SMOOTHED OPTIMIZATION FOR SPARSE OFF-GRID DIRECTIONS-OF-ARRIVAL ESTIMATION

Cheng-Yu Hung and Mostafa Kaveh

Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, USA

ABSTRACT

This paper is concerned with the development of a computationally efficient optimization algorithm for off-grid direction finding using a sparse observation model. The optimization problem can be formulated as one smooth plus two nonsmooth functions. We propose two accelerated smoothing proximal gradient algorithms. The Nesterov smoothing methodology is utilized to reformulate nonsmooth functions into smooth ones, and the accelerated proximal gradient algorithm is adopted to solve the smoothed optimization problem. The computational efficiency and efficacy of the proposed algorithms are demonstrated numerically.

Index Terms— Smoothing, Nonsmooth function, Accelerated proximal gradient, Nondifferentiable, Group sparsity

1. INTRODUCTION

Application of compressed sensing [1, 2] to directions-of-arrival (DoA) estimation has been an active area of investigation. In its original formulation, the compressed sensing approach for DoA estimation assumed a known dictionary formed from the array responses at a grid of candidate directions [3]. In practice, however, the DoAs are most likely not to be located on the model grid, leading to the now well-known off-grid DoA estimation problem, for which a number of model approximations and solutions have been proposed, for example [4–10]. A commonly-used observation for off-grid DoAs follows the noisy structured perturbation model given by:

$$\mathbf{y} = (\mathbf{A} + \mathbf{B}\Gamma)\mathbf{s} + \mathbf{n}, \tag{1}$$

where $\mathbf{A} \in \mathbb{C}^{M\times N}$ is known, and $\mathbf{B} \in \mathbb{C}^{M\times N}$ is known as part of the off-grid approximation. $\Gamma = diag(\boldsymbol{\beta})$, and $\boldsymbol{\beta} = [\beta_1, \dots, \beta_N]^T$ is denoted as the unknown coefficient vector for the approximation. $\mathbf{s}$ is the sparse vector associated with grid points nearest the true DoAs. Equation (1) can be solved by formulating a sparsity promoting constrained nonconvex minimization problem to estimate $\mathbf{s}$ and $\boldsymbol{\beta}$ sequentially by the alternating method [4, 5], but with slow convergence.

Instead of solving a constrained nonconvex minimization of (1), an unconstrained convex optimization problem, which is composed of one smooth and two nonsmooth functions, can be formulated. In [11], a number of primal-dual iterative approaches for solving large-scale nonsmooth optimization problems, such as the M+LFBF (Monotone+Lipschitz Forward Backward Forward) algorithm, are reviewed. In [12,13], subgradient methods are proposed, but their complexity can not be better than than $\mathcal{O}(\frac{1}{\sqrt{k}})$ where k is the number of iterations. Alternatively, smoothing as presented in [14] can be applied to mitigate non-smoothness of the objective function. In [15], a proximal iterative smoothing algorithm was proposed to solve convex nonsmooth optimization problems. In [7], the nondifferentiable function, which is approximated by the Moreau envelope function [16], is used in the column-wise mismatch formulation.

In this paper, an unconstrained off-grid DoA estimator is studied. It consists of one differentiable function and two nonsmooth ones, which are a regularized group-sparsity penalty and an indicator function. Inspired by [14, 17], the Nesterov smoothing methodology is used to reformulate the group-sparsity penalty into a ”max”-structure function and then add a strongly convex term to smooth it. We propose two reformulations for the group-sparsity penalty since ℓ_2/ℓ_1 mixed norm has a two-layer norm structure. Then, the accelerated proximal gradient (APG) [18] method is used on the smoothed optimization case. Note that our first proposed smoothing method is equivalent to the one in [7], as can be deduced from the results of [15]. The performance and computational efficiency of our second proposed method is demonstrated, and compared with the interior point method (CVX), MUSIC, M+LFBF, and CRLB.

2. PROBLEM FORMULATION

2.1. Problem Formulation of DoA Off-Grid Model

Consider a measurement model and its covariance described by

$$\mathbf{v}(t) = \sum_{k=1}^{K} \tilde{s}_k(t)\mathbf{a}(\theta_k) + \mathbf{n}(t) = \tilde{\mathbf{A}}(\boldsymbol{\theta})\tilde{\mathbf{s}}(t) + \mathbf{n}(t), \tag{2}$$

$$\mathbf{R}_\mathbf{v} = E[\mathbf{v}\mathbf{v}^H] = \sum_{k=1}^{K} \sigma_k^2 \mathbf{a}(\theta_k)\mathbf{a}(\theta_k)^H + \sigma_n^2\mathbf{I}, \tag{3}$$

where $\mathbf{v}(t) \in \mathbb{C}^{M\times 1}$ is the measurement vector, $\tilde{s}_k(t)$ is the k-th received signal with power σ_k^2 at time t, θ_k is the k-th unknown parameter of interest. For the direction finding problem in array processing, which is the example treated in this paper, $\mathbf{a}(\theta_k)$ denotes the steering vector for direction θ_k with m-th entry $e^{-j2\pi\frac{d_m}{\lambda}sin\theta_k}$, where λ is wavelength. $\tilde{\mathbf{A}}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_K)]$. $\mathbf{n}(t)$ is i.i.d. noise term with power σ_n^2. $\mathbf{R_v}$ denotes the covariance matrix of $\mathbf{v}(t)$, which is estimated as $\sum_{t=1}^{T}\mathbf{v}(t)\mathbf{v}(t)^H/T$ in practice from T snapshots.

In compressed sensing, the received signal is represented by a linear combination of a few atoms of a dictionary. The dictionary for the array model is composed of a large number of uniformly discretized grid atoms with parameters $\boldsymbol{\phi} = [\phi_1, \dots, \phi_N]$. However, in reality for the DoA estimation problem, the target location θ_k might not be on the grid such that $\beta_i = \theta_k - \phi_i$ if ϕ_i is closest to $\theta_k, \forall k$; otherwise, $\beta_i = 0$. We assume that $0 \le |\beta_i| \le r$ and $r = \frac{|\phi_i - \phi_{i+1}|}{2}$ is the half size of the grid interval. By using Taylor series expansion, the first-order approximate measurement model [19] is

$$\tilde{\mathbf{v}}(t) = (\tilde{\mathbf{A}}(\boldsymbol{\phi}) + \tilde{\mathbf{B}}\Gamma)\bar{\mathbf{s}}(t) + \mathbf{n}(t), \quad (4)$$

where $\tilde{\mathbf{B}} = [\frac{\partial \mathbf{a}(\phi_1)}{\partial \phi_1}, \dots, \frac{\partial \mathbf{a}(\phi_N)}{\partial \phi_N}] \in \mathbb{C}^{M\times N}, \boldsymbol{\beta} = [\beta_1, \dots, \beta_N]^T$, $\Gamma = diag(\boldsymbol{\beta})$, and $\bar{\mathbf{s}}$ is a $\mathbb{C}^{N\times 1}$ sparse vector. By vectorizing the covariance of (4), we have

$$\begin{aligned} \mathbf{y} &= (\mathbf{A}(\boldsymbol{\phi}) + \mathbf{B}\Gamma)\mathbf{s} + \sigma_n \mathbf{1}_n \quad (5)\\ &= (\mathbf{A}(\boldsymbol{\phi})\mathbf{s} + \mathbf{B}\mathbf{p}) + \sigma_n\mathbf{1}_n = [\mathbf{A}(\boldsymbol{\phi}), \mathbf{B}]\mathbf{x} + \sigma_n \mathbf{1}_n, \end{aligned}$$

where $\mathbf{y} = vec(\mathbf{R}_{\tilde{\mathbf{v}}})$, $\mathbf{A}(\boldsymbol{\phi}) = [\mathbf{a}(\phi_1)^H \otimes \mathbf{a}(\phi_1), \dots, \mathbf{a}(\phi_N)^H \otimes \mathbf{a}(\phi_N)] \in \mathbb{C}^{M^2\times N}$ $\mathbf{B} = [\frac{\partial \mathbf{a}(\phi_1)}{\partial\phi_1} \otimes \frac{\partial \mathbf{a}(\phi_1)}{\partial\phi_1}, \dots, \frac{\partial \mathbf{a}(\phi_N)}{\partial\phi_N} \otimes \frac{\partial \mathbf{a}(\phi_N)}{\partial\phi_N}] \in \mathbb{C}^{M^2\times N}$, and $\mathbf{s}$ is a $\mathbb{R}^{N\times 1}$ sparse vector with K nonzero terms σ_k^2's. $vec(\cdot)$ is the vectorizing operator, and $\otimes$ denotes the Kronecker product. $\mathbf{1}_n = [e_1^T, \dots, e_M^T]^T$ where $e_i \in \mathbb{R}^{M\times 1}$ is a all-zero vector except 1 at i-th entry. $\mathbf{x} = [\mathbf{s}^T, \mathbf{p}^T]^T \in \mathbb{R}^{2N\times 1}$, and $\mathbf{p} = \boldsymbol{\beta} \odot \mathbf{s}$ where $\odot$ denotes the Hadamard product. Let $\mathbf{G} = [\mathbf{A}(\boldsymbol{\phi}), \mathbf{B}]$ in the following sections. Note that if r is taken small, then $\mathbf{s} \gg \mathbf{p}$ since the value of β_k is much smaller than σ_k^2 at mild SNRs.

Since $\mathbf{s}, \mathbf{p}$ have the same sparsity pattern, we can solve (5) over a closed convex set $\mathcal{X}$ by group Lasso :

$$\begin{aligned} &\arg\min_{\mathbf{x}\in\mathcal{X}} \ \frac{1}{2}||\mathbf{y} - \mathbf{G}\mathbf{x}||_2^2 + \eta||\mathbf{x}||_{2,1}, \\ &\text{s.t. } \ \mathcal{X} = \{\mathbf{x} = [\mathbf{s}^T, \mathbf{p}^T]^T : \mathbf{s} \ge 0, -r\mathbf{s} \le \mathbf{p} \le r\mathbf{s}\}. \end{aligned} \quad (6)$$

where $\eta > 0$ is a regularization parameter, and r is defined previously. Because the constraint set $\mathcal{X}$ is simple, we can transform it into an unconstrained one by using an indicator function:

$$\arg\min_{\mathbf{x}\in\mathbb{R}^{2N\times 1}} F(\mathbf{x}) = \{\frac{1}{2}||\mathbf{y} - \mathbf{G}\mathbf{x}||_2^2 + \eta||\mathbf{x}||_{2,1} + \iota_{\mathcal{X}}(\mathbf{x})\}, \quad (7)$$

where $\iota_{\mathcal{X}}(\mathbf{x}) = \mathbf{0}$ if $\mathbf{x} \in \mathcal{X}$; otherwise, ∞. Let $f(\mathbf{x}) := \frac{1}{2}||\mathbf{y} - \mathbf{G}\mathbf{x}||_2^2$, $h(\mathbf{x}) := \eta||\mathbf{x}||_{2,1}$, and $g(\mathbf{x}) := \iota_{\mathcal{X}}(\mathbf{x})$. However, two nonsmooth functions in the objective makes this problem difficult to solve efficiently.

3. ALGORITHM

In this section, we will show how to deal with problem (7) by combining the accelerated proximal gradient algorithm with the smoothing technique. We aim to smooth the group-sparsity penalty $h(\mathbf{x}) = \eta||\mathbf{x}||_{2,1}$ so that the APG method can be used. In order to present the idea more clearly, we introduce the notation $||\mathbf{x}||_{2,1} = \sum_{g_i\in\Omega}\|\mathbf{x}_{g_i}\|_2$, where $\mathbf{x}_{g_i} \in \mathbb{R}^{|g_i|}$ denotes the subvector of $\mathbf{x}$ having the same sparse pattern in group g_i, where $|\cdot|$ is the cardinality of a set. Each group g_i represents a subset of index set $\{1, \cdots, 2N\}$ and is disjoint from the others. Denote $\Omega = \{g_1, \dots, g_{|\Omega|}\}$ as the set of groups, and $2N = \sum_{i=1}^{|\Omega|}|g_i|$. In our case, $|\Omega| = N$, $|g_i| = 2, g_i = \{i, i+N\}, \forall i = 1, \cdots, N$, $\mathbf{x}_{g_i} = [x_i, x_{i+N}]^T \in \mathbb{R}^2$ where $x_i = s_i$ and $x_{i+N} = p_i$. Denote x_i, s_i, and p_i as the i-th entry of $\mathbf{x}, \mathbf{s}$, and $\mathbf{p}$, respectively.

3.1. Reformulation of Group-sparsity Penalty

Since $h(\mathbf{x})$ is an ℓ_2/ℓ_1 mixed norm with two layers, i.e., the inner is ℓ_2 norm and the outer is ℓ_1 norm, we can utilize the dual norm property to reformulate it as a maximization of a linear function over an auxiliary variable with "simple" constraints in two different ways.

First, inspired by [17], by using the convex conjugate function and the fact that the dual norm of ℓ_2 norm is ℓ_2 norm, $\|\mathbf{x}_{g_i}\|_2$ has the max-structure as $\max_{\|\mathbf{u}_{g_i}\|_2\le 1}\mathbf{u}_{g_i}^T\mathbf{x}_{g_i}$ where $\mathbf{u}_{g_i} \in \mathbb{R}^{|g_i|}$ denotes an auxiliary vector. Then, $h(\mathbf{x})$ can be written as

$$\begin{aligned} h(\mathbf{x}) &= \eta\sum_{g_i\in\Omega}\|\mathbf{x}_{g_i}\|_2 = \sum_{g_i\in\Omega}\max_{\|\mathbf{u}_{g_i}\|_2\le 1}\{\eta\langle\mathbf{x}_{g_i}, \mathbf{u}_{g_i}\rangle\} \\ &= \max_{\mathbf{u}\in\mathcal{U}_{l_2}}\sum_{g_i\in\Omega}\{\eta\langle\mathbf{x}_{g_i}, \mathbf{u}_{g_i}\rangle\} = \max_{\mathbf{u}\in\mathcal{U}_{l_2}}\{\eta\langle\mathbf{x}, \mathbf{u}\rangle\}, \end{aligned} \quad (8)$$

where $\mathcal{U}_{l_2} = \{\mathbf{u} \in \mathbb{R}^{2N\times 1} : \|\mathbf{u}_{g_i}\|_2 \le 1, \forall g_i \in \Omega\}$ is the set of vectors in the space of the Cartesian product of ℓ_2 norm unit ball. In the Nesterov smoothing technique, if a nonsmooth convex function has the max-structure, then we have its corresponding smoothed function

$$h_\mu^{l_2}(\mathbf{x}) := \max_{\mathbf{u}\in\mathcal{U}_{l_2}}\{\eta\langle\mathbf{x}, \mathbf{u}\rangle - \mu d_{l_2}(\mathbf{u})\} \quad (9)$$

with a smoothing parameter $\mu > 0$. We suppose that a *prox-function* $d_{l_2}(\mathbf{u})$ [14] is continuous and strongly convex on $\mathcal{U}_{l_2}$ with a strong convexity parameter σ. Its *prox-center* of $d(\mathbf{u})$ is denoted by $\mathbf{u}_0 = \arg\min_{\mathbf{u}\in\mathcal{U}_{l_2}}\{d_{l_2}(\mathbf{u})\}$. By the definition of strongly convex, $d_{l_2}(\mathbf{u}) \ge \frac{\sigma}{2}\|\mathbf{u} - \mathbf{u}_0\|_2^2$. Since

$d_{l_2}(\mathbf{u})$ is strongly convex, $h_\mu^{l_2}(\mathbf{x})$ is a smooth and convex function so that its solution is unique and its gradient can be computed easily.

Second, inspired by by the fact that the dual norm of ℓ_1 norm is ℓ_∞ norm, $\|\mathbf{x}\|_1$ has the max-structure as $\max_{\|\mathbf{u}\|_\infty \leq 1} \mathbf{u}^T\mathbf{x}$, where $\mathbf{u}$ denotes an auxiliary vector. Therefore, we propose a second reformulation. Let us define $\nu_i := \|\mathbf{x}_{g_i}\|_2$ and $\boldsymbol{\nu} = [\nu_1, \ldots, \nu_{|\Omega|}]^T \in \mathbb{R}^{N\times 1}$, and then $h(\mathbf{x})$ can be rewritten as

$$h(\mathbf{x}) = \eta \sum_{g_i \in \Omega} \|\mathbf{x}_{g_i}\|_2 = \eta \sum_{i=1}^{|\Omega|} \nu_i = \eta\|\nu\|_1. \tag{10}$$

We define a new function $h(\boldsymbol{\nu})$ as

$$h(\boldsymbol{\nu}) = \eta\|\nu\|_1 = \max_{\mathbf{u}\in\mathcal{U}_{l_1}} \{\eta\langle\boldsymbol{\nu}, \mathbf{u}\rangle\}, \tag{11}$$

where $\mathcal{U}_{l_1} = \{\mathbf{u} \in \mathbb{R}^{N\times 1} : \|\mathbf{u}\|_\infty \leq 1\}$ is the set of vectors in the space of ℓ_∞ norm unit ball. Since it has the max-structure, we have its corresponding smoothed function

$$h_\mu^{l_1}(\boldsymbol{\nu}) := \max_{\mathbf{u}\in\mathcal{U}_{l_1}} \{\eta\langle\boldsymbol{\nu}, \mathbf{u}\rangle - \mu d_{l_1}(\mathbf{u})\} \tag{12}$$

with a smoothing parameter $\mu > 0$. Then, $h_\mu^{l_1}(\boldsymbol{\nu})$ is also a smooth and convex function if a strongly convex function $d_{l_1}(\mathbf{u})$ is chosen. Note that the dimension of $\mathbf{x}$ is twice as many as $\boldsymbol{\nu}$.

Since both $h_\mu^{l_2}(\mathbf{x})$ and $h_\mu^{l_1}(\boldsymbol{\nu})$ are smooth and convex, their gradients can be formed by the following modified theorem [14]

Theorem 1. *For any $\mu > 0$, the functions $h_\mu^{l_2}(\mathbf{x})$ and $h_\mu^{l_1}(\boldsymbol{\nu})$ are well-defined and continuously differentiable in $\mathbf{x}$ and $\boldsymbol{\nu}$, respectively. Moreover, both functions are convex and their gradients:*

$$\nabla h_\mu^{l_2}(\mathbf{x}) = \eta\mathbf{u}^{l_2}, \quad \nabla h_\mu^{l_1}(\boldsymbol{\nu}) = \eta\mathbf{u}^{l_1} \tag{13}$$

are Lipschitz continuous with the same constant $L_\mu = \frac{1}{\mu\sigma}$, where $\mathbf{u}^{l_2}$ and $\mathbf{u}^{l_1}$ are the optimal solutions to (9) and (12), respectively.

Suppose that $\forall\mathbf{u} \in \mathcal{U}_{l_2}$; we choose $d_{l_2}(\mathbf{u}) = \frac{1}{2}\|\mathbf{u}\|_2^2$ with a strong convexity parameter $\sigma = 1$. Then $\forall g_i$, $\mathbf{u}_{g_i}^{l_2}$, which is a subvector of $\mathbf{u}^{l_2}$, can be calculated as $\mathbf{u}_{g_i}^{l_2} = \mathcal{S}_2(\frac{\eta}{\mu}\mathbf{x}_{g_i})$ where $\mathcal{S}_2(\cdot)$ denotes the projection operator of projecting a vector $\mathbf{a}$ to a ℓ_2 unit ball

$$\mathcal{S}_2(\mathbf{a}) = \begin{cases} \frac{\mathbf{a}}{\|\mathbf{a}\|_2}, & \text{if } \|\mathbf{a}\|_2 > 1 \\ \mathbf{a}, & \text{if } \|\mathbf{a}\|_2 \leq 1. \end{cases} \tag{14}$$

Similarly, $\forall\mathbf{u} \in \mathcal{U}_{l_1}$, if we choose $d_{l_1}(\mathbf{u}) = \frac{1}{2}\|\mathbf{u}\|_2^2$, then $\mathbf{u}^{l_1}$ can be computed as $\mathbf{u}^{l_1} = \mathcal{S}_1(\frac{\eta}{\mu}\boldsymbol{\nu})$ where $\mathcal{S}_1(\cdot)$ denotes the projection operator of projecting a vector $\mathbf{a}$ to an ℓ_∞ unit ball

$$\mathcal{S}_1(\mathbf{a}) = \begin{cases} 1, & \text{if } a_i > 1, \forall i \\ a_i, & \text{if } |a_i| \leq 1, \forall i \\ -1, & \text{if } a_i < -1, , \forall i \end{cases} \tag{15}$$

where a_i is the i-th entry of $\mathbf{a}$.

Note that the dimension of $\boldsymbol{\nu}$ is a half of the one of $\mathbf{x}$. Therefore, for the case of $\nabla h_\mu^{l_1}(\boldsymbol{\nu})$, zero-padding is performed such that $\nabla h_\mu^{l_1}(\mathbf{x}) := [\nabla h_\mu^{l_1}(\boldsymbol{\nu})^T, \mathbf{0}^T]^T \in \mathbb{R}^{2N\times 1}$, where $\mathbf{0}$ is a $\mathbb{R}^{N\times 1}$ zero vector, so that a new gradient $\nabla h_\mu^{l_1}(\mathbf{x})$ can be used in the accelerated proximal gradient. This is acceptable only when parameter r is taken small enough. Since $\mathbf{p} \ll \mathbf{s}$ holds in this case, the value of ν_i mainly comes from the contribution of $\mathbf{s}$, so that zero vector can be assigned as the partial derivative of $\mathbf{p}$.

3.2. Accelerated Smoothing Proximal Gradient (ASPG)

Now, we solve two "smoothed" versions of problem (7)

$$\arg\min_{\mathbf{x}\in\mathbb{R}^n} \{H_i(\mathbf{x}) + \iota_{\mathcal{X}}(\mathbf{x})\}, i = 1 \text{ or } 2. \tag{16}$$

where $H_i(\mathbf{x}) := f(\mathbf{x}) + h_\mu^{l_i}(\mathbf{x})$, $i = 1$ or 2, and then its gradient is computed as $\nabla H_i(\mathbf{x}) = \nabla f(\mathbf{x}) + \eta\mathbf{u}^{l_i}$.

Problem (16) is suggested to be solved by the accelerated proximal gradient method [18] in which a proximal operator is used:

$$\operatorname{prox}_\iota(\mathbf{y}) = \arg\min_{\mathbf{x}\in\mathbb{R}^n} \{\frac{1}{2}\|\mathbf{y} - \mathbf{x}\|^2 + \iota(\mathbf{x})\}. \tag{17}$$

In fact, the proximal operator $\operatorname{prox}_{\iota_{\mathcal{X}}}(\mathbf{y})$ of indicator function $\iota_{\mathcal{X}}(\mathbf{x})$ is the projection operator onto the set $\mathcal{X}$, $\Pi_{\mathcal{X}}(\mathbf{x})$.

The ASPG method is summarized in the Algorithm 1. We also show its convergence rate by the following theorem:

Theorem 2. *Suppose $\mathbf{x}^k$ is the k-th iterative solution in Algorithm 1, and $\mathbf{x}^*$ is the optimal solution of problem (7). Assume that ϵ-approximation is required, i.e., $F(\mathbf{x}^k) - F(\mathbf{x}^*) \leq \epsilon$. If we set $\mu = \frac{\epsilon}{2D_i}$, where $D_i = \max_{\mathbf{u}\in\mathcal{U}_{l_i}} d_{l_i}(\mathbf{u})$, then*

$$F(\mathbf{x}^k) - F(\mathbf{x}^*) \leq \frac{\epsilon}{2} + \frac{2(L_f + 2\frac{D_i}{\epsilon\sigma})\|\mathbf{x}^0 - \mathbf{x}^*\|^2}{(k+1)^2}, \tag{18}$$

where L_f is Lipschitz continuous gradient parameter of $f(\mathbf{x})$. The number of iteration k has an upper bound by

$$\sqrt{\frac{4\|\mathbf{x}^0 - \mathbf{x}^*\|^2}{\epsilon}(L_f + \frac{2D_i}{\epsilon\sigma})} - 1 \tag{19}$$

Since space is limited, the proof detail will be provided in the journal version of this paper. This theorem implies its convergence rate is $\mathcal{O}(\frac{1}{k})$. We cannot achieve convergence rate $\mathcal{O}(\frac{1}{k^2})$ of accelerated proximal gradient method due to the smoothing process, but better than the subgradient methods with $\mathcal{O}(\frac{1}{\sqrt{k}})$ [12].

4. NUMERICAL EXAMPLE

In the following numerical example of the off-grid DoA estimation, the proposed two accelerated smoothing proximal

Algorithm 1 Accelerated Smoothing Proximal Gradient

Input: $\mathbf{x}^0 = \mathbf{x}^1 = \mathbf{0}$; $\gamma = 0.5$; $\mu = 10^{-8}$; step-size $\alpha^0 = 1$;

Step k: $(k \geq 1)$ Let $\alpha := \alpha^{k-1}$. Compute
$\mathbf{w}^{k+1} = \mathbf{x}^k + \frac{k}{k+3}(\mathbf{x}^k - \mathbf{x}^{k-1})$

repeat
  Compute $\nabla f(\mathbf{w}^{k+1}) = \mathbf{G}^H(\mathbf{G}\mathbf{w}^{k+1} - \mathbf{y})$,
  Compute $\nabla h_\mu^{l_i}(\mathbf{w}^{k+1}) = \eta \mathbf{u}^{l_2}$ if $i = 2$,
  Compute $\nabla h_\mu^{l_i}(\mathbf{w}^{k+1}) = \eta \mathbf{u}^{l_1}$ if $i = 1$,
  $\mathbf{z} = \Pi_{\mathcal{X}}(\mathbf{w}^{k+1} - \alpha \nabla f(\mathbf{w}^{k+1}) - \alpha \nabla h_\mu^{l_i}(\mathbf{w}^{k+1}))$,
  Break if $F_i(\mathbf{z}) \leq \hat{F}_i^{\alpha}(\mathbf{z}, \mathbf{w}^{k+1}) = F_i(\mathbf{w}^{k+1}) + (\nabla F_i(\mathbf{w}^{k+1}))^T(\mathbf{z} - \mathbf{w}^{k+1}) + \frac{1}{2\alpha}\|\mathbf{z} - \mathbf{w}^{k+1}\|_2^2$,
  Update $\alpha := \gamma \alpha$,
return $\alpha^k := \alpha$, $\mathbf{x}^{k+1} := \mathbf{z}$

Note 1: $\mathbf{u}^{l_2}$ is composed of $\mathbf{u}_{g_i}^{l_2} = \mathcal{S}_2(\frac{\eta}{\mu}\mathbf{w}_{g_i}^{k+1}), \forall g_i$.

Note 2: $\mathbf{u}^{l_1} = [\mathcal{S}_1(\frac{\eta}{\mu}\boldsymbol{\nu})^T, \mathbf{0}^T]^T$ where $\nu_i = \|\mathbf{w}_{g_i}^{k+1}\|_2$, ν_i : i-th entry of $\boldsymbol{\nu}$

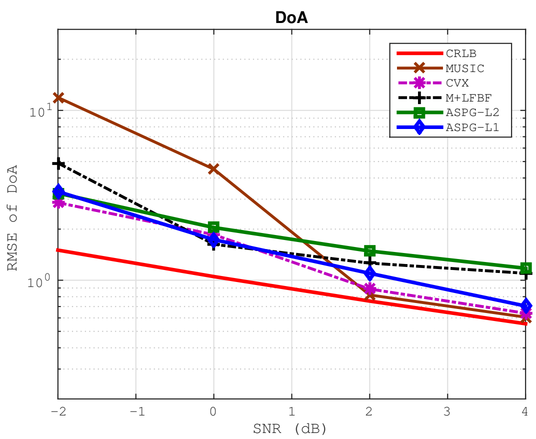

Fig. 1. RMSE of DoA estimation versus SNR.

gradient methods are designated as ASPG-L2 (using $h_\mu^{l_2}(\mathbf{x})$) and ASPG-L1 (using $\nabla h_\mu^{l_1}(\boldsymbol{\nu})$). We also solve problem (6) by using CVX packages. The CVX method can be viewed as a benchmark, which is used to evaluate the estimation performance degradation caused by smoothing in the proposed methods. The estimation errors of these methods are compared with the same for the MUSIC estimator, M+LFBF and the CRLB. Consider $K = 2$ source signals from DoAs $\boldsymbol{\theta} = [13.2220, 28.6022]$ degree impinging on a uniform linear array of $M = 8$ sensors with half-wavelength interelement spacing. The two sources are randomly generated with normal distribution of zero mean and variance σ_s^2. The noise term is i.i.d. AWGN with zero mean and variance σ_n^2. We use one hundred snapshots to estimate the covariance matrix. The size N of search grid is set to 360 with $r = 0.25$ degree, which is used for all methods. In the ASPG method,

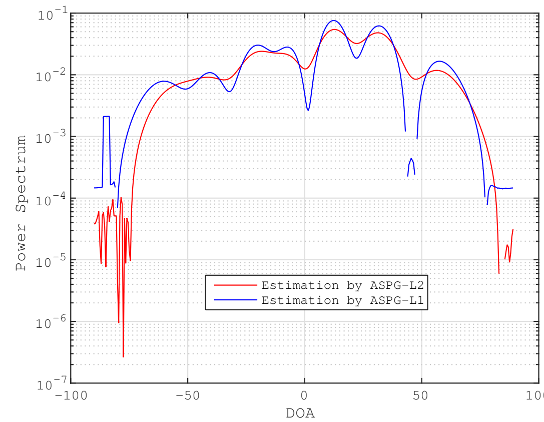

Fig. 2. Power Spectrum versus DoA.

$d_{l_i}(\mathbf{u}) = \frac{1}{2}\|\mathbf{u}\|_2^2, \forall i$, the decreasing factor is $\gamma = 0.5$, and smoothing parameter is chosen as $\mu = 10^{-8}$. The root-mean-square-error (RMSE) of DoA estimation is $(E[\frac{1}{K}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\|_2^2])^{\frac{1}{2}}$. One hundred realizations are performed at each SNR.

In Figure 1, the RMSE of CVX and the proposed methods are almost the same and better than MUSIC and M+LFBF at low SNRs. At high SNRs, the performance of ASPG-L1, CVX, and MUSIC approach CRLB, but not ASPG-L2. The reason is that the sparse property of group-sparsity penalty $\|\mathbf{x}_{g_i}\|_2$ is lost during the smoothing process by only using the property that the dual norm of ℓ_2 norm is also ℓ_2 norm so that sparsity is not promoted in this way. In Figure 2, the estimated power spectrum of ASPG methods is presented at SNR $= 0$ dB. Due to the smoothing process, both have lost their sparsity. However, the two peaks of ASPG-L1 are more separated than ASPG-L2. In other words, ASPG-L1 estimator owns higher DoA resolution.

We also have verified that the computational efficiencies of the proposed methods are better than the CVX method. At SNR $= 0$ dB, the running time at each realization of ASPG-L2, and ASPG-L1 are 2.54s and 2.74s, which are faster than the CVX method with 22.51s, and M+LFBF with 5.59s.

5. CONCLUSION

Two ASPG methods were proposed for the estimation of off-grid DoAs using a sparse model for the observation. The group-sparsity penalty is reformulated and smoothed by the Nesterov smoothing technique so that its gradient can be calculated easily. Then, the accelerated proximal gradient is used to solve the unconstrained optimization problem with the smoothed objective functions plus only one nonsmooth function. The performance and computational efficiencies of the proposed methods were verified by a numerical example.

6. REFERENCES

[1] David L Donoho, "Compressed sensing," *IEEE Transactions on information theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[2] Emmanuel J Candès, Justin Romberg, and Terence Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Transactions on information theory*, vol. 52, no. 2, pp. 489–509, 2006.

[3] Dmitry Malioutov, Müjdat Çetin, and Alan S Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Transactions on Signal Processing*, vol. 53, no. 8, pp. 3010–3022, 2005.

[4] Hao Zhu, Geert Leus, and Georgios B Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2002–2016, 2011.

[5] Zai Yang, Cishen Zhang, and Lihua Xie, "Robustly stable signal recovery in compressed sensing with structured matrix perturbation," *IEEE Transactions on Signal Processing*, vol. 60, no. 9, pp. 4658–4671, 2012.

[6] Jimeng Zheng and Mostafa Kaveh, "Sparse spatial spectral estimation: a covariance fitting algorithm, performance and regularization," *IEEE Transactions on Signal Processing*, vol. 61, no. 11, pp. 2767–2777, 2013.

[7] Zhao Tan, Peng Yang, and Arye Nehorai, "Joint sparse recovery method for compressed sensing with structured dictionary mismatches," *Signal Processing, IEEE Transactions on*, vol. 62, no. 19, pp. 4997–5008, 2014.

[8] Gongguo Tang, Badri Narayan Bhaskar, Parikshit Shah, and Benjamin Recht, "Compressed sensing off the grid," *IEEE Transactions on Information Theory*, vol. 59, no. 11, pp. 7465–7490, 2013.

[9] Rakshith Jagannath and KVS Hari, "Block sparse estimator for grid matching in single snapshot doa estimation," *IEEE Signal Processing Letters*, vol. 20, no. 11, pp. 1040–1043, 2013.

[10] Bo Lin, Jiying Liu, Meihua Xie, and Jubo Zhu, "Super-resolution doa estimation using single snapshot via compressed sensing off the grid," in *Signal Processing, Communications and Computing (ICSPCC), 2014 IEEE International Conference on*. IEEE, 2014, pp. 825–829.

[11] Nikos Komodakis and Jean-Christophe Pesquet, "Playing with duality: An overview of recent primal? dual approaches for solving large-scale optimization problems," *Signal Processing Magazine, IEEE*, vol. 32, no. 6, pp. 31–54, 2015.

[12] Naum Zuselevich Shor, *Minimization methods for non-differentiable functions*, vol. 3, Springer Science & Business Media, 2012.

[13] Yurii Nesterov, *Introductory lectures on convex optimization: A basic course*, vol. 87, Springer Science & Business Media, 2013.

[14] Yurii Nesterov, "Smooth minimization of non-smooth functions," *Mathematical programming*, vol. 103, no. 1, pp. 127–152, 2005.

[15] Francesco Orabona, Andreas Argyriou, and Nathan Srebro, "Prisma: Proximal iterative smoothing algorithm," *arXiv preprint arXiv:1206.2372*, 2012.

[16] Jean-Jacques Moreau, "Proximité et dualité dans un espace hilbertien," *Bulletin de la Société mathématique de France*, vol. 93, pp. 273–299, 1965.

[17] Xi Chen, Qihang Lin, Seyoung Kim, Jaime G Carbonell, Eric P Xing, et al., "Smoothing proximal gradient method for general structured sparse regression," *The Annals of Applied Statistics*, vol. 6, no. 2, pp. 719–752, 2012.

[18] Neal Parikh and Stephen P Boyd, "Proximal algorithms.," *Foundations and Trends in optimization*, vol. 1, no. 3, pp. 127–239, 2014.

[19] Cheng-Yu Hung, Jimeng Zheng, and Mostafa Kaveh, "Directions of arrival estimation by learning sparse dictionaries for sparse spatial spectra," in *Sensor Array and Multichannel Signal Processing Workshop (SAM), 2014 IEEE 8th*. IEEE, 2014, pp. 377–380.