SPARSE RECOVERY OF MULTIPLE MEASUREMENT VECTORS IN IMPULSIVE NOISE: A SMOOTH BLOCK SUCCESSIVE MINIMIZATION ALGORITHM

Zhen-Qing He, Zhi-Ping Shi*, Lei Huang†, Hongbin Li‡, and H. C. So♯*

*National Key Lab of Communications, University of Electronic Science & Technology of China, China
†College of Information Engineering, Shenzhen University, China
‡Department of Electrical & Computer Engineering, Stevens Institute of Technology, USA
♯Department of Electronic Engineering, City University of Hong Kong, Hong Kong, China

ABSTRACT

This paper considers the sparse recovery problem of multiple measurement vector (MMV) model corrupted in impulsive noise. To ensure outlier-robust sparse recovery, we formulate an MMV problem that includes the generalized ℓ_p-norm $(1 < p < 2)$ divergence data-fidelity term added to the $\ell_{2,0}$ joint sparsity-promoting regularizer. The $\ell_{2,0}$ joint sparse penalty, however, is non-continuous and hence non-differentiable, which inevitably raises difficulty in optimization when using a gradient-based method. To address this, we build a smooth approximation for the $\ell_{2,0}$-based sparse metric via the log-sum based sparse-encouraging surrogate function. Then, we propose a block successive upper-bound minimization algorithm for the smooth MMV problem by solving a series of subproblems based on the block coordinate descent (BCD) method. Furthermore, local convergence of the proposed algorithm to a stationary point of the smooth problem is proved. Experiments demonstrate its efficiency and robust recovery performance for suppressing impulsive noise.

***Index Terms*—** Compressed sensing, impulsive noise, multiple measurement vectors, sparse signal recovery.

1. INTRODUCTION

The sparse signal recovery (SSR) problem, an emerging topic bearing a close affinity with compressive sensing (CS) [1, 2], has attracted considerable interest over the past few years. In particular, as CS theories and applications mature, some special structures beyond the classical single measurement vector model have been further exploited within SSR to further, e.g., reduce the number of linear measurements and facilitate more accurate sparse recovery. One such structure is the *multiple measurement vector* (MMV) model [3–5] which aims to jointly recover a set of sparse vectors that share a common support. Mathematically, given a dictionary $\mathbf{\Phi} \in \mathbb{R}^{M\times N}$ $(M < N)$, an MMV recovery problem requires solving for $\mathbf{X} \in \mathbb{R}^{N\times Q}$ to find a joint sparse (approximate) solution from the MMV model $\mathbf{Y} = \mathbf{\Phi X} + \mathbf{E}$ where $\mathbf{E} \in \mathbb{R}^{M\times Q}$ is the additive noise matrix. By appropriately including a joint $\ell_{2,0}$ sparse regularizer, a well-known approach for recovering $\mathbf{X}$ from the observed $\mathbf{Y}$ is the following denoising problem

$$\min_{\mathbf{X}} \ \frac{1}{2}\|\mathbf{Y} - \mathbf{\Phi X}\|_F^2 + \lambda\|\mathbf{X}\|_{2,0} \tag{1}$$

where $\|\mathbf{X}\|_{2,0} \triangleq \sum_{i=1}^{N} \|\mathbf{x}_i\|_2^0$ ($\mathbf{x}_i$ is the ith row of $\mathbf{X}$), $\lambda > 0$ is a regularization parameter, $\|\cdot\|_F$ and $\|\cdot\|_2$ denote the Frobenius norm and ℓ_2-norm, respectively. Indeed, the MMV problem with form (1) has been motivated by a wide range of applications such as magnetoencephalography [6] and direction-of-arrival estimation [7, 8]. In these applications, the authors provide a variety of excellent algorithms by using some approximate sparse-promoting penalties or $\ell_{2,1}$ convex relaxed surrogates for the ideally $\ell_{2,0}$ sparse penalty.

However, most of the aforementioned methods rest upon the squared (ℓ_2) loss metric which is, in general, the optimal metric to quantify the Gaussian background noise. Therefore, the conventional ℓ_2-based MMV recovery algorithms often fall short when the observed noises $\mathbf{E}$ are non-Gaussian distributed and contain outliers. As a kind of very important non-Gaussian noise, the impulsive noise, whose probability density function has heavier tails than the Gaussian distribution, has in fact been reported in many practical applications such as in array processing [9], spectral analysis [10], wireless communications [11], and image processing [12].

Motivated by a variety applications mentioned above, we consider the generalized ℓ_p $(1 < p < 2)$ norm as a divergence metric in the MMV sparse recovery problem, which is particularly suitable for suppressing the impulsive noise. As a result, the outlier-resistant recovery problem is formulated as the generalized $\ell_p - \ell_{2,0}$ problem. To deal with the intractability (non-smoothness) of $\ell_{2,0}$-norm penalty, we build a smooth approximation via the log-sum based sparse-encouraging surrogate function. Then, we propose an effcient block successive minimization algorithm for the smooth MMV problem by solving a series of upper-bound subproblems based on the block coordinate descent (BCD) method. Additionally, we prove that the proposed algorithm can converge a stationary point of the smooth MMV problem. Our experiments show

This work was supported in part by the National 863 project (2014AA01A704), the National Natural Science Foundation of China (61571003, U1501253 and 61271420) and the Guangdong Natural Science Foundation (2015A030311030), and by the Chinese Scholarship Council.

that the proposed algorithm can achieve an accurate joint sparse recovery and suppress the impulsive noise.

2. PROBLEM FORMULATION

To guarantee outlier-resistant sparse reconstruction, we consider the generalized MMV recovery problem

$$\min_{\mathbf{X}}\ G_p(\mathbf{X}) + \lambda \cdot H(\mathbf{X}) \tag{2}$$

where $G_p(\mathbf{X}) \triangleq \frac{1}{p}\|\mathrm{vec}(\mathbf{Y}) - (\mathbf{I}_Q \otimes \mathbf{\Phi})\mathrm{vec}(\mathbf{X})\|_p^p$ and $H(\mathbf{X}) \triangleq \|\mathbf{X}\|_{2,0}^0$. Here, the ℓ_p-norm divergence function $G_p(\mathbf{X})$ can be used to suppress the impulsive noise embedded in $\mathbf{E}$ when $1 < p < 2$.

3. SMOOTH SPARSE APPROXIMATION

Note that he $\ell_{2,0}$ joint sparse penalty $H(\mathbf{X})$ can be expressed as $H(\mathbf{X}) \triangleq \|\mathbf{X}\|_{2,0}^0 = \sum_{i=1}^N \mathrm{sgn}(\|\mathbf{x}_i\|_2)$ where $\mathrm{sgn}(\cdot)$ denotes the sign function. The sign function, i.e., $\mathrm{sgn}(\|\mathbf{x}_i\|_2)$, however, is discontinuous and hence non-differentiable at $\|\mathbf{x}_i\|_2 = 0$ or $\mathbf{x}_i = \mathbf{0}$ where $\mathbf{0}$ is a zero vector with proper dimension. To deal with this, a natural approach is to approximate the problematic function $\mathrm{sgn}(\|\mathbf{x}_i\|_2)$ with some well-chosen continuous surrogate functions. In this paper, we will consider the following log-based surrogate function

$$h_q(\|\mathbf{x}_i\|_2) = \frac{\log(1 + \|\mathbf{x}_i\|_2/q)}{\log(1 + 1/q)} \tag{3}$$

to approximate $\mathrm{sgn}(\|\mathbf{x}_i\|_2)$. In fact, it has been shown that the log-based function $h_q(\|\mathbf{x}_i\|_2)$ exhibits uniform superiority in sparse optimization and usually leads to a kind of iteratively reweighted algorithms [13–15]. Although $h_q(\|\mathbf{x}_i\|_2)$ is now a continuous function, it is still non-differentiable, i.e., nonsmooth, at $\|\mathbf{x}_i\|_2 = 0$, which, usually, leads to an intractable optimization when using a gradient-based method. Therefore following [13] on constructing smooth approximation for single measurement vector sparse optimization, we approximate $h_q(\|\mathbf{x}_i\|_2)$ around $\|\mathbf{x}_i\|_2 = 0$ via a quadratic function, with the following form

$$h_q^\epsilon(\|\mathbf{x}_i\|_2) = \begin{cases} a\|\mathbf{x}_i\|_2^2, & \|\mathbf{x}_i\|_2 \le \epsilon \\ h_q(\|\mathbf{x}_i\|_2) - b, & \|\mathbf{x}_i\|_2 > \epsilon \end{cases} \tag{4}$$

where $\epsilon > 0$ is small parameter. The continuousness and differentiability of $h_q^\epsilon(\|\mathbf{x}_i\|_2)$ at $\|\mathbf{x}_i\|_2 = \epsilon$ indicate that $a\epsilon^2 = h_q(\epsilon) - b$ and $2a\epsilon = \left.\frac{\mathrm{d}h_q(\|\mathbf{x}_i\|_2)}{\mathrm{d}\|\mathbf{x}_i\|_2}\right|_{\|\mathbf{x}_i\|_2=\epsilon}$, which eventually result in the following smooth approximation

$$h_q^\epsilon(\|\mathbf{x}_i\|_2) = \begin{cases} \frac{\|\mathbf{x}_i\|_2^2}{2\epsilon(q+\epsilon)\log(1+1/q)}, & \|\mathbf{x}^i\|_2 \le \epsilon \\ \frac{\log\left(1+\frac{\|\mathbf{x}^i\|_2}{q}\right) - \log\left(1+\frac{\epsilon}{q}\right) + \frac{\epsilon}{2(q+\epsilon)}}{\log(1+q)}, & \|\mathbf{x}^i\|_2 > \epsilon. \end{cases} \tag{5}$$

It can be seen from (5) that when ϵ becomes further smaller (close to zero), $h_q^\epsilon(\|\mathbf{x}_i\|)$ will promote better approximation for $h_q(\|\mathbf{x}_i\|)$. Meanwhile, when $q \to 0$, $h_q^\epsilon(\|\mathbf{x}_i\|)$ can get closer to $\mathrm{sgn}(\|\mathbf{x}^i\|_2)$ and hence $H_q^\epsilon(\mathbf{X}) \triangleq \sum_{i=1}^N h_q^\epsilon(\|\mathbf{x}_i\|_2)$ encourages much more stronger joint (row) sparsity. Armed with the smooth approximation $H_q^\epsilon(\mathbf{X})$, problem (2) can be approximated by

$$\min_{\mathbf{X}}\ F(\mathbf{X}) \triangleq G_p(\mathbf{X}) + \lambda \cdot H_q^\epsilon(\mathbf{X}). \tag{6}$$

Although problem (6) is now a smooth problem and can be solved by exploiting a gradient-based descent method, it is still computationally intensive because it involves optimizing with respect to the matrix variable $\mathbf{X}$. This may be very impractical in, e.g., problems involving the big dictionary matrix where usually $N \gg Q$. In the following, we will develop an efficient block successive minimization algorithm based on the parallel BCD method.

4. BLOCK SUCCESSIVE MINIMIZATION ALGORITHM

4.1. Block Coordinate Descent

Instead of directly solving (6) over $\mathbf{X}$, the BCD method [16, 17], also known as the nonlinear Gauss-Seidel method, decomposes the whole $\mathbf{X}$ as a successive block variables, $\mathbf{x}_i, i = 1, \ldots, N$, and in each iteration, only updates a single block variable while the other blocks are fixed. This decomposition usually leads to a parallel optimization and hence an efficient implementation. More specifically, given the current iteration $\mathbf{X}^{(k)} = \left[\mathbf{x}_1^{(k)}, \ldots, \mathbf{x}_N^{(k)}\right]^T$, the next iteration $\mathbf{X}^{(k+1)} = \left[\mathbf{x}_1^{(k+1)}, \ldots, \mathbf{x}_N^{(k+1)}\right]^T$ is updated by successively solving the subproblems

$$\min_{\mathbf{x}_i}\ F\left(\mathbf{x}_1^{(k+1)}, \ldots, \mathbf{x}_{i-1}^{(k+1)}, \mathbf{x}_i, \ldots, \mathbf{x}_{i+1}^{(k)}, \ldots, \mathbf{x}_N^{(k)}\right) \tag{7}$$

where $F\big(\mathbf{x}_1^{(k+1)}, \ldots, \mathbf{x}_{i-1}^{(k+1)}, \mathbf{x}_i, \ldots, \mathbf{x}_{i+1}^{(k)}, \ldots, \mathbf{x}_N^{(k)}\big) = G_p\big(\mathbf{x}_1^{(k+1)}, \ldots, \mathbf{x}_{i-1}^{(k+1)}, \mathbf{x}_i, \ldots, \mathbf{x}_{i+1}^{(k)}, \ldots, \mathbf{x}_N^{(k)}\big) + \lambda \cdot H_q^\epsilon \big(\mathbf{x}_1^{(k+1)}, \ldots, \mathbf{x}_{i-1}^{(k+1)}, \mathbf{x}_i, \ldots, \mathbf{x}_{i+1}^{(k)}, \ldots, \mathbf{x}_N^{(k)}\big)$. By ignoring the constant terms in the objective function of (7), this problem can be further simplified as

$$\min_{\mathbf{x}_i}\ g_p(\mathbf{x}_i) + \lambda \cdot h_q^\epsilon(\|\mathbf{x}_i\|_2) \tag{8}$$

where $g_p(\mathbf{x}_i) = \frac{1}{p}\|\bar{\mathbf{y}} - \mathbf{\Phi}_i\mathbf{x}_i\|_p^p$, $\mathbf{\Phi}_i = \mathbf{I}_Q \otimes \boldsymbol{\phi}^i \in \mathbb{R}^{MQ\times Q}$, $\mathbf{I}_Q$ is the $Q \times Q$ identity matrix, $\otimes$ symbolizes the kronecker product, $\boldsymbol{\phi}^i$ is the ith column of $\mathbf{\Phi}$, $\bar{\mathbf{y}} = \mathrm{vec}\big(\mathbf{Y} - \sum_{n=1}^{i-1} \boldsymbol{\phi}^n \big(\mathbf{x}_n^{(k+1)}\big)^T - \sum_{n=i+1}^{N} \boldsymbol{\phi}^n \big(\mathbf{x}_n^{(k)}\big)^T\big)$. Despite now a lower complexity of problem (8) when compared to (6), it is still non-convex and not easy to handle directly due to the non-convexity of $h_q^\epsilon(\|\mathbf{x}_i\|_2)$. In contrast to directly solving the non-convex problem (8), the following subsection will concentrate on how to approximately solve it with the successive quadratic upper-bound minimization.

4.2. Successive Quadratic Upper-Bound Minimization

The successive upper-bound minimization method [17], very similar to the majorization-minimization method [18, 19], aims to minimize an upper-bound approximation of the objective function. Usually, the upper-bound function requires much easier to be implemented than the objective function. An appropriate choice is the separable quadratic function since it has a unique closed-form solution which benefits the convergence analysis.

In the following, we will construct the quadratic upper-bound functions $u_p(\mathbf{x}_i, \mathbf{x}_i^{(k)})$ and $u_q^{\epsilon}(\mathbf{x}_i, \mathbf{x}_i^{(k)})$, respectively, for $g_p(\mathbf{x}_i)$ and $h_q^{\epsilon}(\mathbf{x}_i)$. More specifically, at iteration k we need to construct a quadratic approximate function for $g_p(\mathbf{x}_i)$ with the following separable quadratic weighted form

$$u_p\Big(\mathbf{x}_i, \mathbf{x}_i^{(k)}\Big) = [\bar{\mathbf{y}} - \mathbf{\Phi}_i\mathbf{x}_i]^T\, \hat{\mathbf{W}}_i^{(k)}[\bar{\mathbf{y}} - \mathbf{\Phi}_i\mathbf{x}_i] + \alpha_i^{(k)}, \quad (9)$$

where $\hat{\mathbf{W}}_i^{(k)} = \operatorname{diag}\big(\hat{\mathbf{w}}_i^{(k)}\big)$ is a diagonal weighted matrix formed by the vector $\hat{\mathbf{w}}_i^{(k)} \in \mathbb{R}^{MQ}$ and $\alpha_i^{(k)}$ is a constant to be determined. The tightness of the upper-bound function $u_p(\mathbf{x}_i^{(k)})$ at $\mathbf{x}_i^{(k)}$ results in the following two conditions:

$$g_p(\mathbf{x}_i^{(k)}) = u_p(\mathbf{x}_i^{(k)}), \quad (10)$$

$$\nabla g_p(\mathbf{x}_i^{(k)}) = \nabla u_p(\mathbf{x}_i^{(k)}). \quad (11)$$

By defining $\mathbf{z}_i^{(k)} \triangleq \big[\bar{\mathbf{y}} - \mathbf{\Phi}_i\mathbf{x}_i^{(k)}\big]$ and then solving (11), we can calculate each element of the weighted vector $\hat{\mathbf{w}}_i^{(k)}$ as

$$\hat{w}_{i,j}^{(k)} = \frac{1}{2}\big|z_{i_j}^{(k)}\big|^{p-2}, \quad \big|z_{i,j}^{(k)}\big| \neq 0. \quad (12)$$

where $\hat{w}_{i,j}^{(k)}$ and $z_{i,j}^{(k)}$ are the jth elements of $\hat{\mathbf{w}}_i^{(k)}$ and $\mathbf{z}_i^{(k)}$, respectively. Here, we will not explicitly calculate α_i^k with its exact expression since the constant term will not pose any impact for the optimal solution; but with $\hat{\mathbf{w}}_i^{(k)}$ having been determined in (12), it is easy to derive $\alpha_i^{(k)}$ based on the condition (10). Similarly, following the derivation of the upper-bound function $u_p\big(\mathbf{x}_i, \mathbf{x}_i^{(k)}\big)$ of $g_p(\mathbf{x}_i)$ from (9) to (12), the upper-bound approximation of $h_q^{\epsilon}(\mathbf{x}_i)$ is given by

$$u_q^{\epsilon}\left(\mathbf{x}_i, \mathbf{x}_i^{(k)}\right) = \mathbf{x}_i^T \check{\mathbf{W}}_i^{(k)}\mathbf{x}_i + \beta_i^{(k)} \quad (13)$$

where $\beta_i^{(k)}$ is also a constant term, $\check{\mathbf{W}}_i^{(k)} = \operatorname{diag}\big(\check{\mathbf{w}}_i^{(k)}\big) = \check{w}_i^{(k)} \cdot \mathbf{I}_Q$ and $\check{w}_i^{(k)}$ is given as

$$\check{w}_i^{(k)} = \begin{cases} \frac{1}{2\epsilon(q+\epsilon)\log(1+1/q)}, & \big\|\mathbf{x}_i^{(k)}\big\|_2 \leq \epsilon, \\ \frac{1}{2\log(1+1/q)\|\mathbf{x}_i^{(k)}\|_2(\|\mathbf{x}_i^{(k)}\|_2+q)}, & \big\|\mathbf{x}_i^{(k)}\big\|_2 > \epsilon. \end{cases} \quad (14)$$

Now, $u_p\big(\mathbf{x}_i, \mathbf{x}_i^{(k)}\big) + \lambda\, u_q^{\epsilon}\big(\mathbf{x}_i, \mathbf{x}_i^{(k)}\big)$ is a global upper-bound of the objective function of (8) and touches with it at $\mathbf{x}_i = \mathbf{x}_i^{(k)}$ (derivation omitted here because of limited space). By ignoring the constant terms, the upper-bound approximate problem eventually becomes

$$\min_{\mathbf{x}_i} \; [\bar{\mathbf{y}} - \mathbf{\Phi}_i\mathbf{x}_i]^T\, \hat{\mathbf{W}}_i^{(k)}\, [\bar{\mathbf{y}} - \mathbf{\Phi}_i\mathbf{x}_i] + \lambda\, \mathbf{x}_i^T \check{\mathbf{W}}_i^{(k)}\mathbf{x}_i \quad (15)$$

and whose optimal solution, as a preparation for the next iteration, is uniquely given by

$$\mathbf{x}_i^{(k+1)} = \Big[\mathbf{\Phi}_i\hat{\mathbf{W}}_i^{(k)}\mathbf{\Phi}_i^T + \lambda\, \check{\mathbf{W}}_i^{(k)}\Big]^{-1}\mathbf{\Phi}_i^T\hat{\mathbf{W}}_i^{(k)}\bar{\mathbf{y}}. \quad (16)$$

For clarification, the iterative procedure derived above for MMV sparse recovery is referred to the successive quadratic upper-bound minimization (SQUM) algorithm, which is summarized in Algorithm I, where the smooth piece wise function (14) is replaced with the "**if else**" statement.

Algorithm 1: SQUM

Input: $k \leftarrow 0$, $\mathbf{Y}$, $\mathbf{\Phi}$, λ, ϵ, p, q and $\mathbf{X}^{(0)}$;
Output: $\mathbf{X}^{(k)}$;
repeat
- $k \leftarrow k+1$;
- $i \leftarrow (k \mod N) + 1$;
- $\hat{\mathbf{X}} \leftarrow \mathbf{X}^{(k-1)}$;
- $\mathbf{z} \leftarrow \operatorname{vec}(\mathbf{Y} - \mathbf{D}\hat{\mathbf{X}})$;
- $\hat{\mathbf{W}} \leftarrow \operatorname{diag}\left\{\frac{|\mathbf{z}|^{p-2}}{2}\right\}$;
- **if** $\|\hat{\mathbf{x}}_i\|_2 \leq \epsilon$ **then**
  - $\check{w} \leftarrow \frac{1}{2\epsilon(q+\epsilon)\log(1+1/q)}$;
- **else**
  - $\check{w} \leftarrow \frac{1}{2\log(1+1/q)\|\hat{\mathbf{x}}_i\|_2(\|\hat{\mathbf{x}}_i\|_2+q)}$;
- **end**
- $\mathbf{D} \leftarrow \mathbf{I}_Q \otimes \boldsymbol{\phi}^i$;
- $\bar{\mathbf{Y}} \leftarrow \mathbf{Y} - \sum_{n=1}^{i-1}\boldsymbol{\phi}^n\big(\mathbf{x}_n\big)^T - \sum_{n=i+1}^{N}\boldsymbol{\phi}^n\big(\mathbf{x}_n\big)^T$;
- $\hat{\mathbf{x}}_i \leftarrow \big[\mathbf{D}\hat{\mathbf{W}}\mathbf{D}^T + \lambda\check{w}\,\mathbf{I}_Q\big]^{-1}\mathbf{D}^T\hat{\mathbf{W}}_i^{(k)}\operatorname{vec}(\bar{\mathbf{Y}})$;
- $\mathbf{X}^{(k)} \leftarrow \hat{\mathbf{X}}$;

until *satisfy certain stopping criterion or reach certain number of iterations*;

4.3. Local Convergence

The SQUM algorithm is very efficient and simple with the quadratic update (16), especially useful for the big dictionary MMV recovery problems. Despite its simplicity, now a natural question arises: through the smooth block successive minimization, does it guarantee a convergence to a stationary point of problem (6)? Actually, for any given initial point, the SQUM algorithm has a local convergence, which is proved by the following proposition.

Proposition 1. *Given any initial point $\mathbf{X}^{(0)} \in \mathbb{R}^{N\times Q}$, the sequence $\big\{\mathbf{X}^{(k)}\big\}$ generated by the SQUM algorithm converges to a stationary point of the problem (6).*

Proof. Note that $F(\mathbf{X})$ is continuous and coercive[1] over its domain $\mathbf{X} \in \mathbb{R}^{N\times Q}$. According to the Weierstrass's theorem,

[1]We say $F(\mathbf{X})$ is coercive [16] if $\lim_{x_{i,j}\to\infty} F(\mathbf{X}) = \infty, \forall i = 1,\ldots,N, j = 1,\ldots,Q$, where $x_{i,j}$ is the (i,j)th element of $\mathbf{X}$.

for every $\eta \in \mathbb{R}$, the sublevel set $\left\{\mathbf{X} \,|\, F(\mathbf{X}) \leq \eta\right\}$ is a compact (bounded and closed) set. In addition, the upper-bound property implies that $F(\mathbf{X}^{(k+1)}) \leq F(\mathbf{X}^{(k)}), \forall\, k \geq 0$, i.e., the sequence $\{F(\mathbf{X}^{(k)})\}$ is a non-increasing sequence. Naturally, if we assume $\eta_0 = F(\mathbf{X}^{(0)})$, then $\mathbf{X}^{(k)} \in \left\{\mathbf{X} \,|\, F(\mathbf{X}) \leq \eta_0\right\}$ and hence $\left\{\mathbf{X}^{(k)}\right\}$ is a bounded sequence. Therefore, there must exist a convergent subsequence, i.e, a limit point, of $\left\{\mathbf{X}^{(k)}\right\}$ since each bounded sequence has a convergent subsequence. Then according to Corollary 1 of [17], $\left\{\mathbf{X}^{(k)}\right\}$ converges to a stationary point of the problem (6). □

5. SIMULATION RESULTS

We now numerically assess the performance of the proposed SQUM algorithm in impulsive noise. Throughout this section, the measurement matrix $\mathbf{\Phi}$ is assumed to be a Gaussian random matrix whose elements are drawn independently from the standard normal distribution with mean 0 and variance 1, and in which every column vector is normalized to 1. The dimensions of $\mathbf{\Phi}$ are fixed to $M = 20$ and $N = 200$. The non-zero row (block) of the ideally sparse vectors $\mathbf{X}^{\text{true}}$ is also draw independently from the zero-mean normal distribution with variance σ_s^2 and the number of non-zero row (group) is set to be 2. Specifically, to model the impulsive noise, each element of $\mathbf{E}$ is generated according to the generalized Gaussian distribution [9] with probability density function

$$p_e(e) = \frac{\beta\Gamma(4/\beta)}{2\pi\sigma_e^2\Gamma^2(2/\beta)} \exp\left(-\frac{e^2}{c\sigma_e^\beta}\right) \tag{17}$$

where $c = \left(\Gamma(2/\beta)/\Gamma(4/\beta)\right)^{\frac{\beta}{2}}$ and $\beta > 0$ is the shape parameter. When β becomes smaller, the noise will be more impulsive. In our simulations, we fix the parameter $\beta = 0.1$. The signal-to-noise ratio (SNR) is defined as $\text{SNR} = 10\log\left\{\sigma_s^2/\sigma_e^2\right\}$. The dimension of $\mathbf{X}^{\text{true}}$ and the smoothing parameter ϵ are, respectively, set to be 200×10 and 10^{-5}. To promote sparsity, we keep the parameter $q = 0.5$. The initial point of $\mathbf{X}^{(0)}$ is set to be a zero matrix. The regularization parameter λ is set as 2.5 (a large number of experiments show that $\lambda = 2.5$ lead to a better performance than other values).

We first investigate the convergence behavior of the proposed SQUM algorithm under the several values of p: $p = 1.0, 1.1, 1.2, 1.3, 1.4, 1.8, 2.0$. Fig. 1 depicts the the normalized mean square error (NMSE), defined as $\frac{\|\mathbf{X}^{\text{true}}-\hat{\mathbf{X}}\|_2^2}{\|\mathbf{X}^{\text{true}}\|_2^2}$, versus the number of iterations under the case of $\text{SNR} = 5\,\text{dB}$. As can be seen from this figure, different values of p pose a diverse recovery performance. As with increase of p from 1.1 to 2, the NMSE also mostly grows with different values, while the best performance is attained with $p = 1.1$.

To further evaluate the performance of the proposed algorithm for MMV sparse recovery, we conduct 200 independent trials under various SNRs with range from $0\,\text{dB}$ to $10\,\text{dB}$. Meanwhile, based on the above experiment result, we choose the value of p as 1.1. The standard BCD for separable group lasso convex $\ell_{2,1}$ [20] and the simultaneous orthogonal matching pursuit (SOMP) [5] are considered for comparison. The corresponding result on the average NMSE versus the SNR is shown in Fig. 2. It is seen that the proposed SQUM algorithm exhibits the uniform superiority over SOMP and BCD where they are implemented based upon the Euclidean ℓ_2 orthogonality and the ℓ_2 norm fitting criteria, which indicates that ℓ_p-based MMV sparse recovery is robust to impulsive noise.

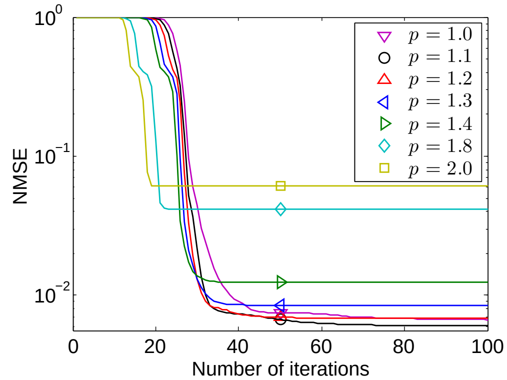

Fig. 1. Objective function $F(\mathbf{X})$ versus number of iterations under various values of p.

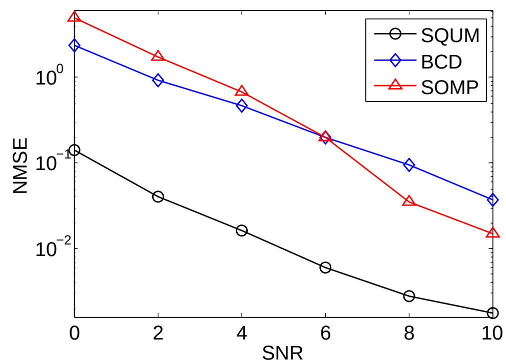

Fig. 2. NMSE versus number of iterations under various values of p.

6. CONCLUSION

In this paper, we have proposed a fast algorithm for M-MV sparse recovery problem in impulsive noise with heavy tailed distribution. Different from the conventional ℓ_2-norm based residual metric function, the generalized ℓ_p-norm $(0 < p < 2)$ divergence function is used to suppress the impulsive noise. To address the non-smoothness of the $\ell_{2,0}$ joint sparse penalty, we have constructed a smooth approximation to facilitate implementation. Then, the smooth approximate problem is solved based on the block coordinate descent method with successive upper-bound minimizations. Simulation results demonstrated that the SQUM algorithm can achieve outlier-robust sparse recovery.

7. REFERENCES

[1] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[2] E. Candès and B. Michael, "An introduction to compressive sampling," *IEEE Signal Proces. Mag.,* vol. 25, no. 2, pp. 21–30, Mar. 2008.

[3] S. F. Cotter, B. D. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, Jul. 2005.

[4] J. Chen and X. Huo, "Theoretical results on sparse representations of multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4634–4643, Dec. 2012.

[5] M. E. Davies, Y. C. Eldar, "Rank awareness in joint sparse recovery," *IEEE Trans. Inf. Theory*, vol. 58, no. 2, pp. 1135–1146, Feb. 2012.

[6] I. F. Gorodnitsky, J. S. George, and B. D. Rao, "Neuromagnetic source imaging with FOCUSS: A recursive weighted minimum norm algorithm," *Electroencephalogr. Clin. Neurophysiol.*, vol. 95, no. 4, pp. 231–251, Oct. 1995.

[7] D. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[8] M. M. Hyder, K. Mahata, "Direction-of-arrival estimation using a mixed norm approximation," *IEEE Trans. Signal Process.*, vol. 58, no. 9, pp. 4646–4655, Sep. 2010.

[9] W.-J. Zeng, H. C. So, and L. Huang, "ℓ_p-MUSIC: Robust direction-of-arrival estimator for impulsive noise environments," *IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4296–4308, Sep. 2013.

[10] Y. Chen, H. C. So, and W. Sun, "ℓ_p-norm based iterative adaptive approach for robust spectral analysis," *Signal Process.*, vol. 94, pp. 144–148, 2014.

[11] K. L. Blackard, T. S. Rappaport, and C. W. Bostian, "Measurements and models of radio frequency impulsive noise for indoor wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 11, no. 7, pp. 991–1001, Sep. 1993.

[12] R. E. Carrillo, K. E. Barner, and T. C. Aysal, "Robust sampling and reconstruction methods for sparse signals in the presence of impulsive noise," *IEEE J. Sel. Top. Signal Proces.*, vol. 4, no. 2, pp. 392–408, Apr. 2010.

[13] J. Song, P. Babu, and D. P. Palomar, "Sparse generalized eigenvalue problem via smooth optimization," *IEEE Trans. Signal Process.*, vol. 63, no. 7, pp. 1627–1642, Apr. 2015.

[14] E. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, pp. 877–905, Dec. 2008.

[15] J. Fang, J. Li, Y. Shen, H. Li, and S. Li, "Super-resolution compressed sensing: An iterative reweighted algorithm for joint parameter learning and sparse signal recovery," *IEEE Signal Process. Lett.*, vol. 21, no. 6, pp. 761–765, Jun. 2014.

[16] D. P. Bertsekas, *Nonlinear Programming.* Athena Scientific, 1999.

[17] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM Journal on Optimization*, vol. 23, no. 2, pp. 1126–1153, 2013.

[18] D. R. Hunter and K. Lange, "A tutorial on MM algorithms," *Amer. Statist.*, vol. 58, no. 1, pp. 30–37, Feb. 2004.

[19] M. A. Figueiredo, J. M. Bioucas-Dias, and R. D. Nowak, "Majorization-minimization algorithms for wavelet-based image restoration," *IEEE Trans. Image Process.*, vol. 16, no. 12, pp. 2980–2991, Dec. 2007.

[20] Z. Qin, K. Scheinberg, D. Goldfarb, "Efficient block-coordinate descent algorithms for the group lasso," *Mathematical Programming Computation*, vol. 52, no. 2, pp. 143–169, 2013.