp. 1010–1017.

[19] Y. Wang, J. Cao, and C. Yang, "Recovery of seismic wavefields based on compressive sensing by an l1-norm constrained trust region method and the piecewise random subsampling," *Geophysical Journal International*, vol. 187, no. 1, pp. 199–213, 2011.

[20] M. Hintermüller and T. Wu, "Nonconvex TV^q-models in image restoration: Analysis and a trust-region regularization–based superlinearly convergent solver," *SIAM Journal on Imaging Sciences*, vol. 6, no. 3, pp. 1385–1415, 2013.

[21] R. Chartrand, "Nonconvex compressed sensing and error correction," in *Proceedings of 2007 IEEE ICASSP*, Honolulu, Hawaii, April 2007.

[22] ——, "Exact reconstruction of sparse signals via nonconvex minimization," *Signal Processing Letters, IEEE*, vol. 14, no. 10, pp. 707–710, Oct 2007.

[23] R. Chartrand and V. Staneva, "Restricted isometry properties and nonconvex compressive sensing," *Inverse Problems*, vol. 24, no. 3, p. 035020, 2008. [Online]. Available: http://stacks.iop.org/0266-5611/24/i=3/a=035020

[24] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, 2008, pp. 3869–3872.

[25] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, "A generalized iterated shrinkage algorithm for non-convex sparse coding," in *2013 IEEE International Conference on Computer Vision (ICCV)*, Dec 2013, pp. 217–224.

[26] R. H. Byrd, J. Nocedal, and R. B. Schnabel, "Representations of quasi-Newton matrices and their use in limited-memory methods," *Math. Program.*, vol. 63, pp. 129–156, 1994.

[27] J. Nocedal and S. Wright, *Numerical optimization*. Springer Science & Business Media, 2006.

[28] A. R. Conn, N. I. M. Gould, and P. L. Toint, *Trust-Region Methods*. Philadelphia, PA: Society for Industrial and Applied Mathematics (SIAM), 2000.

[29] D. M. Gay, "Computing optimal locally constrained steps," *SIAM J. Sci. Statist. Comput.*, vol. 2, no. 2, pp. 186–197, 1981.

[30] J. J. Moré and D. C. Sorensen, "Computing a trust region step," *SIAM J. Sci. and Statist. Comput.*, vol. 4, pp. 553–572, 1983.

[31] M. Figueiredo, R. Nowak, and S. Wright, "Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 4, pp. 586–597, 2007.

[32] R. Plemmons, S. Prasad, and P. Pauca, "Deblurring compressive spectro-polarimetric images trough atmospheric turbulence," in *Proc. Conf. on Signal Recovery & Synthesis (SRS), OSA Technical Digest*, 2014.