# Convex Optimization for Signal Processing and Communications: From Fundamentals to Applications

**Chong-Yung Chi** 

Institute of Communications Engineering & Department of Electrical Engineering National Tsing Hua University, Taiwan 30013 E-mail: cychi@ee.nthu.edu.tw Web: http://www.ee.nthu.edu.tw/cychi/

Invited tutorial talk at Aalborg University, Denmark, 2018/9/20

Acknowledgment: My post doctor, **Dr. Chia-Hsiang Lin** for preparing slides of Part II. My Ph.D. student, **Yao-Rong Syu** for some slides preparation.





## **1** Part I: Fundamentals of Convex Optimization

- Part II: Application in Hyperspectral Image Analysis: (Big Data Analysis and Machine Learning)
- **3** Part III: Application in Wireless Communications (5G Systems)
  - Subsection I: Outage Constrained Robust Transmit Optimization for Multiuser MISO Downlinks
  - Subsection II: Outage Constrained Robust Hybrid Coordinated Beamforming for Massive MIMO Enabled Heterogeneous Cellular Networks

#### • Optimization problem:

 $\begin{array}{ll} \text{minimize} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in \mathcal{C} \end{array} \tag{1}$ 

where  $f(\mathbf{x})$  is the objective function to be minimized and  $\mathcal{C}$  is the feasible set from which we try to find an optimal solution. Let

 $\mathbf{x}^{\star} = \arg\min_{\mathbf{x}\in\mathcal{C}} f(\mathbf{x}) \quad \text{(optimal solution or global minimizer)} \tag{2}$ 

#### • Challenges in applications:

• Local optima; large problem size; decision variable x involving real and/or complex vectors, matrices; feasible set C involving generalized inequalities, etc.

• Computational complexity: NP-hard; polynomial-time solvable.

• Performance analysis: Performance insights, properties, perspectives, proofs (e.g., identifiability and convergence), limits and bounds.

# Convex sets and convex functions-1

• Affine (convex) combination: Provided that C is a nonempty set,

$$\mathbf{x} = \theta_1 \mathbf{x}_1 + \dots + \theta_K \mathbf{x}_K, \ \mathbf{x}_i \in C \ \forall i$$
(3)

is called an *affine (a convex) combination* of  $\mathbf{x}_1, \ldots, \mathbf{x}_K$  (*K* vectors or points of a set) if  $\sum_{i=1}^{K} \theta_i = 1, \ \theta_i \in \mathbb{R}$  ( $\theta_i \in \mathbb{R}_+$ ),  $K \in \mathbb{Z}_{++}$ .

• Affine (convex) set:

• *C* is an *affine (a convex) set* if *C* is closed under the operation of *affine (convex) combination*;

- an affine set is constructed by *lines*;
- a convex set is constructed by *line segments*.

#### • Conic set:

- If  $\theta \mathbf{x} \in C$  for any  $\theta \in \mathbb{R}_+$  and any  $\mathbf{x} \in C$ , then the set C is a *cone* and it is constructed by *rays starting from the origin*;
- the linear combination (3) is called a *conic combination* if  $\theta_i \ge 0 \ \forall i$ ;

## **Convex sets and convex functions-3**



Figure 1: An illustration in  $\mathbb{R}^3$ , where **conv**{ $a_1, a_2, a_3$ } is a simplex defined by the shaded triangle, and **conv**{ $a_1, a_2, a_3, a_4$ } is a simplex (and also a simplest simplex) defined by the tetrahedron with the four extreme points { $a_1, a_2, a_3, a_4$ }.

## Convex Set Examples



• Left plot: conic C (called the *conic hull of* C) is a *convex cone* formed by  $C = {\mathbf{x}_1, \mathbf{x}_2}$  via conic combinations, i.e., *the smallest conic set that contains* C; right plot: conic C formed by another set C (star).

# First-order Condition and Epigraph



Left plot: first-order condition for a convex function f for the one-dimensional case: f(b) ≥ f(a) + f'(a)(b - a), for all a, b ∈ dom f; right plot: the epigraph of a convex function f : ℝ → ℝ.

# **Convex optimization problem**

• Convex problem:

(CVXP) 
$$p^* = \min_{\mathbf{x} \in \mathcal{C}} f(\mathbf{x})$$
 (8)

is a convex problem if the objective function  $f(\cdot)$  is a convex function and C is a convex set (called the feasible set) in standard form as follows:

$$\mathcal{C} = \{ \mathbf{x} \in \mathcal{D} \mid f_i(\mathbf{x}) \le 0, h_j(\mathbf{x}) = 0, i = 1, \dots, m, j = 1, \dots, p \},\$$

where  $f_i(\mathbf{x})$  is convex for all *i* and  $h_j(\mathbf{x})$  is affine for all *j* and

$$\mathcal{D} = \operatorname{\mathbf{dom}} f \cap \left\{ \bigcap_{i=1}^m \operatorname{\mathbf{dom}} f_i \right\} \bigcap \left\{ \bigcap_{i=1}^p \operatorname{\mathbf{dom}} h_i \right\}$$

is called the problem domain.

#### • Advantages:

• *Global optimality*:  $x^*$  can be obtained by closed-form solution, analytically (algorithm), or numerically by convex solvers (e.g., CVX and SeDuMi).

- Computational complexity: Polynomial-time solvable.
- Performance analysis: KKT conditions are the backbone for analysis.

# Global optimality and solution

An optimality criterion: Any suboptimal solution to CVXP (8) is globally optimal. Assume that f is differentiable. Then a point x<sup>\*</sup> ∈ C is optimal if and only if

$$\nabla f(\mathbf{x}^{\star})^{T}(\mathbf{x} - \mathbf{x}^{\star}) \ge 0, \ \forall \mathbf{x} \in \mathcal{C}$$
(9)



(where int  $C \neq \emptyset$  is assumed)

# Global optimality and solution

- Besides the optimality criterion (9), a complementary approach for solving CVXP (8) is founded on the "duality theory".
  - Dual problem:

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) \triangleq f(\mathbf{x}) + \sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \quad \text{(Lagrangian)}$$

$$g(\boldsymbol{\lambda}, \boldsymbol{\nu}) = \inf_{\mathbf{x} \in \mathcal{D}} \mathcal{L}(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\nu}) > -\infty \quad \text{(dual function)}$$

$$d^* = \max \{ g(\boldsymbol{\lambda}, \boldsymbol{\nu}) \mid \boldsymbol{\lambda} \succeq \mathbf{0}, \boldsymbol{\nu} \in \mathbb{R}^p \} \quad \text{(dual problem)}$$

$$\leq p^* = \min \{ f(\mathbf{x}) \mid \mathbf{x} \in \mathcal{C} \} \quad \text{(primal problem CVXP (8))}$$
(10)

where  $\lambda = (\lambda_1, \ldots, \lambda_m)$  and  $\nu = (\nu_1, \ldots, \nu_p)$  are dual variables; " $\succeq$ " stands for an abbreviated generalized inequality defined by the proper cone  $K = \mathbb{R}^m_+$ , i.e., the first orthant, (a closed convex solid and pointed cone), i.e.,  $\lambda \succeq_K \mathbf{0} \Leftrightarrow \lambda \in K$ .

CVXP (8) and its dual can be solved simultaneously by solving the so-called *KKT conditions*.

#### • KKT conditions:

Suppose that  $f, f_1, \ldots, f_m, h_1, \ldots, h_p$  are differentiable and  $\mathbf{x}^*$  is primal optimal and  $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$  is dual optimal to CVXP (8). Under *strong duality*, i.e.,

$$p^{\star} = d^{\star} = \mathcal{L}(\mathbf{x}^{\star}, \boldsymbol{\lambda}^{\star}, \boldsymbol{\nu}^{\star})$$

(which holds true under *Slater's condition: a strictly feasible point exists, i.e.,* relint  $C \neq \emptyset$ ), the KKT conditions for solving  $\mathbf{x}^*$  and  $(\boldsymbol{\lambda}^*, \boldsymbol{\nu}^*)$  are as follows:

$$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^{\star}, \boldsymbol{\lambda}^{\star}, \boldsymbol{\nu}^{\star}) = \mathbf{0}, \tag{11a}$$

$$f_i(\mathbf{x}^*) \le 0, \ i = 1, \dots, m,$$
 (11b)

$$h_i(\mathbf{x}^{\star}) = 0, \ i = 1, \dots, p,$$
 (11c)

$$\lambda_i^* \ge 0, \ i = 1, \dots, m,\tag{11d}$$

 $\lambda_i^{\star} f_i(\mathbf{x}^{\star}) = 0, \ i = 1, \dots, m.$  (complementary slackness) (11e)

The above KKT conditions (11) and the optimality criterion (9) are equivalent under Slater's condition.



• Lagrangian  $\mathcal{L}(x, \lambda)$ , dual function  $g(\lambda)$ , and primal-dual optimal solution  $(x^*, \lambda^*) = (1, 1)$  of the convex problem  $\min\{f_0(x) = x^2 \mid (x - 2)^2 \le 1\}$  with strong duality. Note that  $f_0(x^*) = g(\lambda^*) = \mathcal{L}(x^*, \lambda^*) = 1$ .

• Consider the following convex optimization problem:

$$\min_{\mathbf{x} \in \mathbb{R}^{n}, \mathbf{z} \in \mathbb{R}^{m}} f_{1}(\mathbf{x}) + f_{2}(\mathbf{z})$$
s.t.  $\mathbf{x} \in S_{1}, \ \mathbf{z} \in S_{2}$ 

$$\mathbf{z} = \mathbf{A}\mathbf{x}$$
(18)

where  $f_1: \mathbb{R}^n \mapsto \mathbb{R}$  and  $f_2: \mathbb{R}^m \mapsto \mathbb{R}$  are convex functions,  $\mathbf{A}$  is an  $m \times n$  matrix, and  $\mathcal{S}_1 \subset \mathbb{R}^n$  and  $\mathcal{S}_2 \subset \mathbb{R}^m$  are nonempty convex sets.

• The considered dual problem of (18) is given by

$$\max_{\boldsymbol{\nu} \in \mathbb{R}^{m}} \min_{\mathbf{x} \in \mathcal{S}_{1}, \mathbf{z} \in \mathcal{S}_{2}} \left\{ f_{1}(\mathbf{x}) + f_{2}(\mathbf{z}) + \frac{c}{2} \| \mathbf{A}\mathbf{x} - \mathbf{z} \|_{2}^{2} + \boldsymbol{\nu}^{T} (\mathbf{A}\mathbf{x} - \mathbf{z}) \right\},$$
(19)

where c is a penalty parameter, and  $\nu$  is the dual variable associated with the equality constraint in (18).

# ADMM (Cont'd)

• Inner minimization (convex problems):

$$\mathbf{z}(q+1) = \arg\min_{\mathbf{z}\in\mathcal{S}_2} \left\{ f_2(\mathbf{z}) - \boldsymbol{\nu}(q)^T \mathbf{z} + \frac{c}{2} \|\mathbf{A}\mathbf{x}(q) - \mathbf{z}\|_2^2 \right\},$$
(20a)

$$\mathbf{x}(q+1) = \arg\min_{\mathbf{x}\in\mathcal{S}_1} \left\{ f_1(\mathbf{x}) + \boldsymbol{\nu}(q)^T \mathbf{A}\mathbf{x} + \frac{c}{2} \|\mathbf{A}\mathbf{x} - \mathbf{z}(q+1)\|_2^2 \right\}.$$
 (20b)

#### ADMM Algorithm

- 1: Set q = 0, choose c > 0.
- 2: Initialize  $\nu(q)$  and  $\mathbf{x}(q)$ .
- 3: repeat
- 4: Solve (20a) and (20b) for z(q + 1) and x(q + 1) by two distributed equipments including *the information exchange of* z(q + 1) *and* x(q + 1) *between them;*

5: 
$$\nu(q+1) = \nu(q) + c (\mathbf{Ax}(q+1) - \mathbf{z}(q+1));$$

6: 
$$q := q + 1;$$

- 7: until the predefined stopping criterion is satisfied.
- When  $S_1$  is bounded or  $\mathbf{A}^T \mathbf{A}$  is invertible, ADMM is guaranteed to converge and the obtained  $\{\mathbf{x}(q), \mathbf{z}(q)\}$  is an optimal solution of problem (18).

- **Reformulation into a convex problem:** Equivalent representations (e.g. epigraph representations); function transformation; change of variables, etc.
- Stationary-point solutions: Provided that C is closed and convex but f is nonconvex, a point x\* is a stationary point of the nonconvex problem (1) if

$$f'(\mathbf{x}^{\star}; \mathbf{v}) \triangleq \liminf_{\lambda \downarrow 0} \frac{f(\mathbf{x}^{\star} + \lambda \mathbf{v}) - f(\mathbf{x}^{\star})}{\lambda} \ge 0 \quad \forall \mathbf{x}^{\star} + \mathbf{v} \in \mathcal{C}$$
(21)

 $\Leftrightarrow \ \nabla f(\mathbf{x}^{\star})^{T}(\mathbf{x} - \mathbf{x}^{\star}) \geq 0 \quad \forall \mathbf{x} \in \mathcal{C} \ \text{(when } f \text{ is differentiable)}$ 

where  $f'(\mathbf{x}^*; \mathbf{v})$  is the *directional derivative* of f at a point  $\mathbf{x}^*$  in direction  $\mathbf{v}$ . Block successive upper bound minimization (BSUM) [Razaviyayn'13] guarantees a stationary-point solution under some convergence conditions.

• KKT points (i.e., solutions of KKT conditions) are also stationary points under some mild condition provided that the Slater's condition is satisfied.

<sup>[</sup>Razaviyayn'13] M. Razaviyayn, M. Hong, and Z.-Q. Luo, "A unified convergence analysis of block successive minimization methods for nonsmooth optimization," *SIAM J. Optimiz.*, vol. 23, no. 2, pp. 1126-1153, 2013.

# Stationary points and BSUM

• An illustration of stationary points of problem (1) for a nonconvex *f* and convex *C*; convergence to a stationary point of (1) by BSUM.



# Stationary points for nonconvex feasible set

• An illustration of stationary points of problem (1) when both f and C are nonconvex. If  $y_1, y_2, y_3$  are stationary points of  $\min_{\mathbf{x}\in C} f(\mathbf{x})$  where  $C \subset C$  is convex, then conic  $(C - \{y_i\}) = \{\theta \mathbf{v} \mid \mathbf{v} \in C - \{y_i\}, \theta \ge 0\}$  and

 $\mathcal{C} - \{\boldsymbol{y}_i\} \triangleq \{\mathbf{v} = \mathbf{x} - \{\boldsymbol{y}_i\} \mid \mathbf{x} \in \mathcal{C}\} \subset \operatorname{conic} (C - \{\boldsymbol{y}_i\}), \ i = 1, 2$  $\implies \boldsymbol{y}_1, \boldsymbol{y}_2 \text{ are also stationary points of (1).}$ 



## Nonconvex problem

- Approximate solutions to problem (1) when f is convex but C is nonconvex:
  - Convex restriction to C: Successive convex approximation (SCA)

$$\boldsymbol{x}_{i}^{\star} = \arg\min_{\boldsymbol{\mathbf{x}}\in C_{i}} f(\boldsymbol{\mathbf{x}}) \in C_{i+1}$$
(22)

where  $C_i \subset C$  is convex for all *i*. Then

$$f(\boldsymbol{x}_{i+1}^{\star}) = \min_{\boldsymbol{\mathbf{x}} \in C_{i+1}} f(\boldsymbol{\mathbf{x}}) \le f(\boldsymbol{x}_{i}^{\star})$$
(23)

After convergence, an approximate solution  $x_i^{\star}$  is obtained.

• Convex relaxation to C (e.g., semidefinite relaxation (SDR)):

$$C = \{ \mathbf{X} \in \mathbb{S}^{n}_{+} \mid \operatorname{rank}(\mathbf{X}) = 1 \} \text{ relaxed to conv } C = \mathbb{S}^{n}_{+} \text{ (SDR)}; \\ C = \{-3, -1, +1, +3\} \text{ relaxed to conv } C = [-3, 3]$$
(24)

The obtained  $X^*$  or  $x^*$  may not be feasible to problem (1); for SDR, a good approximate solution can be obtained from  $X^*$  via *Gaussian randomization*.

# Successive Convex Approximation (SCA)

Illustration of SCA for (1) when f is convex but C is nonconvex, yielding a solution x<sup>\*</sup><sub>i</sub> (which is a stationary point under some mild condition).



#### A new book

Convex Optimization for Signal Processing and Communications: From Fundamentals to Applications Chong-Yung Chi, Wei-Chiang Li, Chia-Hsiang Lin (Publisher: CRC Press, 2017, 432 pages, ISBN: 9781498776455)

**Motivation:** Most of mathematical books are hard to read for engineering students and professionals due to *lack of enough fundamental details and tangible linkage* between mathematical theory and applications.

- The book is written in a *causally sequential fashion*; namely, one can *review/peruse the related materials introduced in early chapters/sections again,* to overpass hurdles in reading.
- Covers convex optimization from fundamentals to advanced applications, while holding a strong link from theory to applications.
- Provides comprehensive proofs and perspective interpretations, many insightful figures, examples and remarks to illuminate core convex optimization theory.

#### **Book features**



- Illustrates, by cutting-edge applications, how to apply the convex optimization theory, like a guided journey/exploration rather than pure mathematics.
- Has been used for a 2-week short course under the book title at 9 major universities (Shandong Univ, Tsinghua Univ, Tianjin Univ, BJTU, Xiamen Univ., UESTC, SYSU, BUPT, SDNU) in Mainland China more than 17 times since early 2010.

### Thank you for your attention!

Acknowledgment: Financial support by NTHU; my students, visiting students and scholars, short-course participants for almost uncountable questions, interactions and comments over the last decade.

### **1** Part I: Fundamentals of Convex Optimization

### Part II: Application in Hyperspectral Image Analysis: (Big Data Analysis and Machine Learning)

- **3** Part III: Application in Wireless Communications (5G Systems)
  - Subsection I: Outage Constrained Robust Transmit Optimization for Multiuser MISO Downlinks
  - **Subsection II:** Outage Constrained Robust Hybrid Coordinated Beamforming for Massive MIMO Enabled Heterogeneous Cellular Networks

# Illustration of HU [Ambikapathi'11]

- Red pixel: a mixed pixel (land+vegetation+water)
- Blue pixel: a pure pixel (only water)



[Ambikapathi'11] A. Ambikapathi et al., "Chance constrained robust minimum volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4194-4209, Nov. 2011.

# Craig's HU criterion

- Observation:  $X \triangleq \{\mathbf{x}[1], \dots, \mathbf{x}[L]\} \subset \operatorname{conv}\{\mathbf{a}_1, \dots, \mathbf{a}_N\} \triangleq \mathcal{T}_a \subset \mathbb{R}^M$  (an *N*-vertex simplex due to (A1) and (A2));  $\operatorname{conv} X = \mathcal{T}_a$  if  $\mathbf{a}_i \in X \forall i$ .
- **Craig's belief:** The vertices of the minimum-volume data-enclosing simplex  $\hat{T}_a$  yield good estimate  $\hat{a}_i$  [Craig'94] even without any *pure pixels*,  $a_i$  in X.



Figure 2: Visually,  $vol(\mathcal{T}_a) < vol(\mathcal{T}_i)$ ,  $i \in \mathcal{I}_2$  (the dots are data points  $\mathbf{x}[n]$ ).

<sup>[</sup>Craig'94] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542-552, May 1994.

• Craig's criterion [Craig'94] (an NP-hard problem):

$$\{\widehat{\mathbf{a}}_{1}, \dots, \widehat{\mathbf{a}}_{N}\} \in \arg \min_{\mathbf{b}_{i} \in \mathbb{R}^{M} \forall i} \operatorname{vol}(\operatorname{conv}\{\mathbf{b}_{1}, \dots, \mathbf{b}_{N}\})$$
  
s.t.  $\mathbf{X} \subset \operatorname{conv}\{\mathbf{b}_{1}, \dots, \mathbf{b}_{N}\}$  (26)

• In [Lin'15], we theoretically proved that as long as the *data uniform purity level*  $\gamma$  is above a threshold (a mild condition), i.e.,

 $\gamma \triangleq \max\{r \mid \mathcal{T}_e \cap \mathcal{B}(r) \subseteq \mathsf{conv}\{\mathbf{s}[1], \dots, \mathbf{s}[L]\} > 1/\sqrt{N-1}$ 

where  $\mathcal{T}_e \triangleq \operatorname{conv} \{\mathbf{e}_1, \dots, \mathbf{e}_N\} \subseteq \mathbb{R}^N$  (unit simplex) and  $\mathcal{B}(r) \triangleq \{ \mathbf{x} \in \mathbb{R}^N \mid \|\mathbf{x}\| \leq r \}$ , Craig's criterion can perfectly identify the ground-truth endmembers  $\{\mathbf{a}_1, \dots, \mathbf{a}_N\}$  (i.e., the true simplex  $\mathcal{T}_a$ ).

• Can we devise a super-efficient HU algorithm using Craig's criterion?

<sup>[</sup>Craig'94] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 3, pp. 542-552, May 1994.

<sup>[</sup>Lin'15] C.-H. Lin et al., "Identifiability of the simplex volume minimization criterion for blind hyperspectral unmixing: The no pure-pixel case," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no.10, pp. 5530-5546, Oct. 2015.

# **Illustration of Dimension Reduction**



 Dimension reduction illustration using affine set fitting for N = 3, where the geometric center d of the data cloud X in the M-dimensional space maps to the origin in the (N - 1)-dimensional space.

# Dimension reduction and problem formulation

• Problem (26) can be reformulated in the DR space as:

$$\{\widehat{\boldsymbol{\alpha}}_{1},\ldots,\widehat{\boldsymbol{\alpha}}_{N}\} \in \arg\min_{\boldsymbol{\beta}_{i} \in \mathbb{R}^{N-1} \forall i} \left\{ \operatorname{vol}(\operatorname{conv}\{\boldsymbol{\beta}_{1},\ldots,\boldsymbol{\beta}_{N}\}) = \frac{|\operatorname{det}(\mathbf{B})|}{(N-1)!} \right\}$$
  
s.t.  $\mathcal{X} \subset \operatorname{conv}\{\boldsymbol{\beta}_{1},\ldots,\boldsymbol{\beta}_{N}\}$  (31)

where  $\mathbf{B} = [\boldsymbol{\beta}_1 - \boldsymbol{\beta}_N, \dots, \boldsymbol{\beta}_{N-1} - \boldsymbol{\beta}_N] \in \mathbb{R}^{(N-1) \times (N-1)}.$ 

• Endmember estimates in the original space  $\mathbb{R}^M$ :

$$\widehat{\mathbf{a}}_i = \mathbf{C}\widehat{\boldsymbol{\alpha}}_i + \mathbf{d}, \ \forall i \in \mathcal{I}_N \ \text{(cf. (30))}.$$

#### Existing Challenges

- Pure pixel assumption (PPA) enables various simple and fast blind HU algorithmic schemes (for finding the purest pixels in the data set X or the DR data set X), but it is often seriously infringed.
- Without requiring the PPA, Craig's blind HU criterion identifies the N-vertex minimum-volume data-enclosing simplex T̂<sub>a</sub> ⊂ ℝ<sup>M</sup>, but suffering from heavy simplex volume computations.

# Block diagram of HyperCSI algorithm

An algorithm with parallel processing structure for estimation of normal vectors (b<sub>i</sub>), inner product parameters (h<sub>i</sub>), and abundance maps (s<sub>i</sub>), where the PPA based successive projection algorithm (SPA) is employed to obtain initial estimates α̃<sub>1</sub>,..., α̃<sub>N</sub>.



# Graphical illustration of HyperCSI



• Why  $\mathcal{R}_k^{(i)}$  should be disjoint? Consider  $\{\mathbf{p}_2^{(1)}, \mathbf{q}\}$  identified by (34).

- Why not b<sub>1</sub>? The purest pixel α<sub>3</sub> may not be close to H<sub>1</sub> = aff{α<sub>2</sub>, α<sub>3</sub>}, leading to nontrivial orientation difference between b<sub>1</sub> and b<sub>1</sub>.
- However,  $\{\mathbf{p}_1^{(1)}, \mathbf{p}_2^{(1)}\}$  identified by (35) are very close to  $\mathcal{H}_1$ , so the orientations of  $\hat{\mathbf{b}}_1$  and  $\mathbf{b}_1$  are almost the same.

# Numerical simulaitons (con't)

• Data generation: N = 6 endmembers with M = 224 spectral bands are randomly selected from the US Geological Survey (USGS) library to generate L = 10,000 noiseless synthetic data, and then added by Gaussian noise.

#### • Performance measures:

- (1) Computation time T;
- 2 Root-mean-square (RMS) spectral angle error  $\phi_{en}$  (between  $\mathbf{a}_i$  and  $\widehat{\mathbf{a}}_i$ ):

$$\phi_{en} = \min_{\boldsymbol{\pi} \in \Pi_N} \sqrt{\frac{1}{N} \sum_{i=1}^{N} \left[ \arccos\left(\frac{\mathbf{a}_i^T \hat{\mathbf{a}}_{\pi_i}}{\|\mathbf{a}_i\| \cdot \|\hat{\mathbf{a}}_{\pi_i}\|} \right) \right]^2},$$

where  $\Pi_N$  is the set of all the permutations of  $\{1, \ldots, N\}$ . 3 RMS angle error  $\phi_{ab}$  (between  $s_i$  and  $\hat{s}_i$ ):

$$\phi_{ab} = \min_{\boldsymbol{\pi} \in \Pi_N} \sqrt{\frac{1}{N} \sum_{i=1}^N \left[ \arccos\left(\frac{\boldsymbol{s}_i^T \hat{\boldsymbol{s}}_{\pi_i}}{\|\boldsymbol{s}_i\| \cdot \| \hat{\boldsymbol{s}}_{\pi_i} \|} \right) \right]^2}$$

# Numerical simulaitons (con't)



(b) Ground truth abundance maps of SYN2

Two sets of sparsely, non-i.i.d. and non-Dirichlet distributed maps are used to generate two synthetic datasets [lordache'12], denoted by SYN1 and SYN2, for performance evaluation, where SYN1 contains L = 10,000 pixels and SYN2 contains L = 16,900pixels, resp. [lordache'12].

<sup>[</sup>lordache'12] M.-D. lordache et al., "Total variation spatial regularization for sparse hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 11, pp. 4484-4502, 2012.

# Numerical simulaitons (con't)

# • Each simulation result for different SNRs, is obtained by averaging over 100 realizations.

|      |          | $\phi_{en}$ (degrees) |          |      | $\phi_{ab}$ (degrees) |          |       |       |       |             |       |         |
|------|----------|-----------------------|----------|------|-----------------------|----------|-------|-------|-------|-------------|-------|---------|
|      | Methods  |                       | SNR (dB) |      |                       | SNR (dB) |       |       |       | T (seconds) |       |         |
|      |          | 20                    | 25       | 30   | 35                    | 40       | 20    | 25    | 30    | 35          | 40    |         |
| SYN1 | MVC-NMF  | 3.23                  | 1.97     | 1.05 | 0.55                  | 0.25     | 13.87 | 8.51  | 4.79  | 2.65        | 1.34  | 1.74E+2 |
|      | MVSA     | 10.65                 | 6.12     | 3.38 | 1.88                  | 1.05     | 22.93 | 15.13 | 9.34  | 5.52        | 3.19  | 3.53E+0 |
|      | MVES     | 9.55                  | 5.49     | 3.60 | 1.96                  | 1.22     | 23.89 | 17.35 | 14.49 | 7.78        | 5.66  | 3.42E+1 |
|      | SISAL    | 4.43                  | 2.89     | 1.81 | 1.18                  | 0.86     | 15.85 | 10.39 | 6.89  | 5.29        | 4.65  | 2.66E+0 |
|      | ipMVSA   | 11.62                 | 6.82     | 3.38 | 2.01                  | 1.05     | 24.05 | 16.28 | 9.34  | 5.98        | 3.19  | 1.65E+0 |
|      | HyperCSI | 1.55                  | 1.22     | 0.79 | 0.52                  | 0.35     | 12.03 | 6.92  | 4.16  | 2.49        | 1.46  | 5.56E-2 |
| SYN2 | MVC-NMF  | 2.86                  | 1.71     | 0.97 | 0.54                  | 0.23     | 22.86 | 15.52 | 9.39  | 5.27        | 2.67  | 2.48E+2 |
|      | MVSA     | 10.21                 | 5.55     | 3.08 | 1.71                  | 0.95     | 29.86 | 22.72 | 15.57 | 9.78        | 5.83  | 5.65E+0 |
|      | MVES     | 10.12                 | 5.19     | 3.15 | 2.04                  | 3.77     | 29.43 | 22.13 | 15.66 | 10.42       | 13.17 | 2.22E+1 |
|      | SISAL    | 3.25                  | 2.18     | 1.48 | 0.96                  | 0.63     | 24.79 | 17.49 | 11.51 | 7.00        | 4.21  | 4.45E+0 |
|      | ipMVSA   | 11.34                 | 8.26     | 3.34 | 1.94                  | 1.01     | 30.23 | 30.38 | 16.29 | 10.30       | 6.39  | 8.14E-1 |
|      | HyperCSI | 1.48                  | 1.08     | 0.71 | 0.44                  | 0.31     | 22.64 | 15.98 | 11.10 | 7.25        | 4.40  | 7.48E-2 |

## **Real data experiments**

- Real hyperspectral imaging data experiments: Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) taken over the Cuprite mining site, Nevada, in 1997 [AVIRIS'97].
- The number of sources for this dataset is estimated to be N = 9 using an information-theoretic minimum description length (MDL) criterion [Lin'16-2].
- The proposed HyperCSI algorithm, along with the following two benchmark algorithms (for analyzing the hyperspectral imaging data), are used to process the AVIRIS data:
  - MVC-NMF algorithm [Miao'07] (based on Craig's criterion);
  - **VCA** algorithm [Nascimento'05] (based on pure-pixel assumption).

<sup>[</sup>AVIRIS'97] AVIRIS Free Standard Data Products. [Online]. Available: http://aviris.jpl.nasa.gov/html/aviris.freedata.html [Lin'16-2] C.-H. Lin et al., "Detection of sources in non-negative blind source separation by minimum description length criterion," *IEEE Trans. Neural Networks and Learning Systems*, vol. 29, no. 9, pp. 4022-4037, Sept. 2018. [Nascimento'05] J. Nascimento et al., "Vertex component analysis: A fast algorithm to unmix hyperspectral data," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 4, pp. 898-910, Apr. 2005.

# Real data experiments (con't)



Endmembers extracted by HyperCSI algorithm show better resemblance to their counterparts in library. For instance, the endmember of Alunite extracted by HyperCSI shows much clearer absorption feature than MVC-NMF and VCA, in the bands approximately from 2.3 to 2.5  $\mu$ m.

# Real data experiments (con't)

 The average RMS spectral angle error φ between endmember estimates and their corresponding library spectra, is used for quantitative comparison:

|                          | HyperCSI | MVC-NMF | VCA           |
|--------------------------|----------|---------|---------------|
| Muscovite                | 3.03     | 3.96    | 4.54          |
| Alunite                  | 7.48     | 6.23    | 6.57          |
| Desert Varnish           | 9.49     | 4.91    | 7.92          |
| Hematite                 | 7.83     | 12.94   | 7.24          |
| Montmorillonite          | 4.84     | 7.44    | 6.59          |
| Kaolinite #1             | 8.63     | 7.56    | 13.80 (11.71) |
| Kaolinite #2             | 7.39     | -       | -             |
| Buddingtonite            | 6.55     | 8.16    | 6.46          |
| Chalcedony               | 5.92     | 7.97    | 8.25          |
| Andradite                | -        | 7.43    | -             |
| Average $\phi$ (degrees) | 6.80     | 7.40    | 8.12          |
| T (seconds)              | 0.12     | 988.67  | 5.40          |

- As the pure pixels may not be present in the selected subscene, the two Craig criterion based algorithms outperform VCA as expected.
- In terms of the computation time *T*, in spite of parallel processing not yet applied, HyperCSI is much faster than the other two algorithms.

# Real data experiments (con't)









Desert varnish



Muscovite

Hematite



Montmorillonite



Kaolinite #1



Kaolinite #2



Buddingtonite



Chalcedony

The good performance of HyperCSI in the experiment also implies that the requirement of sufficient (i.e., N(N-1) = 72) active pixels lying close to the hyperplanes of the actual endmembers' simplex, has been met for the considered hyperspectral scene.

Based on the hyperplane representation for a simplest simplex, the presented HyperCSI algorithm has the following remarkable characteristics:

- **O** Craig's simplex is *reconstructed from* N(N-1) *pixels* (regardless of the data length L), without involving any simplex volume computations.
- It is *reproducible* (without involving random initialization and tuning of regularization parameters) and not data-dependent, regardless of the existence of pure pixels.
- Its superior performance over state-of-the-art methods has been demonstrated by analysis, simulations and real data experiments.
- It only involves simple linear algebraic computations, with a complexity  $\mathcal{O}(NL)$  with or  $\mathcal{O}(N^2L)$  without parallel implementation, thereby sustaining its practical applicability.

## **1** Part I: Fundamentals of Convex Optimization

Part II: Application in Hyperspectral Image Analysis: (Big Data Analysis and Machine Learning)

# **3** Part III: Application in Wireless Communications (5G Systems)

- Subsection I: Outage Constrained Robust Transmit Optimization for Multiuser MISO Downlinks
- Subsection II: Outage Constrained Robust Hybrid Coordinated Beamforming for Massive MIMO Enabled Heterogeneous Cellular Networks

• Multiuser multiple-input single-output (MISO) downlink:

A practical scenario in wireless communications where one base station (BS) equipped with  $N_t$  antennas sends independent messages to K single-antenna users.



• Transmit signal from BS:

$$\boldsymbol{x}(t) = \sum_{i=1}^{K} \boldsymbol{x}_i(t).$$
(43)

- $\boldsymbol{x}_i(t) \in \mathbb{C}^{N_t}$ : information signal for user i;  $\boldsymbol{x}_i(t) \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{S}_i)$  with  $\boldsymbol{S}_i \succeq \mathbf{0}$  denoting the signal covariance matrix.
- Assuming that  $\operatorname{rank}(\boldsymbol{S}_i) = d$ ,  $\boldsymbol{x}_i(t)$  can be expressed as [Vu07]

$$\boldsymbol{x}_{i}(t) = \sum_{k=1}^{d} \sqrt{\lambda_{k}(\boldsymbol{S}_{i})} \boldsymbol{w}_{k} s_{ik}(t).$$
(44)

- $\lambda_k(S_i)$ : the kth largest eigenvalue of  $S_i$ ;
- $s_{ik}(t) \sim C\mathcal{N}(0,1)$ : kth independent data stream for user i;
- $\boldsymbol{w}_k \in \mathbb{C}^{N_t}$ : orthonormal eigenvectors of  $\boldsymbol{S}_i$ .

# • When d = 1, the transmit strategy for $x_i(t)$ reduces to transmit beamforming.

<sup>[</sup>Vu07] M. Vu and A. Paulraj, "MIMO wireless linear precoding," *IEEE Signal Process. Magazine*, vol. 24, no. 5, pp. 86–105, Sep. 2007.

• Received signal of user *i*:

$$y_i(t) = \boldsymbol{h}_i^H \boldsymbol{x}(t) + n_i(t).$$
(45)

- $\boldsymbol{h}_i \in \mathbb{C}^{N_t}$ : the channel of user *i*;
- $n_i(t) \sim \mathcal{CN}(0, \sigma_i^2)$ : additive noise at the user i.
- Achievable rate of user i (in bits/sec/Hz), assuming single-user detection with perfect h<sub>i</sub> at the receiver i [Telatar99]:

$$\mathsf{R}_{i}(\{\boldsymbol{S}_{k}\}_{k=1}^{K};\boldsymbol{h}_{i}) = \log_{2}\left(1 + \underbrace{\frac{\boldsymbol{h}_{i}^{H}\boldsymbol{S}_{i}\boldsymbol{h}_{i}}{\sum_{k\neq i}^{K}\boldsymbol{h}_{i}^{H}\boldsymbol{S}_{k}\boldsymbol{h}_{i} + \sigma_{i}^{2}}}_{\mathsf{SINR}}\right), \ i = 1, \dots, K,$$
(46)

where SINR denotes the signal-to-interference-plus-noise ratio associated with user i.

<sup>[</sup>Telatar99] E. Telatar, "Capacity of multi-antenna Gaussian channels," Bell Labs Tech. J., vol. 10, no. 6, pp. 585-595, Nov./Dec. 1999.

#### Rate Constrained Problem under Perfect Channel State Information (CSI)

With the given CSI  $h_1, \ldots, h_K$  that are known to the BS,

$$\min_{\boldsymbol{S}_1,\dots,\boldsymbol{S}_K \in \mathbb{H}^{N_t}} \sum_{i=1}^K \operatorname{Tr}(\boldsymbol{S}_i)$$
(47a)

s.t. 
$$\mathsf{R}_i(\{\boldsymbol{S}_k\}_{k=1}^K; \boldsymbol{h}_i) \geq \boldsymbol{r}_i, \quad i = 1, \dots, K,$$
 (47b)  
 $\boldsymbol{S}_1, \dots, \boldsymbol{S}_K \succeq \boldsymbol{0},$  (47c)

where each  $r_i \ge 0$  is the required information rate (target rate) for user *i*.

 Problem (47) can be reformulated as a convex semidefinite program (SDP), which is polynomial-time solvable [Bengtsson01] [Gershman10].

<sup>[</sup>Bengtsson01] M. Bengtsson and B. Ottersten, "Handbook of Antennas in Wireless Communications," L. C. Godara, Ed., CRC Press, Aug. 2001.

<sup>[</sup>Gershman10] A. B. Gershman and N. D. Sidiropoulos and S. Shahbazpanahi and M. Bengtsson and B. Ottersten, "Convex optimization-based beamforming," *IEEE Signal Process. Mag.*, vol. 27, no. 3, pp. 62-75, May 2010.

- Unfortunately, the BS cannot acquire perfect CSI *h<sub>i</sub>* (used in the conventional formulation in (47)) due to imperfect channel estimation and limited feedback [Love08].
- CSI error model:

$$\boldsymbol{h}_i = \bar{\boldsymbol{h}}_i + \boldsymbol{e}_i, \qquad i = 1, \dots, K, \tag{48}$$

where  $\bar{h}_i \in \mathbb{C}^{N_t}$  is the presumed channel at the BS, and  $e_i \in \mathbb{C}^{N_t}$  is the channel error vector.

 Gaussian channel error model [Marco05] [Shenouda08] (suitable for imperfect channel estimation at BS):

$$\boldsymbol{e}_i \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{C}_i) \tag{49}$$

for some known error covariance  $C_i \succeq 0$ .

<sup>[</sup>Love08] D. J. Love, R. Heath, V. K. N. Lau, D. Gesbert, B. Rao, and M. Andrews, "An overview of limited feedback in wireless communication systems," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 8, pp. 1341-1365, Oct. 2008.

<sup>[</sup>Marco05] D. Marco and D. L. Neuhoff, "The validity of the additive noise model for uniform scalar quantizers," IEEE Trans. Inform. Theory, vol. 51, no. 5, pp. 1739-1755, May 2005.

<sup>[</sup>Shenouda08] M. B. Shenouda and T. N. Davidson, "Probabilistically-constrained approaches to the design of the multiple antenna downlink," in *Proc. 42nd Asilomar Conference 2008*, Pacific Grove, October 26-29, 2008, pp. 1120-1124.

#### **Rate Outage Constrained Problem**

Given rate requirements  $r_1, \ldots, r_K > 0$  and maximum tolerable outage probabilities  $\rho_1, \ldots, \rho_K \in (0, 1]$ ,

$$\min_{\boldsymbol{S}_{1},...,\boldsymbol{S}_{K} \in \mathbb{H}^{N_{t}}} \sum_{i=1}^{K} \operatorname{Tr}(\boldsymbol{S}_{i})$$
(50a)  
s.t. Prob  $\left\{ \mathsf{R}_{i}(\{\boldsymbol{S}_{k}\}_{k=1}^{K}; \bar{\boldsymbol{h}}_{i} + \boldsymbol{e}_{i}) \leq r_{i} \right\} \leq \rho_{i}, \ i = 1, \dots, K,$ (50b)  
 $\boldsymbol{S}_{1}, \dots, \boldsymbol{S}_{K} \succeq \boldsymbol{0},$ (50c)  
where  $\mathsf{R}_{i}(\{\boldsymbol{S}_{k}\}_{k=1}^{K}, \bar{\boldsymbol{h}}_{i} + \boldsymbol{e}_{i})$  is defined in (46).

 Problem (50) is hard to solve since rate outage probabilities in (50b) have no closed-form expressions and are unlikely to be efficiently computable in general.

#### 2. A Restriction Approach for Problem (50)

• The rate outage constraints in (50b) can be expressed as

 $Prob\{e_{i}^{H}Q_{i}e_{i}+2Re\{e_{i}^{H}r_{i}\}+s_{i}<0\}\leq\rho_{i}, \quad i=1,\ldots,K,$ (51)

where for notational simplicity,  $e_i \sim C\mathcal{N}(\mathbf{0}, I_{N_t})$  (originally denoting channel error), and

$$\boldsymbol{Q}_{i} = \boldsymbol{C}_{i}^{1/2} \left( \frac{1}{\gamma_{i}} \boldsymbol{S}_{i} - \sum_{k \neq i} \boldsymbol{S}_{k} \right) \boldsymbol{C}_{i}^{1/2}, \quad \boldsymbol{r}_{i} = \boldsymbol{C}_{i}^{1/2} \left( \frac{1}{\gamma_{i}} \boldsymbol{S}_{i} - \sum_{k \neq i} \boldsymbol{S}_{k} \right) \bar{\boldsymbol{h}}_{i},$$
(52a)

$$s_{i} = \bar{\boldsymbol{h}}_{i}^{H} \left( \frac{1}{\gamma_{i}} \boldsymbol{S}_{i} - \sum_{k \neq i} \boldsymbol{S}_{k} \right) \bar{\boldsymbol{h}}_{i} - \sigma_{i}^{2},$$
(52b)

in which

 $\gamma_i = 2^{r_i} - 1$  (target SINR for user *i* (cf. (46)))

corresponding to the rate requirement  $r_i = \log_2(1 + \gamma_i)$ .

### 2. Convex Restriction Methods (Bernstein-Type Inequality)

• Alternative expression for the outage probability constraint in (55):

 $\operatorname{Prob}\{\boldsymbol{e}^{H}\boldsymbol{Q}\boldsymbol{e}+2\operatorname{Re}\{\boldsymbol{e}^{H}\boldsymbol{r}\}+s\geq0\}\geq1-\rho.$ 

Lemma 2 (Bernstein-Type Inequality) [Bechar09]

Let  $e \sim \mathcal{CN}(\mathbf{0}, I_n)$ , and let  $Q \in \mathbb{H}^n$  and  $r \in \mathbb{C}^n$  be given. Then, for any  $\eta > 0$ ,

$$\operatorname{Prob}\left\{\boldsymbol{e}^{H}\boldsymbol{Q}\boldsymbol{e}+2\operatorname{Re}\left\{\boldsymbol{e}^{H}\boldsymbol{r}\right\}\geq\boldsymbol{\Upsilon}(\boldsymbol{\eta})\right\}\geq1-e^{-\boldsymbol{\eta}},$$
(56)

where  $\Upsilon:\mathbb{R}_{++}\rightarrow\mathbb{R}$  is defined by

$$\Upsilon(\boldsymbol{\eta}) = \operatorname{Tr}(\boldsymbol{Q}) - \sqrt{2\eta} \sqrt{\|\boldsymbol{Q}\|_F^2 + 2\|\boldsymbol{r}\|_2^2} - \eta \lambda^+(\boldsymbol{Q}),$$

and  $\lambda^+(\boldsymbol{Q}) = \max\{\lambda_{\max}(-\boldsymbol{Q}), 0\}.$ 

<sup>[</sup>Bechar09] I. Bechar, "A Bernstein-type inequality for stochastic processes of quadratic forms of Gaussian variables," 2009, preprint, available on http://arxiv.org/abs/ 0909.3595.

#### Method II (Bernstein-Type Inequality)

A convex restriction approximation of problem (50):

$$\min_{\substack{\boldsymbol{S}_{i} \in \mathbb{H}^{N_{t}}, x_{i}, y_{i} \in \mathbb{R}, \\ i=1,\dots,K}} \sum_{i=1}^{K} \operatorname{Tr}(\boldsymbol{S}_{i}) \tag{58a}$$
s.t.  $\operatorname{Tr}(\boldsymbol{Q}_{i}) - \sqrt{2 \ln(1/\rho_{i})} \cdot x_{i} + \ln(\rho_{i}) \cdot y_{i} + s_{i} \ge 0, \forall i, \tag{58b}$ 

$$\left\| \begin{bmatrix} \operatorname{vec}(\boldsymbol{Q}_{i}) \\ \sqrt{2}\boldsymbol{r}_{i} \end{bmatrix} \right\|_{2} \le x_{i}, \ i = 1,\dots,K, \tag{58c}$$

$$y_i \boldsymbol{I}_{N_t} + \boldsymbol{Q}_i \succeq \boldsymbol{0}, \ i = 1, \dots, K,$$
 (58d)

$$y_1, \ldots, y_K \ge 0, \ \boldsymbol{S}_1, \ldots, \boldsymbol{S}_K \succeq \boldsymbol{0},$$
 (58e)

where  $Q_i$ ,  $r_i$  and  $s_i$  are defined in (52),  $i = 1, \ldots, K$ .

Simulation Setting (spatially i.i.d. Gaussian CSI errors):

- $N_t = K = 3;$
- Users' noise powers:  $\sigma_1^2 = \cdots = \sigma_K^2 = 0.1;$
- Preset outage probabilities:  $\rho_1 = \cdots = \rho_K = 0.1$ ;
- SINR requirements:  $\gamma_1 = \cdots = \gamma_K \triangleq \gamma$  (recall that  $\gamma_i = 2^{r_i} 1$ );
- Spatially i.i.d. Gaussian CSI errors  $C_1 = \cdots = C_K = 0.002 I_{N_t}$ ;
- In each simulation, 500 sets of the presumed channels  $\{\bar{h}_i\}_{i=1}^K$  are randomly and independently generated with  $\bar{h}_i \sim C\mathcal{N}(\mathbf{0}, I_{N_i})$ ;
- Performance comparisons with probabilistic SOCP [Shenouda08].

<sup>[</sup>Shenouda08] M. B. Shenouda and T. N. Davidson, "Probabilistically-constrained approaches to the design of the multiple antenna downlink," in *Proc. 42nd Asilomar Conference 2008*, Pacific Grove, October 26-29, 2008, pp. 1120-1124.

### 3. Simulation Results

Transmit power performance of the various methods.



### 3. Simulation Results

- In the simulation, a "rank-1 solution" for  $(\widehat{S}_1, \ldots, \widehat{S}_K)$  is obtained if  $\frac{\lambda_{\max}(\widehat{S}_i)}{\operatorname{Tr}(\widehat{S}_i)} \geq 0.9999$  for all i.
- Ratio of rank-one solution  $\triangleq \frac{\text{no. of realizations yielding a rank-one solution}}{\text{no. of realizations yielding a feasible solution}}$ .

| ρ             | 0.1     |         |         |         |  |  |  |
|---------------|---------|---------|---------|---------|--|--|--|
| $\gamma$ (dB) | 3       | 7       | 11      | 15      |  |  |  |
| Method I      | 464/464 | 448/448 | 404/404 | 292/292 |  |  |  |
| Method II     | 489/489 | 475/475 | 441/441 | 363/363 |  |  |  |
| Method III    | 488/488 | 449/449 | 372/372 | 251/251 |  |  |  |

Table 1: Ratios of rank-one solutions.

| ρ             | 0.01    |         |         |         |  |  |  |
|---------------|---------|---------|---------|---------|--|--|--|
| $\gamma$ (dB) | 3       | 7       | 11      | 15      |  |  |  |
| Method I      | 450/450 | 424/424 | 343/343 | 225/225 |  |  |  |
| Method II     | 477/480 | 463/463 | 428/428 | 322/322 |  |  |  |
| Method III    | 473/473 | 418/418 | 301/301 | 124/124 |  |  |  |

### 4. Conclusions

- We considered the multiuser MISO downlink scenario with Gaussian CSI errors and studied a rate outage constrained optimization problem.
- Bernstein-type inequality based method for efficiently computable convex restriction of the probabilistic constraints using analytic tools from probability theory was presented [Wang'14].
- Simulation results demonstrated that Bernstein-type inequality based method significantly improve upon the existing state-of-the-art method [Shenouda08] in terms of both computational complexity and solution accuracy.

<sup>[</sup>Wang'14] K.-Y. Wang, A. M.-C. So, T.-H. Chang, W.-K. Ma, and Chong-Yung Chi, "Outage constrained robust transmit optimization for multiuser MISO downlinks: Tractable approximations by conic optimization," IEEE Trans. Signal Processing, vol. 62, no. 21, pp. 5690-5705, Nov. 2014. (*Citations:* 166 by Google Scholar)

## **1** Part I: Fundamentals of Convex Optimization

- Part II: Application in Hyperspectral Image Analysis: (Big Data Analysis and Machine Learning)
- Part III: Application in Wireless Communications (5G Systems)
  - Subsection I: Outage Constrained Robust Transmit Optimization for Multiuser MISO Downlinks
  - Subsection II: Outage Constrained Robust Hybrid Coordinated Beamforming for Massive MIMO Enabled Heterogeneous Cellular Networks

#### 1. System Model and Problem Formulation

• Massive MIMO enabled two-tier heterogeneous network (HetNet): A macrocell base station (MBS) equipped with large-scale  $N_{\rm MBS}$  antennas, and a femtocell base station (FBS) equipped with  $N_{\rm FBS}$  antennas, serve K single-antenna macrocell user equipments (MUEs) and Jsingle-antenna femtocell user equipments (FUEs), respectively.



#### 1. System Model and Problem Formulation

• Hybrid coordinated beamforming (HyCoBF) structure at MBS for the HetNet, with  $N_{\rm RF}$  radio frequency (RF) chains satisfying  $N_{\rm MBS} \gg N_{\rm RF} \ge K$ , and  $N_{\rm RF} \times N_{\rm MBS}$  analog phase shifters [Molisch'16].



<sup>[</sup>Molisch'16] A. F. Molisch, V. V. Ratnam, S. Han, Z. Li, S. Nguyen, S. Li, and K. Haneda, "Hybrid beamforming for massive MIMO-A survey," http://arxiv.org/pdf/1609.05078v1.pdf, Sep. 2016.

### 3. Simulation Results

#### Simulation Setting:

- Users' AWGN powers:  $\sigma_k^2 = \sigma_j^2 = \sigma^2, \ \forall k \in \mathcal{I}_K, \ \forall j \in \mathcal{I}_J;$
- Target SINRs:  $\gamma_{Mk} = \gamma_{Fj} = \gamma, \forall k, j;$
- SINR outage probabilities:  $\rho_{Mk} = \rho_{Fj} = \rho$ ;
- CSI error covariance matrices:  $\mathbf{C}_{MMk} = \mathbf{C}_{MFj} = \varepsilon^2 \mathbf{I}_{N_{\text{MBS}}}$ ,  $\mathbf{C}_{FFj} = \mathbf{C}_{FMk} = \varepsilon^2 \mathbf{I}_{N_{\text{FBS}}}$ ;
- The performance evaluations were performed using CVX for the proposed HyCoBF design.
- The digital CoBF of the proposed robust HyCoBF design reduces to full digital (FD) CoBF as  $N_{\rm MBS} = N_{\rm RF}$ .

#### 3. Simulation Results (Power Performance)

•  $N_{\text{MBS}} = 16$ , K = 4,  $N_{\text{FBS}} = J = 2$ ,  $N_{\text{RF}} = \{4, 8, 16\}$ ;  $\rho = 0.1$ ,  $\varepsilon^2 = \{0.001, 0.002\}$ .



#### 3. Simulation Results (Power Performance)

•  $N_{\text{MBS}} = 64, K = 4, N_{\text{FBS}} \in \{2, 4\}, N_{\text{RF}} = \{4, 8, 16\}, J = 2; \rho = 0.1, \varepsilon^2 = 0.002.$ 



 $\gamma(dB)$ 

### 3. Simulation Results (Rank-one Solution)

 In simulation, a "rank-one solution" for W<sup>\*</sup><sub>k</sub> and U<sup>\*</sup><sub>j</sub> is obtained if the following conditions hold:

$$\frac{\lambda_{\max}(\mathbf{w}_{k}^{\star})}{\operatorname{Tr}(\mathbf{w}_{k}^{\star})} \geq 0.9999, \frac{\lambda_{\max}(\mathbf{U}_{j}^{\star})}{\operatorname{Tr}(\mathbf{U}_{j}^{\star})} \geq 0.9999, k \in \mathcal{I}_{\mathcal{K}}, j \in \mathcal{I}_{\mathcal{J}}.$$
(78)

• Counts of rank-one solutions and all the feasible solutions for each simulation case for  $\varepsilon^2 \in \{0.01, 0.002\}$  and  $\gamma \in \{1, 5, 9, 13\}$  dB.

| $\varepsilon^2$ | 0.01       |            |            |            |  |
|-----------------|------------|------------|------------|------------|--|
| $\gamma$ (dB)   | 1          | 5          | 9          | 13         |  |
| Robust FD       | (447, 447) | (396, 396) | (286, 286) | (102, 102) |  |
| HyCoBF-PRM      | (418, 418) | (250, 250) | (63, 63)   | (35, 35)   |  |

| $\varepsilon^2$ | 0.002      |            |            |            |  |  |
|-----------------|------------|------------|------------|------------|--|--|
| $\gamma$ (dB)   | 1 5        |            | 9          | 13         |  |  |
| Robust FD       | (490, 490) | (481, 481) | (460, 460) | (401, 401) |  |  |
| HyCoBF-PRM      | (413, 413) | (402, 402) | (336, 336) | (191, 191) |  |  |

- Simulation results have demonstrated that the proposed robust HyCoBF design algorithm can yield promising performance and, most importantly, it can achieve comparable performance to the FD beamforming scheme with much fewer RF chains.
- Recently, a distributed implementation for the CoBF solution using ADMM in the proposed HyCoBF has been finished [Xu'17].

<sup>[</sup>Xu'17] G.-X. Xu, C.-H. Lin, W.-G. Ma, S.-Z. Chen, and Chong-Yung Chi, "Outage constrained robust hybrid coordinated beamforming for massive MIMO enabled heterogeneous cellular networks," *IEEE Access*, vol. 5, pp. 13601-13616, Mar. 2017.