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

Cheng-Yu Hung
Mostafa Kaveh

ECE Department, University of Minnesota

March 7, 2017

Outline

Problem Description

Sparse Off-grid DoA Model

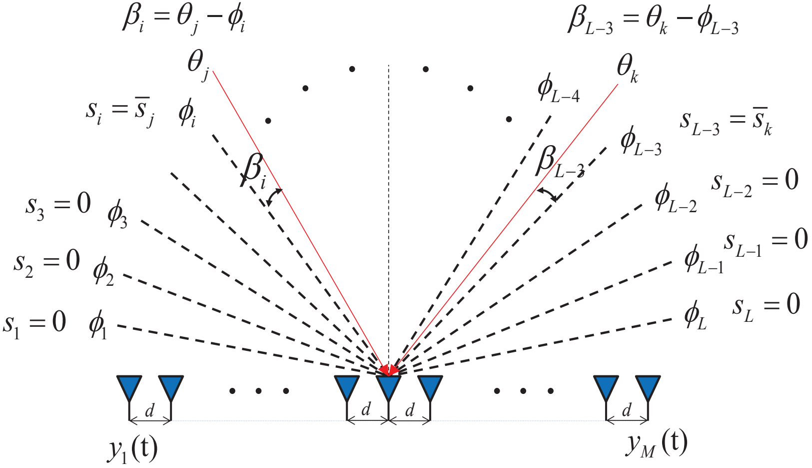
The Proposed Methods

Numerical Rresults

Conclusions

Problem Description (1): Pictorial Sketch

Figure: Sparse Off-grid DoA.



Problem Description (2)

- Off-grid DoA estimation problem can be formulated as a sparse model with a structured perturbation
- Can be solved by basis pursuit denoising (BPDN) solver w/ linear constraints
  - pros: easy to implement, e.g. use CVX
  - cons: large computational complexity
- Aim to solve this problem efficiently
- Contributions of this work
  - Propsed an efficient algorithm
  - Analyzed convergence rate

Outline

Problem Description

Sparse Off-grid DoA Model

The Proposed Methods

Numerical Rresults

Conclusions

Sparse Off-grid DoA Model (1)

- True DoAs not on the grid
- Structured perturbation $\mathbf{E} = \mathbf{B}\Gamma$ is first-order term in Taylor expansion

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

  - Sensing matrix $\mathbf{A}(\phi) \in \mathbb{C}^{M \times N}$ is known and parametrized by $\phi = [\phi_1, \ldots, \phi_N]$
  - Mismatch matrix $\mathbf{B} \in \mathbb{C}^{M \times N}$ is known
  - Diagonal matrix $\Gamma = diag(\boldsymbol{\beta})$ with off-grid $\boldsymbol{\beta} = [\beta_1, \ldots, \beta_N]^T$ is unknown, $0 \leq |\beta_i| \leq r$ and $r = \frac{|\phi_i - \phi_{i+1}|}{2}$ is the half size of the grid interval
  - K-sparse signal of interest $\mathbf{s} \in \mathbb{R}^{N \times 1}$
- Let $\mathbf{p} = \boldsymbol{\beta} \odot \mathbf{s}$, where $\odot$: Hadamard product. $\mathbf{G} = [\mathbf{A}(\phi), \mathbf{B}]$, $\mathbf{x} = [\mathbf{s}^T, \mathbf{p}^T]^T \in \mathbb{R}^{2N \times 1}$, where $\mathbf{s}, \mathbf{p}$ share the same sparsity.

$$\mathbf{y} = (\mathbf{A}(\phi)\mathbf{s} + \mathbf{B}\mathbf{p}) + \mathbf{n} = \mathbf{G}\mathbf{x} + \mathbf{n}, \tag{2}$$

Sparse Off-grid DoA Model (2)

- Solve (2) by BPDN w/ linear inequalities constraints

$$\begin{aligned} &\arg\min_{\mathbf{x}\in\mathcal{X}} \ \frac{1}{2}||\mathbf{y}-\mathbf{G}\mathbf{x}||_2^2+\eta||\mathbf{x}||_{2,1}, \\ &\text{s.t. } \ \mathcal{X}=\{\mathbf{x}=[\mathbf{s}^T,\mathbf{p}^T]^T:\mathbf{s}\geq 0,-r\mathbf{s}\leq\mathbf{p}\leq r\mathbf{s}\}. \end{aligned} \tag{3}$$

- Unconstrained problem

$$\arg\min_{\mathbf{x}\in\mathbb{R}^{2N\times 1}} F(\mathbf{x})=\{f(\mathbf{x})+h(\mathbf{x})+\iota_{\mathcal{X}}(\mathbf{x})\}, \tag{4}$$

$f(\mathbf{x}):=\frac{1}{2}||\mathbf{y}-\mathbf{G}\mathbf{x}||_2^2$, $h(\mathbf{x}):=\eta||\mathbf{x}||_{2,1}=\eta\sum_{g_i\in\Omega}\|\mathbf{x}_{g_i}\|_2$.
$\iota_{\mathcal{X}}(\mathbf{x})$: indicator function.

- Smooth one of nonsmooth functions

Sparse Off-grid DoA Model (3)

Examples of Related Work

- Off-grid DoA estimation algorithms
  - (Zhu, Leus, and Giannakis 2011)
  - (Yang, Zhang, and Xie 2012)
  - (Tan, Yang, and Nehorai 2013)
  - (Hung, Zheng, and Kaveh 2014)
- Subgradient method
  - (Ben-Tal and Teboulle 1989)
  - (Shor 2012)
- Smoothing technique
  - Nesterov smoothing (Nesterov 2005)
  - Moreau envelope (Moreau 1965)
- Accelerated proximal gradient
  - (Beck and Teboulle 2009)
- Forward-backward-forward method
  - (Combettes and Pesquet 2012)

Outline

Problem Description

Sparse Off-grid DoA Model

The Proposed Methods

Numerical Rresults

Conclusions

The Proposed Methods (1): l_2-based Smoothing

- Method 1: Reformulation of group-sparsity penalty $h(\mathbf{x})$ by considering the dual norm of l_2 norm

$$
\begin{aligned}
h(\mathbf{x}) = \eta \sum_{g_i \in \Omega} \|\mathbf{x}_{g_i}\|_2 &= \sum_{g_i \in \Omega} \max_{\|\mathbf{u}_{g_i}\|_2 \leq 1} \{\eta \langle \mathbf{x}_{g_i}, \mathbf{u}_{g_i} \rangle\} \\
&= \max_{\mathbf{u} \in \mathcal{U}_{l_2}} \sum_{g_i \in \Omega} \{\eta \langle \mathbf{x}_{g_i}, \mathbf{u}_{g_i} \rangle\} = \max_{\mathbf{u} \in \mathcal{U}_{l_2}} \{\eta \langle \mathbf{x}, \mathbf{u} \rangle\}, \qquad (5)
\end{aligned}
$$

  where $\mathcal{U}_{l_2} = \{\mathbf{u} \in \mathbb{R}^{2N \times 1} : \|\mathbf{u}_{g_i}\|_2 \leq 1, \forall g_i \in \Omega\}$
- Inspired by (Nesterov 2005), use a *prox-function* $d_{l_2}(\mathbf{u})$ (continuous and strongly convex on $\mathcal{U}_{l_2}$, i.e., $d_{l_2}(\mathbf{u}) \geq \frac{\sigma}{2}\|\mathbf{u} - \mathbf{u}_0\|_2^2$)
- $h_\mu^{l_2}(\mathbf{x})$: smooth and convex

$$
h_\mu^{l_2}(\mathbf{x}) := \max_{\mathbf{u} \in \mathcal{U}_{l_2}} \{\eta \langle \mathbf{x}, \mathbf{u} \rangle - \mu d_{l_2}(\mathbf{u})\} \qquad (6)
$$

The Proposed Methods (2): l_1-based Smoothing

- Method 2: Consider the dual norm of l_1 norm
- Define $\nu_i := \|\mathbf{x}_{g_i}\|_2$ and $\boldsymbol{\nu} \in \mathbb{R}^{N\times 1}$

$$h(\mathbf{x}) = \eta \sum_{g_i \in \Omega} \|\mathbf{x}_{g_i}\|_2 = \eta \sum_{i=1}^{|\Omega|} \nu_i = \eta\|\nu\|_1. \tag{7}$$

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

$$h(\boldsymbol{\nu}) = \eta\|\nu\|_1 = \max_{\mathbf{u}\in\mathcal{U}_{l_1}} \{\eta\langle\boldsymbol{\nu}, \mathbf{u}\rangle\}, \tag{8}$$

where $\mathcal{U}_{l_1} = \{\mathbf{u} \in \mathbb{R}^{N\times 1} : \|\mathbf{u}\|_\infty \leq 1\}$

- $h_\mu^{l_2}(\mathbf{x})$: smooth and convex w/ a strongly convex function $d_{l_1}(\mathbf{u})$

$$h_\mu^{l_1}(\boldsymbol{\nu}) := \max_{\mathbf{u}\in\mathcal{U}_{l_1}} \{\eta\langle\boldsymbol{\nu}, \mathbf{u}\rangle - \mu d_{l_1}(\mathbf{u})\} \tag{9}$$

The Proposed Methods (3)

Theorem

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:

$$\nabla h_\mu^{l_2}(\mathbf{x}) = \eta \mathbf{u}^{l_2}, \qquad \nabla h_\mu^{l_1}(\boldsymbol{\nu}) = \eta \mathbf{u}^{l_1} \tag{10}$$

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 (6) and (9), respectively.

The Proposed Methods (4): Examples

- Choose $d_{l_2}(\mathbf{u}) = \frac{1}{2}\|\mathbf{u}\|_2^2$
- $\nabla h_\mu^{l_2}(\mathbf{x}) = \eta \mathbf{u}^{l_2} \in \mathbb{R}^{2N\times 1}$ with $\mathbf{u}_{g_i}^{l_2} = \mathcal{S}_2(\frac{\eta}{\mu}\mathbf{x}_{g_i}), \forall g_i$ where $\mathcal{S}_2(\cdot)$ denotes the projection operator onto a ℓ_2 unit ball

$$\mathcal{S}_2(\mathbf{a}) = \begin{cases} \frac{\mathbf{a}}{\|\mathbf{a}\|_2}, & \text{if } \|\mathbf{a}\|_2 > 1 \\ \mathbf{a}, & \text{if } \|\mathbf{a}\|_2 \leq 1. \end{cases} \tag{11}$$

- $\nabla h_\mu^{l_1}(\boldsymbol{\nu}) = \eta \mathbf{u}^{l_1} \in \mathbb{R}^{N\times 1}$ with $\mathbf{u}^{l_1} = \mathcal{S}_1(\frac{\eta}{\mu}\boldsymbol{\nu})$ where $\mathcal{S}_1(\cdot)$ denotes the projection operator onto an ℓ_∞ unit ball

$$\mathcal{S}_1(\mathbf{a}) = \begin{cases} 1, & \text{if } a_i > 1, \forall i \\ a_i, & \text{if } |a_i| \leq 1, \forall i \\ -1, & \text{if } a_i < -1, , \forall i \end{cases} \tag{12}$$

  - 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
  - Feasible if $\mathbf{p} \ll \mathbf{s}$ holds

The Proposed Methods (5): Accelerated Smoothing Proximal Gradient

Accelerated Smoothing Proximal Gradient (ASPG)

- Solve

$$\arg\min_{\mathbf{x}\in\mathbb{R}^n}\{H_i(\mathbf{x}) + \iota_{\mathcal{X}}(\mathbf{x})\}, i = 1 \text{ or } 2. \tag{13}$$

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

- Apply accelerated proximal gradient method (Parikh and Boyd 2014) in which a proximal operator is used for any function $\iota(\mathbf{x})$:

$$\text{prox}_\iota(\mathbf{y}) = \arg\min_{\mathbf{x}\in\mathbb{R}^n}\{\frac{1}{2}\|\mathbf{y} - \mathbf{x}\|^2 + \iota(\mathbf{x})\}. \tag{14}$$

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

The Proposed Methods (6): Algorithm

Algorithm: 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

$\mathbf{w}^{k+1} = \mathbf{x}^k + \frac{k}{k+3}(\mathbf{x}^k - \mathbf{x}^{k-1})$

1: **repeat**

2: Compute $\nabla f(\mathbf{w}^{k+1}) = \mathbf{G}^H(\mathbf{G}\mathbf{w}^{k+1} - \mathbf{y})$,

3: Compute $\nabla h_\mu^{l_i}(\mathbf{w}^{k+1}) = \eta \mathbf{u}^{l_2}$ if $i = 2$,

4: Compute $\nabla h_\mu^{l_i}(\mathbf{w}^{k+1}) = \eta \mathbf{u}^{l_1}$ if $i = 1$,

5: $\mathbf{z} = \Pi_{\mathcal{X}}(\mathbf{w}^{k+1} - \alpha \nabla f(\mathbf{w}^{k+1}) - \alpha \nabla h_\mu^{l_i}(\mathbf{w}^{k+1}))$,

6: Break if $F_i(\mathbf{z}) \leq \hat{F}_i^{\alpha}(\mathbf{z}, \mathbf{w}^{k+1}) =$
$F_i(\mathbf{w}^{k+1}) + (\nabla F_i(\mathbf{w}^{k+1}))^T(\mathbf{z} - \mathbf{w}^{k+1}) + \frac{1}{2\alpha}\|\mathbf{z} - \mathbf{w}^{k+1}\|_2^2$,

7: Update $\alpha := \gamma\alpha$,

8: **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 $\boldsymbol{\nu}$

The Proposed Methods (7): Convergence Analysis

Theorem

Suppose $\mathbf{x}^k$ *is the* k*-th iterative solution in Algorithm, and* $\mathbf{x}^*$ *is the optimal solution of problem (4). 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*

$$F(\mathbf{x}^k) - F(\mathbf{x}^*) \leq \frac{\epsilon}{2} + \frac{2(L_f + 2\frac{D_i}{\epsilon\sigma})\|\mathbf{x}^0 - \mathbf{x}^*\|^2}{(k+1)^2}, \qquad (15)$$

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

$$\sqrt{\frac{4\|\mathbf{x}^0 - \mathbf{x}^*\|^2}{\epsilon}(L_f + \frac{2D_i}{\epsilon\sigma})} - 1 \qquad (16)$$

- Convergence rate is $\mathcal{O}(\frac{1}{k})$, better than subgradient methods with $\mathcal{O}(\frac{1}{\sqrt{k}})$

Outline

Problem Description

Sparse Off-grid DoA Model

The Proposed Methods

Numerical Rresults

Conclusions

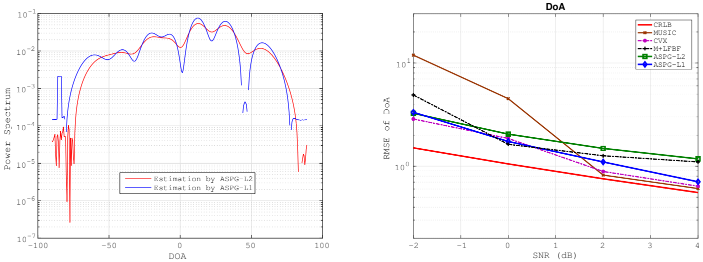
Numerical Rresults (1)

DoA estimation problem

- ULA with $M = 8$ sensors with $d/\lambda = 0.5$, steering vector: $\mathbf{g}(\theta_i) = [e^{-j(-(M-1)/2)2\pi \frac{d}{\lambda} sin\theta_i}, \ldots, e^{-j((M-1)/2)2\pi \frac{d}{\lambda} sin\theta_i}]^T$
- $K = 2$ plane waves with the actual DoAs =[$13.2220^\circ, 28.6022^\circ$].
- $L = 100$ multiple sanpshots.
- 100 realizations at SNRs
- DoA search grid is $-90^\circ \sim 90^\circ$ with 1° separation, N=180
- smoothing parameter $\mu = 10^{-8}$

Numerical Rresults (2)

DoA estimation problem: Performance comparison



(a) Power spectrum vs DoA at SNR=0 dB

(b) RMSE of DoA

Figure: DoA estimation.

Numerical Rresults (3)

DoA estimation problem: Computational complexity

Table: CPU Time (seconds) of Methods at SNR=0 dB

Algorithm	CVX	MUSIC	ASPG-L2	ASPG-L1	M+LFBF
M=8	22.52s	0.0061s	2.54s	2.74s	5.59s

Outline

Problem Description

Sparse Off-grid DoA Model

The Proposed Methods

Numerical Rresults

Conclusions

Conclusion

- Proposed two methods to smooth group sparsity penalty
  - Resolution comparison in power spectrum
  - l_1-based smoothing method better than L2-based
- RMSE performance
  - l_1-based smoothing method approaches CVX method
- Computational complexity
  - lower than CVX method and M+LFBF
- Convergence rate
  - better than subgradient method
- Have to choose the smoothing parameter carefully

Thank You

References I

Beck, Amir and Marc Teboulle (2009). “Gradient-based algorithms with applications to signal recovery”. In: *Convex optimization in signal processing and communications*, pp. 42–88.

Ben-Tal, A and M Teboulle (1989). “A smoothing technique for nondifferentiable optimization problems”. In: *Optimization*. Springer, pp. 1–11.

Combettes, Patrick L and Jean-Christophe Pesquet (2012). “Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators”. In: *Set-Valued and variational analysis* 20.2, pp. 307–330.

References II

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

Moreau, Jean-Jacques (1965). “Proximité et dualité dans un espace hilbertien”. In: *Bulletin de la Société mathématique de France* 93, pp. 273–299.

Nesterov, Yurii (2005). “Smooth minimization of non-smooth functions”. In: *Mathematical programming* 103.1, pp. 127–152.

Parikh, Neal and Stephen P Boyd (2014). “Proximal Algorithms.” In: *Foundations and Trends in optimization* 1.3, pp. 127–239.

Shor, Naum Zuselevich (2012). *Minimization methods for non-differentiable functions*. Vol. 3. Springer Science & Business Media.

References III

Tan, Zhao, Peng Yang, and Arye Nehorai (2013). "Joint-sparse recovery in compressed sensing with dictionary mismatch". In: *Computational Advances in Multi-Sensor Adaptive Processing, IEEE 5th International Workshop on*.

Yang, Zai, Cishen Zhang, and Lihua Xie (2012). "Robustly stable signal recovery in compressed sensing with structured matrix perturbation". In: *IEEE Transactions on Signal Processing* 60.9, pp. 4658–4671.

Zhu, Hao, Geert Leus, and Georgios B Giannakis (2011). "Sparsity-cognizant total least-squares for perturbed compressive sampling". In: *IEEE Transactions on Signal Processing* 59.5, pp. 2002–2016.