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 ITCOM Summer School, NTHU, Taiwan, 2017/8/1-4

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.

Outline

1. **Part I: Fundamentals of Convex Optimization**
2. **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

- **Optimization problem:**

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

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 $\mathbf{x}$ involving real and/or complex vectors, matrices; feasible set $\mathcal{C}$ 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,

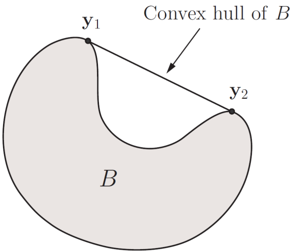
$$\mathbf{x} = \theta_1 \mathbf{x}_1 + \cdots + \theta_K \mathbf{x}_K, \ \mathbf{x}_i \in C \ \forall i \tag{3}$$

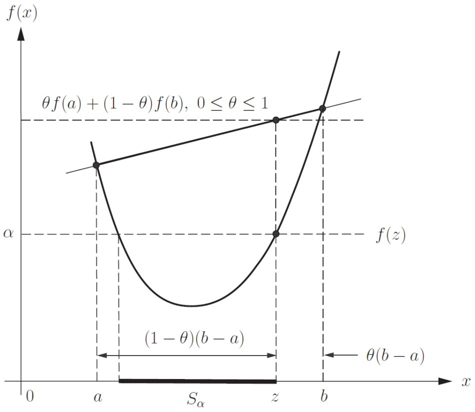
  is called an *affine (a convex) combination* of $\mathbf{x}_1, \ldots, \mathbf{x}_K$ (K vectors or points of a set) if $\sum_{i=1}^{K} \theta_i = 1$, $\theta_i \in \mathbb{R}$ ($\theta_i \in \mathbb{R}_+$), $K \in \mathbb{Z}_+$.

- **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 $\theta \geq 0$ and any $\mathbf{x} \in C$, then C is a cone; a cone is constructed by *rays starting from the origin*;
  • the linear combination (3) is called a *conic combination* if $\theta_i \geq 0 \ \forall i$;
  • C is a *convex cone* if C is closed under the operation of *conic combination*.

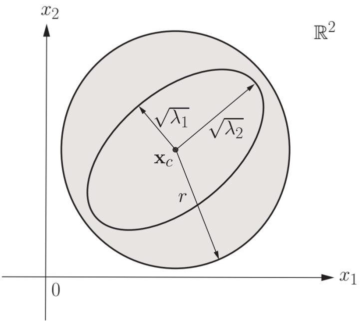
Convex sets and convex functions-2

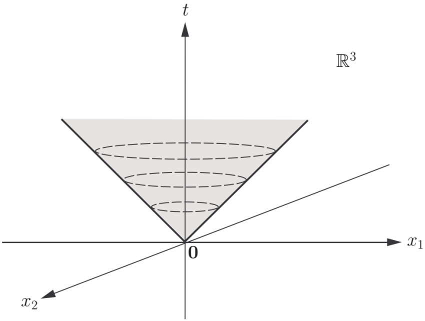




- Left plot: $\frac{\mathbf{y}_1+\mathbf{y}_2}{2} \notin B$, implying that B is not a convex set; right plot: $f(x)$ is a convex function (by (4)).

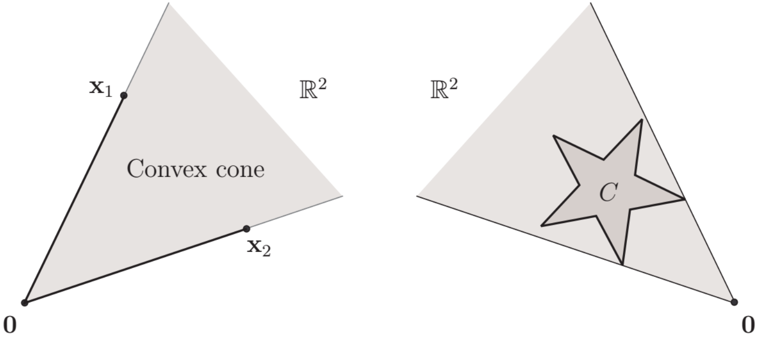
Convex Set Examples





- Left plot: An ellipsoid centered at $\mathbf{x}_c$ with semiaxes $\sqrt{\lambda_1}, \sqrt{\lambda_2}$, and an Euclidean ball centered at $\mathbf{x}_c$ with radius $r > \max\{\sqrt{\lambda_1}, \sqrt{\lambda_2}\}$ in $\mathbb{R}^2$; right plot: Second-order cone $C = \left\{(\mathbf{x}, t) \in \mathbb{R}^{n+1} \mid \|\mathbf{x}\|_2 \leq t\right\}$ in $\mathbb{R}^3$.

Convex Set Examples



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

Convex sets and convex functions-3

- Let $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_N\} \subset \mathbb{R}^M$. Affine hull of $\mathcal{A}$ *(the smallest affine set containing $\mathcal{A}$)* is defined as

$$\begin{aligned} \text{aff}\, \mathcal{A} &= \left\{ \mathbf{x} = \sum_{i=1}^{N} \theta_i \mathbf{a}_i \mid \sum_{i=1}^{N} \theta_i = 1,\ \theta_i \in \mathbb{R}\ \forall i \right\} \\ &= \left\{ \mathbf{x} = \mathbf{C}\boldsymbol{\alpha} + \mathbf{d} \mid \boldsymbol{\alpha} \in \mathbb{R}^p \right\} \quad \text{(affine set representation)} \end{aligned}$$

where $\mathbf{C} \in \mathbb{R}^{M \times p}$ is of full column rank, $\mathbf{d} \in \text{aff}\, \mathcal{A}$, and

$$\text{affdim}\, \mathcal{A} = p \leq \min\{N-1, M\}.$$

- $\mathcal{A}$ is *affinely independent* $(\mathcal{A}.\mathcal{I}.)$ with $\text{affdim}\, \mathcal{A} = N-1$ if the set $\{\mathbf{a} - \mathbf{a}_i \mid \mathbf{a} \in \mathcal{A}, \mathbf{a} \neq \mathbf{a}_i\}$ is linearly independent for any i; moreover,

$$\text{aff}\, \mathcal{A} = \{\mathbf{x} \mid \mathbf{b}^T\mathbf{x} = h\} \triangleq \mathcal{H}(\mathbf{b}, h) \quad \text{(when } M = N)$$

is a hyperplane, where $(\mathbf{b}, h)$ can be determined from $\mathcal{A}$ (closed-form expressions available).

Convex sets and convex functions-4

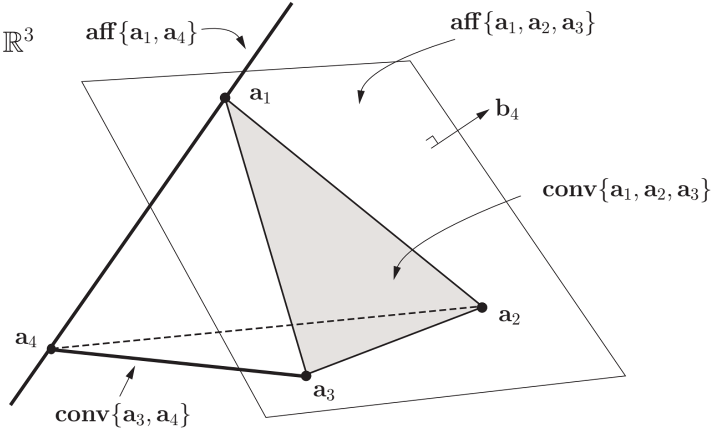


Figure 1: An illustration in $\mathbb{R}^3$, where $\mathbf{conv}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3\}$ is a simplex defined by the shaded triangle, and $\mathbf{conv}\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$ is a simplex (and also a simplest simplex) defined by the tetrahedron with the four extreme points $\{\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3, \mathbf{a}_4\}$.

Convex sets and convex functions-5

- Let $\mathcal{A} = \{\mathbf{a}_1, \ldots, \mathbf{a}_N\} \subset \mathbb{R}^M$. Convex hull of $\mathcal{A}$ *(the smallest convex set containing $\mathcal{A}$)* is defined as

$$\text{conv}\,\mathcal{A} = \left\{ \mathbf{x} = \sum_{i=1}^{N} \theta_i \mathbf{a}_i \mid \sum_{i=1}^{N} \theta_i = 1,\ \theta_i \in \mathbb{R}_+\ \forall i \right\} \subset \text{aff}\,\mathcal{A}$$

  • $\text{conv}\,\mathcal{A}$ is called a *simplex* (a polytope with N vertices) if $\mathcal{A}$ is $\mathcal{A}.\mathcal{I}.$

  • When $\mathcal{A}$ is $\mathcal{A}.\mathcal{I}.$ and $M = N - 1$, $\text{conv}\,\mathcal{A}$ is called a *simplest simplex of N vertices* (i.e., $\mathbf{a}_1, \ldots, \mathbf{a}_N$);

$$\text{aff}\,(\mathcal{A} \setminus \{\mathbf{a}_i\}) = \mathcal{H}(\mathbf{b}_i, h_i),\ i \in \mathcal{I}_N = \{1, \ldots, N\}$$

  is a *hyperplane*, where the N boundary hyperplane parameters $\{(\mathbf{b}_i, h_i), i = 1, \ldots, N\}$ can be uniquely determined from $\mathcal{A}$ (*closed-form expressions available*) and vice versa; this geometry fact plays an essential role in *hyperspectral unmixing* (a cutting edge research in remote sensing) to be introduced in Part II.

Convex sets and convex functions-6

- **Convex function:** f is convex if $\textbf{dom}\ f$ (the domain of f) is a convex set, and for all $\mathbf{x}, \mathbf{y} \in \textbf{dom}\ f$,

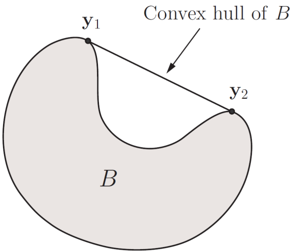
$$f(\theta\mathbf{x} + (1-\theta)\mathbf{y}) \leq \theta f(\mathbf{x}) + (1-\theta) f(\mathbf{y}),\ \forall\, 0 \leq \theta \leq 1. \tag{4}$$

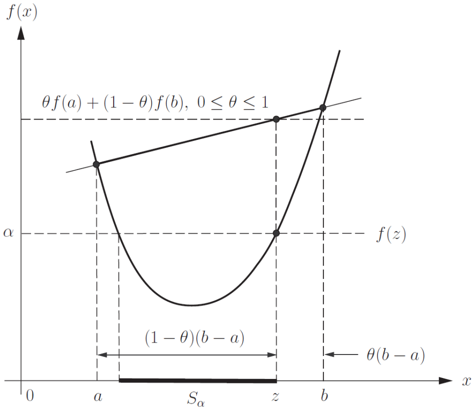
  - f is *concave* if $-f$ is convex.

Some Examples of Convex Functions

- An *affine function* $f(\mathbf{x}) = \mathbf{a}^T\mathbf{x} + b$ is both convex and concave on $\mathbb{R}^n$.
- $f(\mathbf{x}) = \mathbf{x}^T\mathbf{P}\mathbf{x} + 2\mathbf{q}^T\mathbf{x} + r$, where $\mathbf{P} \in \mathbb{S}^n, \mathbf{q} \in \mathbb{R}^n$ and $r \in \mathbb{R}$ is convex if and only if $\mathbf{P} \in \mathbb{S}^n_+$.
- Every norm on $\mathbb{R}^n$ (e.g., $\|\cdot\|_p$ for $p \geq 1$) is convex.
- Linear function $f(\mathbf{X}) = \mathrm{Tr}(\mathbf{A}\mathbf{X})$ (where $\mathrm{Tr}(\mathbf{V})$ denotes the trace of a square matrix $\mathbf{V}$) is both convex and concave on $\mathbb{R}^{n\times n}$.
- $f(\mathbf{X}) = -\log\ \det(\mathbf{X})$ is convex on $\mathbb{S}^n_{++}$.

Convex sets and convex functions-2





- Left plot: $\frac{\mathbf{y}_1+\mathbf{y}_2}{2} \notin B$, implying that B is not a convex set; right plot: $f(x)$ is a convex function (by (4)).

Ways Proving Convexity of a Function

First-order Condition

Suppose that f is differentiable. f is a convex function if and only if $\mathbf{dom}\ f$ is a convex set and

$$f(\mathbf{y}) \geq f(\mathbf{x}) + \nabla f(\mathbf{x})^T(\mathbf{y} - \mathbf{x}) \quad \forall \mathbf{x}, \mathbf{y} \in \mathbf{dom}\ f. \tag{5}$$

Second-order Condition

Suppose that f is twice differentiable. f is a convex function if and only if $\mathbf{dom}\ f$ is a convex set and

$$\nabla^2 f(\mathbf{x}) \succeq \mathbf{0} \text{ (positive-semidefinite)}, \forall \mathbf{x} \in \mathbf{dom}\ f. \tag{6}$$

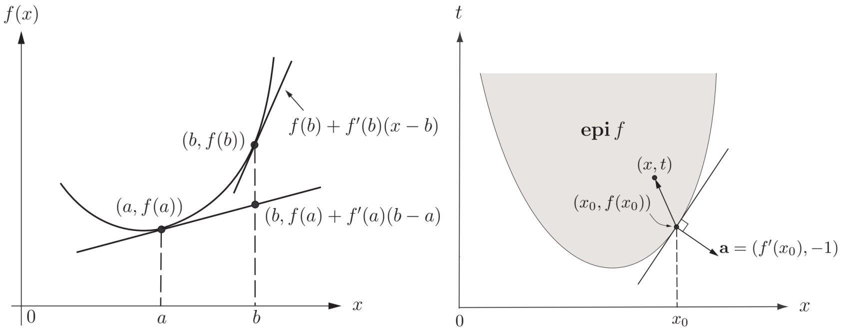
Epigraph

The *epigraph* of a function $f : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\mathbf{epi}\ f = \{(\mathbf{x}, t) \mid \mathbf{x} \in \mathbf{dom}\ f,\ f(\mathbf{x}) \leq t\} \subseteq \mathbb{R}^{n+1}. \tag{7}$$

A function f is convex if and only if $\mathbf{epi}\ f$ is a convex set.

First-order Condition and Epigraph



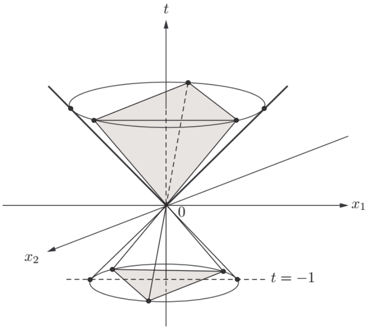
- Left plot: first-order condition for a convex function f for the one-dimensional case: $f(b) \geq f(a) + f'(a)(b - a)$, for all $a, b \in \mathbf{dom}\ f$; right plot: the epigraph of a convex function $f : \mathbb{R} \to \mathbb{R}$.

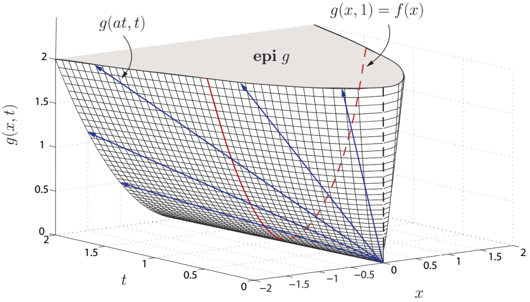
Convex Functions (Cont.)

Convexity Preserving Operations

- *Intersection, $\bigcap_i C_i$ of convex sets C_i is a convex set;*
  Nonnegative weighted sum, $\sum_i \theta_i f_i$ (where $\theta_i \geq 0$) of convex functions f_i is a convex function.
- *Image, $h(C) = \{h(\mathbf{x}) \mid \mathbf{x} \in C)\}$, of a convex set C via affine mapping $h(\mathbf{x}) \triangleq \mathbf{A}\mathbf{x} + \mathbf{b}$, is a convex set;*
  Composition $f(h(\mathbf{x}))$ of a convex function f with affine mapping h is a convex function;
- *Image, $p(C)$, of a convex set C via perspective mapping $p(\mathbf{x}, t) \triangleq \mathbf{x}/t$ is a convex set;*
  Perspective, $g(\mathbf{x}, t) = tf(\mathbf{x}/t)$ (where $t > 0$) of a convex function f is a convex function.

Perspective Mapping & Perspective of a Function





- Left plot: pinhole camera interpretation of the perspective mapping $p(x, t) = x/t, \ t > 0$;
  right plot: epigraph of the perspective $g(x, t) = tf(x/t), \ t > 0$ of $f(x) = x^2$, where each ray is associated with $g(at, t) = a^2 t$ for a different value of a.

Convex optimization problem

- **Convex problem:**

$$\text{(CVXP)} \quad p^\star = \min_{\mathbf{x}\in\mathcal{C}} \; f(\mathbf{x})$$

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

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

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

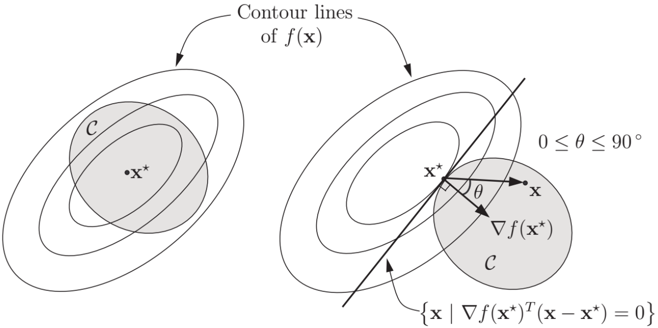
$$\mathcal{D} = \mathbf{dom}\, f \;\cap\; \left\{\bigcap_{i=1}^{m} \mathbf{dom}\ f_i\right\} \bigcap \left\{\bigcap_{i=1}^{p} \mathbf{dom}\ h_i\right\}$$

is called the problem domain.

- **Advantages:**
  - *Global optimality*: $\mathbf{x}^\star$ 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.

- **An optimality criterion:** Any suboptimal solution to (CVXP) is *globally optimal*. Assume that f is differentiable. Then a point $\mathbf{x}^\star \in \mathcal{C}$ is optimal *if and only if*

$$\nabla f(\mathbf{x}^\star)^T(\mathbf{x} - \mathbf{x}^\star) \geq 0, \ \forall \mathbf{x} \in \mathcal{C} \tag{8}$$



Case 1: $\nabla f(\mathbf{x}^\star) = \mathbf{0}_n$ Case 2: $\nabla f(\mathbf{x}^\star)^T(\mathbf{x} - \mathbf{x}^\star) \geq 0, \ \forall \mathbf{x} \in \mathcal{C}$

Global optimality and solution

- Because the optimality criterion (8) can only solve limited convex problems, a complementary approach for solving problem (CVXP) is founded on the "*duality theory*".

  - **Dual problem:**

$$
\begin{aligned}
&\mathcal{L}(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\nu}) \triangleq f(\mathbf{x}) + \textstyle\sum_{i=1}^{m} \lambda_i f_i(\mathbf{x}) + \sum_{i=1}^{p} \nu_i h_i(\mathbf{x}) \quad \text{(Lagrangian)}\\
&g(\boldsymbol{\lambda},\boldsymbol{\nu}) = \inf_{\mathbf{x}\in\mathcal{D}} \mathcal{L}(\mathbf{x},\boldsymbol{\lambda},\boldsymbol{\nu}) > -\infty \quad \text{(dual function)}\\
&d^\star = \max\ \{g(\boldsymbol{\lambda},\boldsymbol{\nu}) \mid \boldsymbol{\lambda} \succeq \mathbf{0}, \boldsymbol{\nu} \in \mathbb{R}^p\} \quad \text{(dual problem)}\\
&\quad \leq p^\star = \min\ \{f(\mathbf{x}) \mid \mathbf{x} \in \mathcal{C})\} \quad \text{(primal problem (CVXP))}
\end{aligned}
\tag{9}
$$

where $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)$ and $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_p)$ are dual variables.

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

Global optimality and solution

- **KKT conditions:**
  Suppose that $f, f_1, \ldots, f_m, h_1, \ldots, h_p$ are differentiable and $\mathbf{x}^\star$ is primal optimal and $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ is dual optimal to problem (CVXP). 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.,* $\text{relint}\ \mathcal{C} \neq \emptyset$), the KKT conditions for solving $\mathbf{x}^\star$ *and* $(\boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star)$ are as follows:

  $$\nabla_{\mathbf{x}} \mathcal{L}(\mathbf{x}^\star, \boldsymbol{\lambda}^\star, \boldsymbol{\nu}^\star) = \mathbf{0}, \tag{10a}$$

  $$f_i(\mathbf{x}^\star) \leq 0,\ i = 1, \ldots, m, \tag{10b}$$

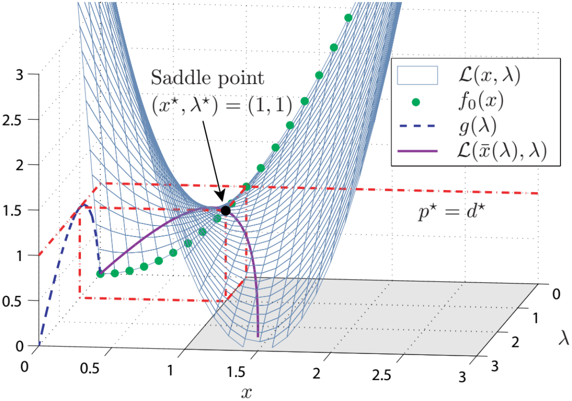
  $$h_i(\mathbf{x}^\star) = 0,\ i = 1, \ldots, p, \tag{10c}$$

  $$\lambda_i^\star \geq 0,\ i = 1, \ldots, m, \tag{10d}$$

  $$\lambda_i^\star f_i(\mathbf{x}^\star) = 0,\ i = 1, \ldots, m. \quad \textit{(complementary slackness)} \tag{10e}$$

  The above KKT conditions (10) and the optimality criterion (8) are equivalent under Slater's condition.

Strong Duality



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

Standard Convex Optimization Problems

Linear Programming (LP) - Inequality Form

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{G}\mathbf{x} \preceq \mathbf{h}, \quad (\preceq \text{ stands for componentwise inequality}) \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{11}$$

where $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{p\times n}$, $\mathbf{b} \in \mathbb{R}^p$, $\mathbf{G} \in \mathbb{R}^{m\times n}$, $\mathbf{h} \in \mathbb{R}^m$, and $\mathbf{x} \in \mathbb{R}^n$ is the unknown vector variable.

LP- Standard Form

$$\begin{aligned} \min \quad & \mathbf{c}^T\mathbf{x} \\ \text{s.t.} \quad & \mathbf{x} \succeq \mathbf{0}, \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{12}$$

where $\mathbf{c} \in \mathbb{R}^n$, $\mathbf{A} \in \mathbb{R}^{p\times n}$, $\mathbf{b} \in \mathbb{R}^p$, and $\mathbf{x} \in \mathbb{R}^n$ is the unknown vector variable.

Standard Convex Optimization Problems (Cont.)

Quadratic Programming (QP): Convex if and only if $\mathbf{P} \succeq \mathbf{0}$ (i.e., $\mathbf{P}$ is positive semidefinite)

$$\begin{aligned} \min \ & \frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{q}^T\mathbf{x} + r \\ \text{s.t. } & \mathbf{A}\mathbf{x} = \mathbf{b}, \ \mathbf{G}\mathbf{x} \preceq \mathbf{h}, \end{aligned} \tag{13}$$

where $\mathbf{P} \in \mathbb{S}^n$, $\mathbf{G} \in \mathbb{R}^{m\times n}$, and $\mathbf{A} \in \mathbb{R}^{p\times n}$.

Quadratically constrained QP (QCQP): Convex if and only if $\mathbf{P}_i \succeq \mathbf{0}, \ \forall i$

$$\begin{aligned} \min \ & \frac{1}{2}\mathbf{x}^T\mathbf{P}_0\mathbf{x} + \mathbf{q}_0^T\mathbf{x} + r_0 \\ \text{s.t. } & \frac{1}{2}\mathbf{x}^T\mathbf{P}_i\mathbf{x} + \mathbf{q}_i^T\mathbf{x} + r_i \leq 0, \ i = 1, \ldots, m, \\ & \mathbf{A}\mathbf{x} = \mathbf{b}, \end{aligned} \tag{14}$$

where $\mathbf{P}_i \in \mathbb{S}^n$, $i = 0, 1, \ldots, m$, and $\mathbf{A} \in \mathbb{R}^{p\times n}$.

Standard Convex Optimization Problems (Cont.)

Second-order cone programming (SOCP)

$$\begin{aligned} \min \ & \mathbf{c}^T \mathbf{x} \\ \text{s.t.} \ & \|\mathbf{A}_i \mathbf{x} + \mathbf{b}_i\|_2 \leq \mathbf{f}_i^T \mathbf{x} + d_i, \ i = 1, \ldots, m, \\ & \mathbf{F}\mathbf{x} = \mathbf{g}, \end{aligned} \tag{15}$$

where $\mathbf{A}_i \in \mathbb{R}^{n_i \times n}$, $\mathbf{b}_i \in \mathbb{R}^{n_i}$, $\mathbf{f}_i \in \mathbb{R}^n$, $d_i \in \mathbb{R}$, $\mathbf{F} \in \mathbb{R}^{p \times n}$, $\mathbf{g} \in \mathbb{R}^p$, $\mathbf{c} \in \mathbb{R}^n$, and $\mathbf{x} \in \mathbb{R}^n$ is the vector variable.

Semidefinite programming (SDP) - Standard Form

$$\begin{aligned} \min \ & \mathrm{Tr}(\mathbf{C}\mathbf{X}) \\ \text{s.t.} \ & \mathbf{X} \succeq \mathbf{0}, \\ & \mathrm{Tr}(\mathbf{A}_i \mathbf{X}) = b_i, \ i = 1, \ldots, p, \end{aligned} \tag{16}$$

with variable $\mathbf{X} \in \mathbb{S}^n$, where $\mathbf{A}_i \in \mathbb{S}^n$, $\mathbf{C} \in \mathbb{S}^n$, and $b_i \in \mathbb{R}$.

Alternating direction method of multipliers (ADMM)

- Consider the following convex optimization problem:

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^n,\mathbf{z}\in\mathbb{R}^m} \quad & f_1(\mathbf{x}) + f_2(\mathbf{z}) \\ \text{s.t.} \quad & \mathbf{x}\in\mathcal{S}_1,\ \mathbf{z}\in\mathcal{S}_2 \\ & \mathbf{z} = \mathbf{A}\mathbf{x} \end{aligned} \tag{17}$$

  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 (17) is given by

$$\max_{\boldsymbol{\nu}\in\mathbb{R}^m} \min_{\mathbf{x}\in\mathcal{S}_1,\mathbf{z}\in\mathcal{S}_2} \left\{ f_1(\mathbf{x}) + f_2(\mathbf{z}) + \frac{c}{2}\|\mathbf{A}\mathbf{x}-\mathbf{z}\|_2^2 + \boldsymbol{\nu}^T(\mathbf{A}\mathbf{x}-\mathbf{z}) \right\}, \tag{18}$$

  where c is a penalty parameter, and $\boldsymbol{\nu}$ is the dual variable associated with the equality constraint in (17).

ADMM (Conti)

- Inner minimization (convex problems):

$$\mathbf{z}(q+1) = \arg\min_{\mathbf{z}\in\mathcal{S}_2} \left\{ f_2(\mathbf{z}) - \boldsymbol{\nu}(q)^T\mathbf{z} + \frac{c}{2}\left\|\mathbf{A}\mathbf{x}(q) - \mathbf{z}\right\|_2^2 \right\}, \quad (19a)$$

$$\mathbf{x}(q+1) = \arg\min_{\mathbf{x}\in\mathcal{S}_1} \left\{ f_1(\mathbf{x}) + \boldsymbol{\nu}(q)^T\mathbf{A}\mathbf{x} + \frac{c}{2}\left\|\mathbf{A}\mathbf{x} - \mathbf{z}(q+1)\right\|_2^2 \right\}. \quad (19b)$$

ADMM Algorithm

1: Set $q = 0$, choose $c > 0$.
2: Initialize $\boldsymbol{\nu}(q)$ and $\mathbf{x}(q)$.
3: **repeat**
4: Solve (19a) and (19b) for $\mathbf{z}(q+1)$ and $\mathbf{x}(q+1)$ by two distributed equipments including *the information exchange of $\mathbf{z}(q+1)$ and $\mathbf{x}(q+1)$ between them;*
5: $\boldsymbol{\nu}(q+1) = \boldsymbol{\nu}(q) + c\,(\mathbf{A}\mathbf{x}(q+1) - \mathbf{z}(q+1))$;
6: $q := q+1$;
7: **until** the predefined stopping criterion is satisfied.

- When $\mathcal{S}_1$ 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 (17).

Nonconvex problem

- **Reformulation into a convex problem:**
  Equivalent representations (e.g. epigraph representations); function transformation; change of variables, etc.

- **Stationary-point solutions**: Suppose that $\mathcal{C}$ is closed and convex but f is nonconvex. A point $\mathbf{x}^\star$ is a *stationary point* of problem (1) if

$$f'(\mathbf{x}^\star;\mathbf{v}) \triangleq \liminf_{\lambda\downarrow 0} \frac{f(\mathbf{x}^\star+\lambda\mathbf{v})-f(\mathbf{x}^\star)}{\lambda} \geq 0 \quad \forall \mathbf{x}^\star+\mathbf{v}\in\mathcal{C} \tag{20}$$

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

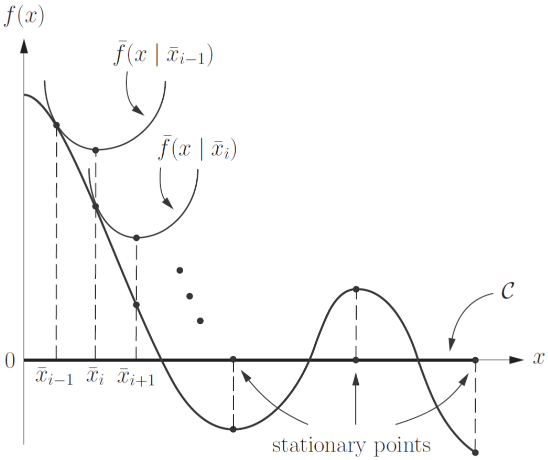
  where $f'(\mathbf{x}^\star;\mathbf{v})$ is the *directional derivative* of f at a point $\mathbf{x}^\star$ in direction $\mathbf{v}$. Block successive upper bound minimization (BSUM) [Razaviyayn'13] guarantees a stationary-point solution under some convergence conditions.

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

[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. 11261153, 2013.

Stationary points and BSUM

- Illustration of stationary points of a nonconvex function and BSUM for one-dimensional case.



Nonconvex problem

- Approximate solutions when f is a convex function but $\mathcal{C}$ is a nonconvex set:
  - Convex *restriction* to $\mathcal{C}$: *Successive convex approximation (SCA)*

$$\boldsymbol{x}_i^\star = \arg\min_{\mathbf{x}\in C_i} f(\mathbf{x}) \in C_{i+1} \tag{21}$$

where $C_i \subset \mathcal{C}$ is convex for all i. Then

$$f(\boldsymbol{x}_{i+1}^\star) = \min_{\mathbf{x}\in\mathcal{C}_{i+1}} f(\mathbf{x}) \leq f(\boldsymbol{x}_i^\star), \quad \mathcal{C}_{i+1} \triangleq \underline{\cup_{j=1}^{i+1} C_j \subset \mathcal{C}} \tag{22}$$

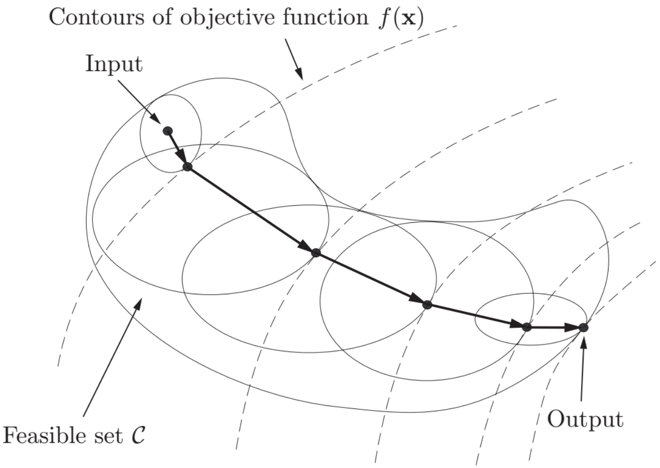
After convergence, a stationary-point solution may be obtained depending on the problem under consideration and the choice of the convex set C_i.

  - Convex *relaxation* to $\mathcal{C}$ (e.g., *semidefinite relaxation (SDR)*):

$$\begin{aligned} \mathcal{C}' &= \{\mathbf{X} \in \mathbb{S}_+^n \mid \operatorname{rank}(\mathbf{X}) = 1\} \subset \mathcal{C} \text{ relaxed to conv } \mathcal{C}' = \mathbb{S}_+^n \text{ (SDR)}; \\ \mathcal{C}' &= \{-3, -1, +1, +3\} \subset \mathcal{C} \text{ relaxed to conv } \mathcal{C}' = [-3, 3] \end{aligned} \tag{23}$$

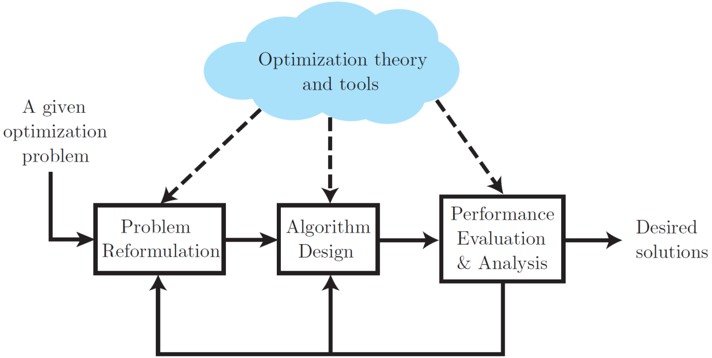
The obtained $\boldsymbol{X}^\star$ or $\boldsymbol{x}^\star$ may not be feasible to problem (1); for SDR, a good approximate solution can be obtained from $\boldsymbol{X}^\star$ via *Gaussian randomization*.

Successive Convex Approximation (SCA)



Algorithm development

- **Foundamental theory and tools:** Calculus, linear algebra, matrix analysis and computations, convex sets, convex functions, convex problems (e.g., geometric program (GP), LP, QP, SOCP, SDP), duality, interior-point method; CVX and SeDuMi.



Outline

1. Part I: Fundamentals of Convex Optimization
2. **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

Introduction to hyperspectral unmixing (HU)

A **hyperspectral sensor** records electromagnetic ‘fingerprints’ (scattering patterns) of substances (materials) in a scene, known as *spectral signatures*, over hundreds of spectral bands from visible to short-wave infrared wavelength region.

Airborne Sensors. Courtesy: http://masterweb.jpl.nasa.gov/

Issue: *mixed pixel spectra* in the hyperspectral image/data (due to limited spatial resolution of the hypersepctral sensor) must be decomposed for identifying the underlying materials *(i.e., spectral unmixing)* [Keshava’02].

[Keshava’02] N. Keshava et al., “Spectral unmixing,” *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44-57, Jan. 2002.

Introduction to hyperspectral unmixing (HU)

Airborne Sensors. Courtesy: http://masterweb.jpl.nasa.gov/

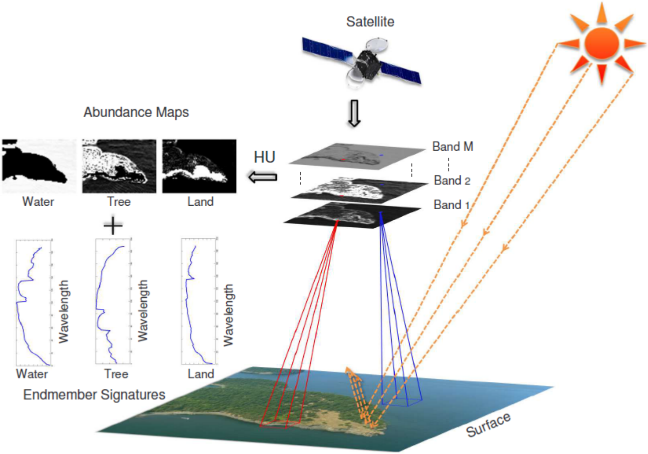
Hyperspectral unmixing (HU) is a signal processing procedure of extracting hidden *spectral signatures of substances* (called *endmembers*) and *the corresponding proportions* (called *abundances*) (i.e., *distribution map of substances*) in a scene, from the hyperspectral observations [Keshava'02].

Applications: Terrain classification, mineral identification and quantification, agricultural monitoring, military surveillance, space object identification, etc.

[Keshava'02] N. Keshava et al., "Spectral unmixing," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 44-57, Jan. 2002.

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.

Signal model and assumptions

Linear Mixing Model

The pixel vector $\mathbf{x}[n]$ of M spectral bands can be represented as a linear combination of N endmember signatures $\{\mathbf{a}_1, \ldots, \mathbf{a}_N\}$, i.e.,

$$\mathbf{x}[n] = \begin{cases} \mathbf{A}\mathbf{s}[n] = \sum_{i=1}^{N} s_i[n]\mathbf{a}_i \in \mathbb{R}^M, \;\; n \in \mathcal{I}_L \triangleq \{1, \ldots, L\}. \\ s_j[n]\mathbf{a}_j \;\; \text{(a pure pixel) if } s_i[n] = 0 \; \forall i \neq j \end{cases} \tag{24}$$

- $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N] \in \mathbb{R}^{M \times N}$ (spectral signature matrix or mixing matrix).
- $\mathbf{s}[n] \triangleq [s_1[n], \ldots, s_N[n]]^T \in \mathbb{R}^N$ is the nth abundance vector;
  $\boldsymbol{s}_i = \{s_i[n] \mid n \in \mathcal{I}_L\}$ is called source i or abundance map i.

Standard Assumptions (Non-statistical) [Keshava'02]

(A1) $\mathbf{s}[n] \succeq \mathbf{0}_N$ and $\sum_{i=1}^{N} s_i[n] = 1, \; \forall n \in \mathcal{I}_L$.

(A2) $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N] \succeq \mathbf{0}_{M \times N}$ has full column rank and $\min\{L, M\} \geq N$.

Craig's HU criterion

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

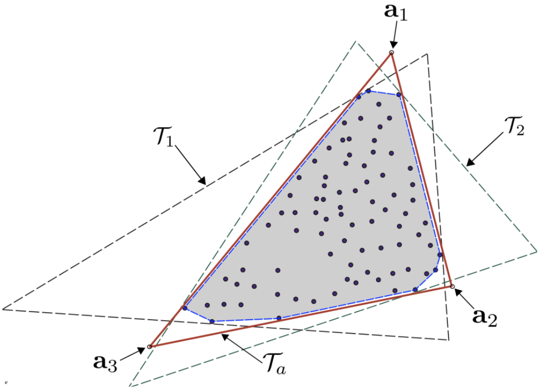


Figure 2: Visually, $\text{vol}(\mathcal{T}_a) < \text{vol}(\mathcal{T}_i)$, $i \in \mathcal{I}_2$ (the dots are data points $\mathbf{x}[n]$).

[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 HU criterion (Cont.)

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

$$\{\widehat{\mathbf{a}}_1,\ldots,\widehat{\mathbf{a}}_N\} \in \arg\min_{\boldsymbol{b}_i\in\mathbb{R}^M\,\forall i} \text{vol}(\text{conv}\{\boldsymbol{b}_1,\ldots,\boldsymbol{b}_N\}) \quad \text{s.t. } \boldsymbol{X}\subset \text{conv}\{\boldsymbol{b}_1,\ldots,\boldsymbol{b}_N\} \tag{25}$$

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

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

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

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

[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.
[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.

Dimension reduction and problem formulation

- Dimension-reduced (DR) representation for noise suppression and computational complexity reduction: Due to $\mathrm{affdim}\ \mathcal{T}_a = N-1$,

$$\mathbf{x}[n] = \mathbf{C}\widetilde{\mathbf{x}}[n] + \mathbf{d} \in \underline{\boldsymbol{X} \subset \mathcal{T}_a \subset \text{aff } \mathcal{T}_a \subset \mathbb{R}^M} \quad \text{(affine set fitting)} \tag{26}$$

$$\Rightarrow \widetilde{\mathbf{x}}[n] = \mathbf{C}^\dagger(\mathbf{x}[n] - \mathbf{d})$$

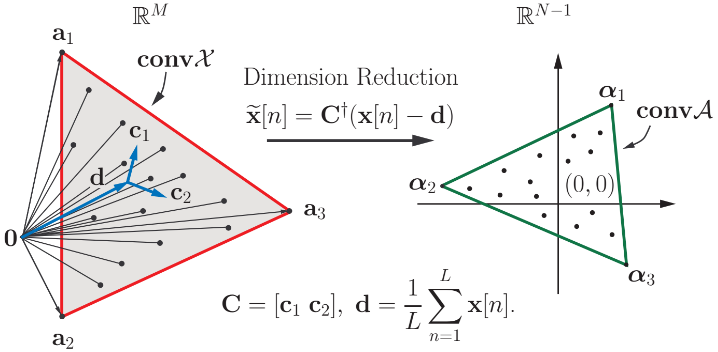
$$= \sum_{i=1}^{N} s_i[n]\boldsymbol{\alpha}_i \in \mathcal{T}_\alpha = \text{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\} \subset \text{aff } \mathcal{T}_\alpha \subset \mathbb{R}^{N-1} \tag{27}$$

  - $\mathbf{C} \triangleq [\boldsymbol{q}_1(\mathbf{U}\mathbf{U}^T), \ldots, \boldsymbol{q}_{N-1}(\mathbf{U}\mathbf{U}^T)]$ (a semiunitary matrix), where $\boldsymbol{q}_i(\cdot)$ is the ith principal eigenvector, and $\mathbf{U} \triangleq [\mathbf{x}[1] - \mathbf{d}, \ldots, \mathbf{x}[L] - \mathbf{d}] \in \mathbb{R}^{M \times L}$ (mean removed data matrix);
  - $\mathbf{d} \triangleq \frac{1}{L}\sum_{n=1}^{L} \mathbf{x}[n]$ (mean of the data set $\boldsymbol{X}$);
  - $\mathcal{T}_\alpha$ is an N-vertex simplest simplex;

$$\mathcal{X} = \{\widetilde{\mathbf{x}}[1], \ldots, \widetilde{\mathbf{x}}[L]\} \subset \mathcal{T}_\alpha \subset \mathbb{R}^{N-1} \quad \textit{(DR data set)} \tag{28}$$

$$\boldsymbol{\alpha}_i \triangleq \mathbf{C}^\dagger(\mathbf{a}_i - \mathbf{d}) \in \mathbb{R}^{N-1} \quad \textit{(DR endmembers)}. \tag{29}$$

Illustration of Dimension Reduction



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

Dimension reduction and problem formulation

- Problem (25) can be reformulated in the DR space as:

$$\{\widehat{\boldsymbol{\alpha}}_1, \ldots, \widehat{\boldsymbol{\alpha}}_N\} \in \arg \min_{\boldsymbol{\beta}_i \in \mathbb{R}^{N-1} \forall i} \left\{ \text{vol}(\text{conv}\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\}) = \frac{|\det(\mathbf{B})|}{(N-1)!} \right\}$$
$$\text{s.t. } \mathcal{X} \subset \text{conv}\{\boldsymbol{\beta}_1, \ldots, \boldsymbol{\beta}_N\} \qquad (30)$$

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

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

$$\widehat{\mathbf{a}}_i = \mathbf{C}\widehat{\boldsymbol{\alpha}}_i + \mathbf{d}, \ \forall i \in \mathcal{I}_N \ \text{ (cf. (29))}.$$

Existing Challenges

- Pure pixel assumption (PPA) enables various simple and fast blind HU algorithmic schemes *(for finding the purest pixels in the data set $\boldsymbol{X}$ or the DR data set $\mathcal{X}$)*, but it is often seriously infringed.
- Without requiring the PPA, Craig's blind HU criterion identifies the N-vertex minimum-volume data-enclosing simplex $\widehat{\mathcal{T}}_a \subset \mathbb{R}^M$, but suffering from heavy simplex volume computations.

New Philosophy: Hyperplane-based CSI Algorithm

Geometry Fact 1 [Lin'16]:

$$s_i[n] = 0 \text{ if and only if } \tilde{\mathbf{x}}[n] \in \mathcal{H}_i,$$

where $\mathcal{H}_i \triangleq \text{aff}(\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\} \setminus \{\boldsymbol{\alpha}_i\}) \subset \mathbb{R}^{N-1}$ is the boundary hyperplane of the DR endmembers' *simplest simplex* $\mathcal{T}_\alpha \subset \mathbb{R}^{N-1}$.

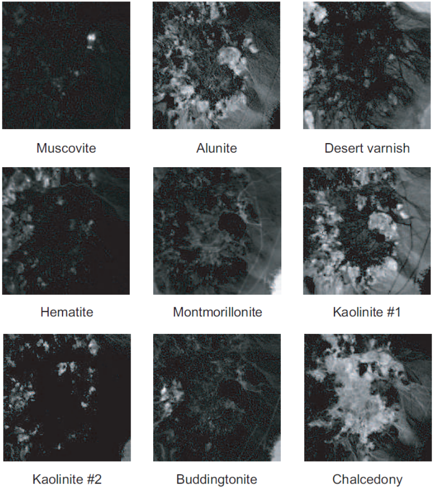
Observation

The abundance maps often show *large sparseness* [Cichocki'09], and many pixels lying on or close to the boundary hyperplanes of the Craig's simplex.

Geometry Fact 2:

A *simplest simplex* of N vertices can be defined by its N boundary hyperplanes $\mathcal{H}_i$ (each containing $N-1$ vertices) and vice versa.

Real data experiments



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.

Super-fast HU algorithm: HyperCSI

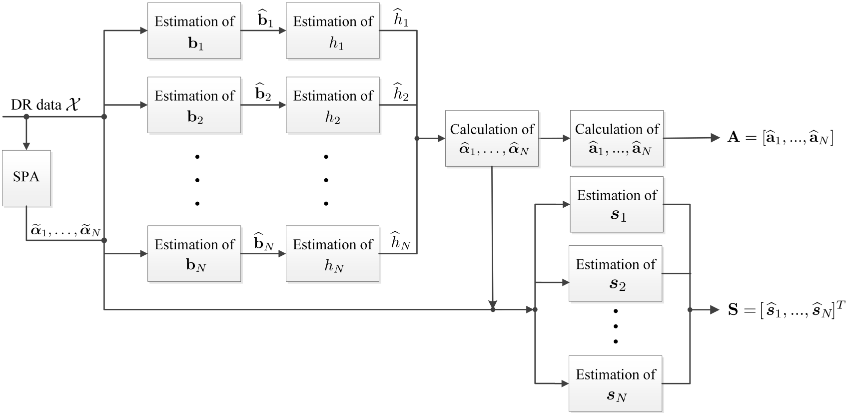
HyperCSI Algorithm: A novel practical HU algorithm

- HyperCSI algorithm tries to estimate the N boundary hyperplanes of the N-vertex simplest simplex $\widehat{\mathcal{T}}_\alpha \subset \mathbb{R}^{N-1}$ (in DR space), without any simplex volume computations.
- Each hyperplane is constructed by $N-1$ $\mathcal{A}.\mathcal{I}.$ data pixels that are identified via simple linear algebraic computations.
- With better performance than state-of-the-art algorithms, HyperCSI is at least computationally efficient as PPA based HU algorithms.

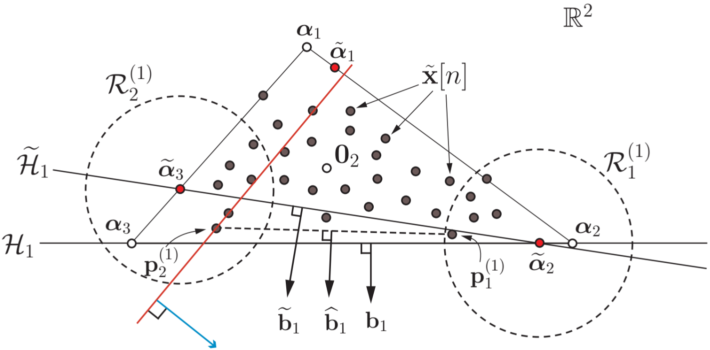
[Lin'16] C.-H. Lin et al., "A fast hyperplane-based minimum-volume enclosing simplex algorithm for blind hyperspectral unmixing," *IEEE Trans. Signal Processing*, vol. 64, no.8, pp. 1946-1961, Apr. 2016.

Block diagram of HyperCSI algorithm

- *An algorithm with parallel processing structure* for estimation of normal vectors $(\mathbf{b}_i)$, inner product parameters (h_i), and abundance maps $(\boldsymbol{s}_i)$, where the PPA based successive projection algorithm (SPA) is employed to obtain initial estimates $\widetilde{\alpha}_1, \ldots, \widetilde{\alpha}_N$.



Graphical illustration of HyperCSI



- Why $\mathcal{R}_k^{(i)}$ should be disjoint? Consider $\{\mathbf{p}_2^{(1)}, \mathbf{q}\}$ identified by (33).
- Why not $\widetilde{\mathbf{b}}_1$? The purest pixel $\widetilde{\boldsymbol{\alpha}}_3$ may not be close to $\mathcal{H}_1 = \text{aff}\{\boldsymbol{\alpha}_2, \boldsymbol{\alpha}_3\}$, leading to nontrivial orientation difference between $\widetilde{\mathbf{b}}_1$ and $\mathbf{b}_1$.
- However, $\{\mathbf{p}_1^{(1)}, \mathbf{p}_2^{(1)}\}$ identified by (34) are very close to $\mathcal{H}_1$, so the orientations of $\widehat{\mathbf{b}}_1$ and $\mathbf{b}_1$ are almost the same.

HyperCSI Algorithm

Pseudocode for the HyperCSI Algorithm

1. **Given** Hyperspectral data $\{\mathbf{x}[1], \ldots, \mathbf{x}[L]\}$, number of endmembers N, and $\eta = 0.9$.
2. Obtain DR dataset $\mathcal{X} = \{\widetilde{\mathbf{x}}[1], \ldots, \widetilde{\mathbf{x}}[L]\}$ by (27). (DR processing)
3. Obtain purest pixels $\{\widetilde{\boldsymbol{\alpha}}_1, \ldots, \widetilde{\boldsymbol{\alpha}}_N\}$ by SPA. (preprocessing)
4. Obtain $\widetilde{\mathbf{b}}_i = \boldsymbol{v}_i(\widetilde{\boldsymbol{\alpha}}_1, \ldots, \widetilde{\boldsymbol{\alpha}}_N)$ by (31), and then obtain $\mathcal{A}.\mathcal{I}.$ set $\{\mathbf{p}_1^{(i)}, \ldots, \mathbf{p}_{N-1}^{(i)}\}$ by (34) $\forall i \in \mathcal{I}_N$. ($\mathcal{A}.\mathcal{I}.$ w.p.1)
5. Obtain $\widehat{\mathbf{b}}_i$ and $\widehat{h}_i$ by (35) and (36), respectively, $\forall i \in \mathcal{I}_N$.
6. Obtain c' by (37), and set $c = c'/\eta$. (further denoising)
7. Obtain $\widehat{\boldsymbol{\alpha}}_i$ by (38) and then obtain $\widehat{\mathbf{a}}_i$ by (39) $\forall i \in \mathcal{I}_N$.(identifiable w.p.1)
8. Calculate $\widehat{\mathbf{s}}[n] = [\widehat{s}_1[n], \ldots, \widehat{s}_N[n]]^T$ by (41) for all $n \in \mathcal{I}_L$. $\mathcal{O}(N^2L)$
9. **Output** The endmember estimates $\{\widehat{\mathbf{a}}_1, \ldots, \widehat{\mathbf{a}}_N\}$ and abundance maps $\{\widehat{\boldsymbol{s}}_1, \ldots, \widehat{\boldsymbol{s}}_N\}$.

Hyperplane representation of Simplest Simplex

- Since $\mathcal{T}_\alpha$ is a *simplest simplex* in the DR space $\mathbb{R}^{N-1}$, enabling the following handy parameterization for $\mathcal{H}_i$:

$$\mathcal{H}_i \triangleq \text{aff}(\mathcal{T}_\alpha \setminus \{\boldsymbol{\alpha}_i\}) = \{\mathbf{x} \in \mathbb{R}^{N-1} \mid \mathbf{b}_i^T \mathbf{x} = h_i\} \equiv \mathcal{H}(\mathbf{b}_i, h_i),$$

where $\mathbf{b}_i \in \mathbb{R}^{N-1}$ and $h_i \in \mathbb{R}$ respectively denote the outward-pointing normal vector of $\mathcal{H}_i$ and the inner product constant of $\mathcal{H}_i$.

Proposition 1 [Lin'16]

The endmembers $\boldsymbol{\alpha}_i$ can then be reconstructed by the hyperplanes $\mathcal{H}_i$:

$$\boldsymbol{\alpha}_i = \mathbf{B}_{-i}^{-1}\, \mathbf{h}_{-i}, \ \forall\, i \in \mathcal{I}_N,$$

where $\mathbf{B}_{-i} \triangleq [\mathbf{b}_1, \ldots, \mathbf{b}_{i-1}, \mathbf{b}_{i+1}, \ldots, \mathbf{b}_N]^T$;
$\mathbf{h}_{-i} \triangleq [h_1, \ldots, h_{i-1}, h_{i+1}, \ldots, h_N]^T$.

- Problem (30) can be decoupled into N parallel subproblems of estimating $(\mathbf{b}_i, h_i), \ \forall i \in \mathcal{I}_N$. Next, let us focus on how to estimate $(\mathbf{b}_i, h_i), \ \forall i \in \mathcal{I}_N$.

Normal vector $\mathbf{b}_i$ estimation

- The normal vector $\mathbf{b}_i$ can be reconstructed by $N-1$ $\mathcal{A}.\mathcal{I}.$ points on $\mathcal{H}_i$.

Proposition 2 [Lin'16]

Given $\mathcal{A}.\mathcal{I}.$ $\{\mathbf{p}_1^{(i)}, \ldots, \mathbf{p}_{N-1}^{(i)}\} \subseteq \mathcal{H}_i$, we have

$$\begin{aligned}\mathbf{b}_i &= \boldsymbol{v}_i(\mathbf{p}_1^{(i)}, \ldots, \mathbf{p}_{i-1}^{(i)}, \mathbf{0}_{N-1}, \mathbf{p}_i^{(i)}, \ldots, \mathbf{p}_{N-1}^{(i)}), \\ &\triangleq \left(\mathbf{I}_{N-1} - \mathbf{P}(\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\right) \cdot \mathbf{p}_k^{(i)},\end{aligned} \tag{31}$$

for arbitrary $k \in \mathcal{I}_{N-1}$, where

$$\begin{aligned}\mathbf{P} &\triangleq \mathbf{Q} - \mathbf{p}_k^{(i)} \cdot \mathbf{1}_{N-2}^T \in \mathbb{R}^{(N-1)\times(N-2)}, \\ \mathbf{Q} &\triangleq [\mathbf{p}_1^{(i)}, \ldots, \mathbf{p}_{k-1}^{(i)}, \mathbf{p}_{k+1}^{(i)}, \ldots, \mathbf{p}_{N-1}^{(i)}] \in \mathbb{R}^{(N-1)\times(N-2)}.\end{aligned}$$

- Motivated by Proposition 2, we aim at extracting $N-1$ $\mathcal{A}.\mathcal{I}.$ data pixels $\mathbf{p}_k^{(i)}$ from the DR dataset $\mathcal{X}$ *that are closest to* $\mathcal{H}_i$.

Normal vector $\mathbf{b}_i$ Estimation (Cont.)

Proposition 3 [Lin'16]

Observing that all the points $\mathbf{p} \in \mathcal{X}$ lie on the same side of $\mathcal{H}_i$, i.e.,

$$\mathbf{b}_i^T\mathbf{p} \leq h_i, \ \forall \mathbf{p} \in \mathcal{X}, \text{ and } \mathrm{dist}(\mathbf{p}, \mathcal{H}_i) = |h_i - \mathbf{b}_i^T\mathbf{p}|/\|\mathbf{b}_i\|,$$

we see that the point $\mathbf{p} \in \mathcal{X}$ closest to $\mathcal{H}_i$ is exactly the one with maximum of $\mathbf{b}_i^T\mathbf{p}$, because $\mathbf{b}_i$ is outward-pointing.

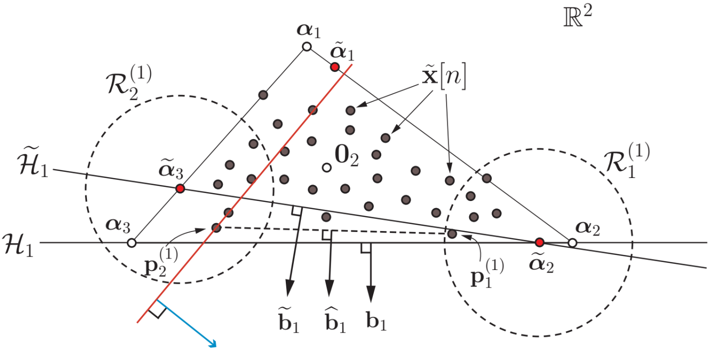
- Let $\widetilde{\mathcal{H}}_i \triangleq \mathrm{aff}(\{\widetilde{\boldsymbol{\alpha}}_1, \ldots, \widetilde{\boldsymbol{\alpha}}_N\} \setminus \{\widetilde{\boldsymbol{\alpha}}_i\}) \equiv \mathcal{H}(\widetilde{\mathbf{b}}_i, \widetilde{h}_i)$, where $\widetilde{\boldsymbol{\alpha}}_i$ are the purest pixels extracted by the *successive projection algorithm* [Arora'12].
- By Proposition 2 and Proposition 3, a naive way for estimating $\mathbf{p}_k^{(i)}$ is

$$\{\mathbf{p}_1^{(i)}, \ldots, \mathbf{p}_{N-1}^{(i)}\} \in \arg\max \{\widetilde{\mathbf{b}}_i^T(\mathbf{p}_1 + \cdots + \mathbf{p}_{N-1}) \mid \mathbf{p}_k \in \mathcal{X}\}. \quad (32)$$

$\Longrightarrow$ The identified $\mathbf{p}_k^{(i)}$ can be quite close to each other.

[Arora'12] S. Arora et al., "A practical algorithm for topic modeling with provable guarantees," arXiv preprint arXiv:1212.4777, 2012.

Normal vector $\mathbf{b}_i$ Estimation (Cont.)



- By Proposition 2 and Proposition 3, a naive way for estimating $\mathbf{p}_k^{(i)}$ is

$$\{\mathbf{p}_1^{(i)}, \ldots, \mathbf{p}_{N-1}^{(i)}\} \in \arg\max \{\widetilde{\mathbf{b}}_i^T(\mathbf{p}_1 + \cdots + \mathbf{p}_{N-1}) \mid \mathbf{p}_k \in \mathcal{X}\}. \quad (33)$$

$\Longrightarrow$ The identified $\mathbf{p}_k^{(i)}$ can be quite close to each other.

[Arora'12] S. Arora et al., "A practical algorithm for topic modeling with provable guarantees," arXiv preprint arXiv:1212.4777, 2012.

Normal vector $\mathbf{b}_i$ Estimation (Cont.)

- Accordingly, $\mathbf{p}_k^{(i)}$ for $\mathcal{H}_i$ are sifted by

$$\mathbf{p}_k^{(i)} \in \arg\max\{\widetilde{\mathbf{b}}_i^T \mathbf{p} \mid \mathbf{p} \in \mathcal{X} \cap \mathcal{R}_k^{(i)}\}, \ \forall k \in \mathcal{I}_{N-1} \tag{34}$$

where $\mathcal{R}_k^{(i)}$ are disjoint regions (norm balls with the same radius) defined by

$$\mathcal{R}_k^{(i)} \triangleq \begin{cases} \mathcal{B}(\widetilde{\boldsymbol{\alpha}}_k, r), \text{ if } k < i, \\ \mathcal{B}(\widetilde{\boldsymbol{\alpha}}_{k+1}, r), \text{ if } k \geq i, \end{cases}$$

$$\mathcal{B}(\widetilde{\boldsymbol{\alpha}}_k, r) \triangleq \{\mathbf{x} \in \mathbb{R}^N \mid \|\mathbf{x} - \widetilde{\boldsymbol{\alpha}}_k\| < r\},$$

$$r \triangleq (1/2) \cdot \min\{\|\widetilde{\boldsymbol{\alpha}}_i - \widetilde{\boldsymbol{\alpha}}_j\| \mid 1 \leq i < j \leq N\} > 0.$$

- **Normal vector estimates:**

$$\widehat{\mathbf{b}}_i = \boldsymbol{v}_i(\mathbf{p}_1^{(i)}, \ldots, \mathbf{p}_{i-1}^{(i)}, \mathbf{0}_{N-1}, \mathbf{p}_i^{(i)}, \ldots, \mathbf{p}_{N-1}^{(i)}) \ \ \forall i. \tag{35}$$

Inner product constant h_i Estimation

- Craig's simplex must tightly enclose $\mathcal{X}$:

$$\widehat{h}_i = \max \{ \widehat{\mathbf{b}}_i^T \mathbf{p} \mid \mathbf{p} \in \mathcal{X} \}. \quad (36)$$

- Considering the simplex volume expansion caused by noise, the estimated hyperplanes need to be properly shifted closer to the origin, or equivalently $\widehat{h}_i$ should be scaled down as $\widehat{h}_i/c$ for some $c \geq 1$ in general.

Non-data dependent choice of the parameter c

As $\mathbf{A} \succeq \mathbf{0}_{M\times N}$, Proposition 1 indicates $c \geq c'$ (empirically, $c = c'/\eta$ and $\eta = 0.9$), where

$$\begin{aligned} c' &\triangleq \min_{c''\geq 1} \{c'' \mid \mathbf{C}\,(\widehat{\mathbf{B}}_{-i}^{-1} \cdot \hat{\mathbf{h}}_{-i}) + c'' \cdot \mathbf{d} \succeq \mathbf{0}_M,\ \forall\, i \in \mathcal{I}_N\} \\ &= \max\{1, \max\{-v_{ij}/d_j \mid i \in \mathcal{I}_N,\ j \in \mathcal{I}_M\}\}, \end{aligned} \quad (37)$$

in which v_{ij} is the jth entry of $\mathbf{C}\,(\widehat{\mathbf{B}}_{-i}^{-1} \cdot \widehat{\mathbf{h}}_{-i})$ and d_j is the jth entry of $\mathbf{d}$.

- **Endmember estimates:**

$$\widehat{\boldsymbol{\alpha}}_i = \widehat{\mathbf{B}}_{-i}^{-1}\, \widehat{\mathbf{h}}_{-i}/c \in \mathbb{R}^{N-1}, \quad (38)$$

$$\widehat{\mathbf{a}}_i = \mathbf{C}\widehat{\boldsymbol{\alpha}}_i + \mathbf{d} \succeq \mathbf{0}_M,\ \forall i \in \mathcal{I}_N. \quad (39)$$

Abundance vector $\mathbf{s}[n]$ estimation

Lemma 3 [Lin'16] (Closed-form Expression of $\mathbf{s}[n]$)

Assume that (A1) and (A2) hold true. Then, $\mathbf{s}[n] = [s_1[n], \ldots, s_N[n]]^T$ has the following closed-form expression:

$$s_i[n] = \frac{h_i - \mathbf{b}_i^T \tilde{\mathbf{x}}[n]}{h_i - \mathbf{b}_i^T \boldsymbol{\alpha}_i}, \quad \forall\, i \in \mathcal{I}_N,\ \forall\, n \in \mathcal{I}_L. \tag{40}$$

$$\Rightarrow\ \widehat{s}_i[n] = \max\left\{ \frac{\widehat{h}_i - \widehat{\mathbf{b}}_i^T \tilde{\mathbf{x}}[n]}{\widehat{h}_i - \widehat{\mathbf{b}}_i^T \widehat{\boldsymbol{\alpha}}_i}, 0 \right\}, \quad \forall\, i \in \mathcal{I}_N,\ \forall\, n \in \mathcal{I}_L. \tag{41}$$

Theorem 1 (Computational Complexity Analysis) [Lin'16]

The computational complexity of HyperCSI is $\mathcal{O}(NL)$ *with parallel processing* and $\mathcal{O}(NL^2)$ *without parallel processing*.

Statistical analysis

(A3) $\mathbf{s}[1], \ldots, \mathbf{s}[L]$ are independent and identically distributed (i.i.d.) with a continuous probability density function (p.d.f.) whose support is the unit simplex $\mathcal{T}_e \triangleq \mathrm{conv}\{\mathbf{e}_1, \ldots, \mathbf{e}_N\} \subseteq \mathbb{R}^N$ (e.g., Dirichlet distribution).

Theorem 2 (Affine Independence Analysis) [Lin'16]

Under (A1)-(A3), the pixels $\{\widehat{\mathbf{p}}_1^{(i)}, \ldots, \widehat{\mathbf{p}}_{N-1}^{(i)}\}$ extracted by (34) are $\mathcal{A}.\mathcal{I}.$ with probability 1 (w.p.1), $\forall i \in \mathcal{I}_N$.

Theorem 3 (Identifiability of HyperCSI Algorithm) [Lin'16]

Under (A1)-(A3), the noiseless assumption and $L \to \infty$, the simplex identified by the HyperCSI algorithm with $c = 1$ is exactly the Craig's simplex and the true endmembers' simplex $\mathrm{conv}\{\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_N\}$ in the DR space w.p.1.

Numerical simulaitons

- HyperCSI is compared with five state-of-the-art algorithms [Ma'14]:
  1. Minimum-volume simplex analysis (**MVSA**) algorithm;
  2. Interior-point method based MVSA (**ipMVSA**) algorithm [Li'15];
  3. Minimum-volume enclosing simplex (**MVES**) algorithm [Chan'09];
  4. Simplex identification via split augmented Lagrangian (**SISAL**) algorithm;
  5. Minimum-volume constrained non-negative matrix factorization (**MVC-NMF**) algorithm [Miao'07].

[Ma'14] W.-K. Ma et al., "A signal processing perspective on hyperspectral unmixing," *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67-81, Jan. 2014.

[Li'15] J. Li et al., "Minimum volume simplex analysis: A fast algorithm for linear hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 9, pp. 5067-5082, Apr. 2015.

[Chan'09] T.-H. Chan, et al., "A convex analysis-based minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Signal Processing*, vol. 57, no. 11, pp. 4418-4432, Nov. 2009. (*Citations:* **245** *by Google Scholar*)

[Miao'07] L. Miao et al., "Endmember extraction from highly mixed data using minimum volume constrained nonnegative matrix factorization," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 3, pp. 765-777, 2007.

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 ϕ_{en} (between $\mathbf{a}_i$ and $\widehat{\mathbf{a}}_i$):

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

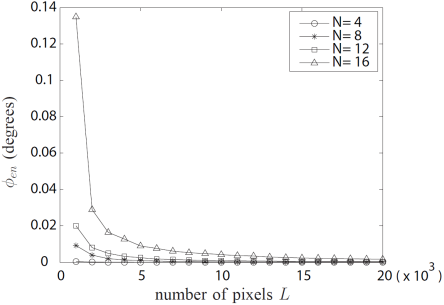
     where Π_N is the set of all the permutations of $\{1, \ldots, N\}$.

  3. RMS angle error ϕ_{ab} (between $\boldsymbol{s}_i$ and $\widehat{\boldsymbol{s}}_i$):

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

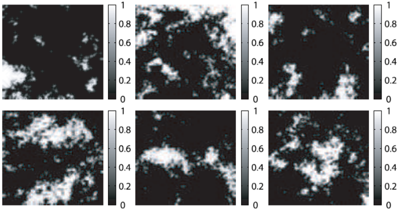
Numerical simulaitons (con't)

- The asymptotic analysis in Theorem 3 (noiseless and $c = 1$) is illustrated below:

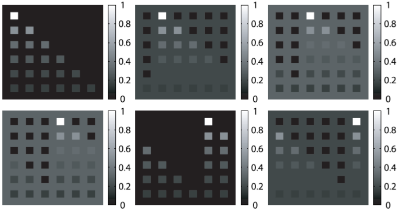


- HyperCSI performs well even with a moderately finite number of pixels L, i.e., several tens of thousands, which is typical in HU.
- The curve for larger N converges at a slower rate as we need more pixels (i.e., $N(N-1)$) lying close to the boundary hyperplanes for larger N.

Numerical simulaitons (con't)



(a) Ground truth abundance maps of SYN1



(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 [Iordache'12], denoted by SYN1 and SYN2, for performance evaluation, where SYN1 contains $L = 10,000$ pixels and SYN2 contains $L = 16,900$ pixels, resp. [Iordache'12].

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

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

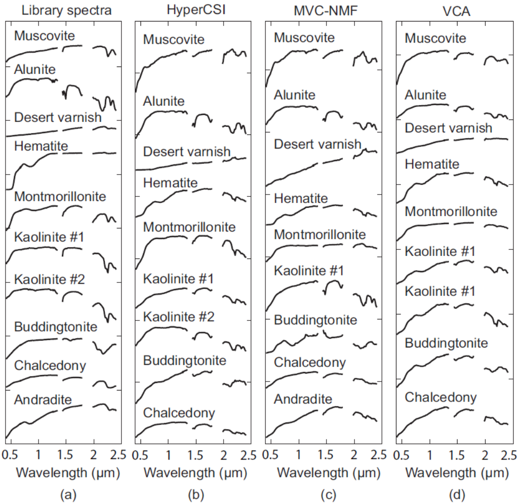
	Methods	ϕ_{en} (degrees)					ϕ_{ab} (degrees)					T (seconds)
		SNR (dB)					SNR (dB)					
		20	25	30	35	40	20	25	30	35	40	
SYN1	MVC-NMF	3.23	1.97	1.05	0.55	**0.25**	13.87	8.51	4.79	2.65	**1.34**	1.74E+2
	MVSA	10.65	6.12	3.38	1.88	1.05	22.93	15.13	9.34	5.52	3.19	3.53E+0
	MVES	9.55	5.49	3.60	1.96	1.22	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:
  1. MVC-NMF algorithm [Miao'07] (based on Craig's criterion);
  2. VCA algorithm [Nascimento'05] (based on pure-pixel assumption).

[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," submitted to *IEEE Trans. Neural Networks and Learning Systems* (acceptable subject to minor revision).
[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 μm.

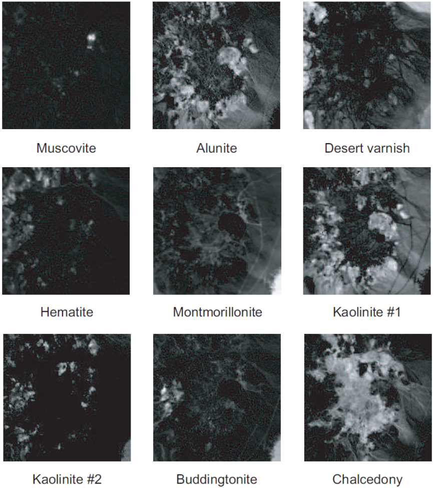
Real data experiments (con't)

- 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 ϕ (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)



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.

Summary

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

1. Craig's simplex is *reconstructed from $N(N-1)$ pixels* (regardless of the data length L), without involving any simplex volume computations.
2. It is *reproducible* (without involving random initialization and tuning of regularization parameters) and not data-dependent, regardless of the existence of pure pixels.
3. Its *superior performance* over state-of-the-art methods has been demonstrated *by analysis, simulations and real data experiments.*
4. 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.

Outline

Outline

1. **Part I: Fundamentals of Convex Optimization**
2. **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

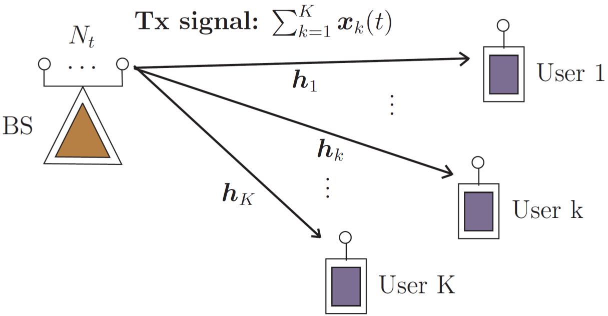
Outline (Subsection I)

1. System Model and Problem Statement
2. Convex Restriction Methods to the Rate Outage Constrained Problem: Bernstein-type Inequality
3. Simulation Results
4. Conclusions

Outline (Subsection I)

1. System Model and Problem Statement
2. Convex Restriction Methods to the Rate Outage Constrained Problem: Bernstein-Type Inequality
3. Simulation Results
4. Conclusions

- 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.



1. System Model and Problem Statement

- Transmit signal from BS:

$$\boldsymbol{x}(t) = \sum_{i=1}^{K} \boldsymbol{x}_i(t). \tag{42}$$

  - $\boldsymbol{x}_i(t) \in \mathbb{C}^{N_t}$: 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 $\mathrm{rank}(\boldsymbol{S}_i) = d$, $\boldsymbol{x}_i(t)$ can be expressed as [Vu07]

$$\boldsymbol{x}_i(t) = \sum_{k=1}^{d} \sqrt{\lambda_k(\boldsymbol{S}_i)}\boldsymbol{w}_k s_{ik}(t). \tag{43}$$

  - $\lambda_k(\boldsymbol{S}_i)$: the kth largest eigenvalue of $\boldsymbol{S}_i$;
  - $s_{ik}(t) \sim \mathcal{CN}(0, 1)$: 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 $\boldsymbol{x}_i(t)$ reduces to transmit beamforming.

[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:

$$y_i(t) = \boldsymbol{h}_i^H \boldsymbol{x}(t) + n_i(t). \tag{44}$$

  - $\boldsymbol{h}_i \in \mathbb{C}^{N_t}$: 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 $\boldsymbol{h}_i$ at the receiver i [Telatar99]:

$$\mathsf{R}_i(\{\boldsymbol{S}_k\}_{k=1}^K; \boldsymbol{h}_i) = \log_2\left(1 + \underbrace{\frac{\boldsymbol{h}_i^H \boldsymbol{S}_i \boldsymbol{h}_i}{\sum_{k\neq i}^K \boldsymbol{h}_i^H \boldsymbol{S}_k \boldsymbol{h}_i + \sigma_i^2}}_{\mathsf{SINR}}\right),\ i = 1, \ldots, K, \tag{45}$$

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

[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 $\boldsymbol{h}_1, \ldots, \boldsymbol{h}_K$ that are known to the BS,

$$\min_{\boldsymbol{S}_1,\ldots,\boldsymbol{S}_K \in \mathbb{H}^{N_t}} \sum_{i=1}^{K} \mathrm{Tr}(\boldsymbol{S}_i) \tag{46a}$$

$$\text{s.t. } \mathsf{R}_i(\{\boldsymbol{S}_k\}_{k=1}^{K}; \boldsymbol{h}_i) \geq r_i, \quad i = 1, \ldots, K, \tag{46b}$$

$$\boldsymbol{S}_1, \ldots, \boldsymbol{S}_K \succeq \boldsymbol{0}, \tag{46c}$$

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

- Problem (46) can be reformulated as a convex semidefinite program (SDP), which is polynomial-time solvable [Bengtsson01] [Gershman10].

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

[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 $\boldsymbol{h}_i$ (used in the conventional formulation in (46)) due to imperfect channel estimation and limited feedback [Love08].

- CSI error model:

$$\boldsymbol{h}_i = \bar{\boldsymbol{h}}_i + \boldsymbol{e}_i, \qquad i = 1, \ldots, K, \tag{47}$$

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

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

$$\boldsymbol{e}_i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{C}_i) \tag{48}$$

  for some known error covariance $\boldsymbol{C}_i \succeq \mathbf{0}$.

[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.
[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.
[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.

Effect of CSI errors on QoS requirement (information rate)

Rate outage probabilities:

$$\rho_i \triangleq \text{Prob}\{\mathsf{R}_i \leq r_i\}, \ i = 1, \ldots, K.$$

Consider a simulation example for $N_t = K = 3$, target rates $r_i = r$, $\rho_i \leq \rho = 0.1$ for all i, with the common target SINR value $\gamma \triangleq 2^r - 1 = 11$ dB (cf. (45)).

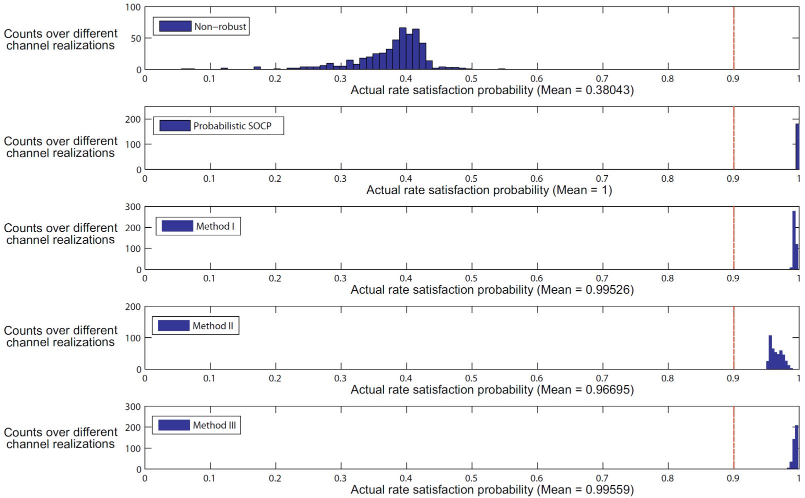
- 500 sets of $\{\bar{\boldsymbol{h}}_i\}_{i=1}^K$ are randomly generated with $\bar{\boldsymbol{h}}_i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{I}_{N_t})$;
- For each $\{\bar{\boldsymbol{h}}_k\}_{k=1}^K$,
  - obtain $\{\hat{\boldsymbol{S}}_k\}_{k=1}^K$ by solving (46) (as if $\{\bar{\boldsymbol{h}}_k\}_{k=1}^K$ were perfect), and using 4 robust designs, probabilistic SOCP [Shenouda08] and Methods I-III (to be presented later) with $\boldsymbol{C}_i = 0.002\boldsymbol{I}_{N_t}$;
  - generate $10,000$ Gaussian CSI errors $\boldsymbol{e}_i \sim \mathcal{CN}(\mathbf{0}, \boldsymbol{C}_i)$, $i = 1, \ldots, K$, and then obtain the actual $\hat{\rho}_i = N_i/10000$, where N_i denotes the no. of error realizations for which

$$\mathsf{R}_i(\{\hat{\boldsymbol{S}}_k\}_{k=1}^K; \bar{\boldsymbol{h}}_i + \boldsymbol{e}_i) \leq r, \ i = 1, \ldots, K.$$

[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.

Effect of CSI errors on QoS requirement (information rate)

- Histograms of actual satisfaction probabilities $\{1-\hat{\rho}_i\}_{i=1}^{K}$ over the 500 channel realizations:



1. System Model and Problem Statement

Rate Outage Constrained Problem

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

$$\min_{\boldsymbol{S}_1,\ldots,\boldsymbol{S}_K \in \mathbb{H}^{N_t}} \sum_{i=1}^{K} \mathrm{Tr}(\boldsymbol{S}_i) \tag{49a}$$

$$\text{s.t. } \mathrm{Prob}\left\{ \mathsf{R}_i(\{\boldsymbol{S}_k\}_{k=1}^{K}; \bar{\boldsymbol{h}}_i + \boldsymbol{e}_i) \leq r_i \right\} \leq \rho_i, \; i = 1, \ldots, K, \tag{49b}$$

$$\boldsymbol{S}_1, \ldots, \boldsymbol{S}_K \succeq \boldsymbol{0}, \tag{49c}$$

where $\mathsf{R}_i(\{\boldsymbol{S}_k\}_{k=1}^{K}, \bar{\boldsymbol{h}}_i + \boldsymbol{e}_i)$ is defined in (45).

- Problem (49) is hard to solve since rate outage probabilities in (49b) have no closed-form expressions and are unlikely to be efficiently computable in general.

Outline (Subsection I)

1. System Model and Problem Statement
2. Convex Restriction Methods to the Rate Outage Constrained Problem: Bernstein-Type Inequality
3. Simulation Results
4. Conclusions

- The rate outage constraints in (49b) can be expressed as

$$\text{Prob}\{\boldsymbol{e}_i^H \boldsymbol{Q}_i \boldsymbol{e}_i + 2\text{Re}\{\boldsymbol{e}_i^H \boldsymbol{r}_i\} + s_i < 0\} \leq \rho_i, \quad i = 1, \ldots, K, \tag{50}$$

where for notational simplicity, $\boldsymbol{e}_i \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{N_t})$ (originally denoting channel error), and

$$\boldsymbol{Q}_i = \boldsymbol{C}_i^{1/2} \left(\frac{1}{\gamma_i} \boldsymbol{S}_i - \sum_{k \neq i} \boldsymbol{S}_k \right) \boldsymbol{C}_i^{1/2}, \quad \boldsymbol{r}_i = \boldsymbol{C}_i^{1/2} \left(\frac{1}{\gamma_i} \boldsymbol{S}_i - \sum_{k \neq i} \boldsymbol{S}_k \right) \bar{\boldsymbol{h}}_i, \tag{51a}$$

$$s_i = \bar{\boldsymbol{h}}_i^H \left(\frac{1}{\gamma_i} \boldsymbol{S}_i - \sum_{k \neq i} \boldsymbol{S}_k \right) \bar{\boldsymbol{h}}_i - \sigma_i^2, \tag{51b}$$

in which

$$\gamma_i = 2^{r_i} - 1 \qquad \text{(target SINR for user } i \text{ (cf. (45)))}$$

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

Challenge 1

Let $\boldsymbol{e} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_n)$ and $(\boldsymbol{Q}, \boldsymbol{r}, s) \in \mathbb{H}^n \times \mathbb{C}^n \times \mathbb{R}$ be an arbitrary 3-tuple of (deterministic) variables. Find an *efficiently computable convex function* $f : \mathbb{H}^n \times \mathbb{C}^n \times \mathbb{R} \to \mathbb{R}$ such that

$$\text{Prob}\left\{\boldsymbol{e}^H\boldsymbol{Q}\boldsymbol{e} + 2\text{Re}\{\boldsymbol{e}^H\boldsymbol{r}\} + s < 0\right\} \leq f(\boldsymbol{Q}, \boldsymbol{r}, s). \qquad (52)$$

- Once a function f satisfying (52) is found,

$$f(\boldsymbol{Q}, \boldsymbol{r}, s) \leq \rho \qquad (53)$$
$$\Longrightarrow \text{Prob}\{\boldsymbol{e}^H\boldsymbol{Q}\boldsymbol{e} + 2\text{Re}\{\boldsymbol{e}^H\boldsymbol{r}\} + s < 0\} \leq \rho. \qquad (54)$$

- Constraint (53) gives a *convex restriction* of constraint (54), because the associated $\boldsymbol{Q}_i, \boldsymbol{r}_i, s_i$ defined by (51) are affine in $\boldsymbol{S}_1, \ldots, \boldsymbol{S}_K$, implying that both f and (53) are convex with respect to decision variable $\boldsymbol{S}_i$.

- Alternative expression for the outage probability constraint in (54):

$$\text{Prob}\{\boldsymbol{e}^H\boldsymbol{Q}\boldsymbol{e} + 2\text{Re}\{\boldsymbol{e}^H\boldsymbol{r}\} + s \geq 0\} \geq 1 - \rho.$$

Lemma 2 (Bernstein-Type Inequality) [Bechar09]

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

$$\text{Prob}\left\{\boldsymbol{e}^H\boldsymbol{Q}\boldsymbol{e} + 2\text{Re}\{\boldsymbol{e}^H\boldsymbol{r}\} \geq \Upsilon(\eta)\right\} \geq 1 - e^{-\eta}, \tag{55}$$

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

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

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

[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.

2. Convex Restriction Methods (Bernstein-Type Inequality)

- A sufficient condition for achieving the outage constraint (54):

$$f(\boldsymbol{Q}, \boldsymbol{r}, s) = e^{-\Upsilon^{-1}(-s)} \leq \rho,$$

or equivalently $\Upsilon(\ln(1/\rho)) \geq -s$, i.e.,

$$\mathrm{Tr}(\boldsymbol{Q}) - \sqrt{2\ln(1/\rho)}\sqrt{\|\boldsymbol{Q}\|_F^2 + 2\|\boldsymbol{r}\|_2^2} + \ln(\rho) \cdot \lambda^+(\boldsymbol{Q}) + s \geq 0. \quad (56)$$

Method II (Bernstein-Type Inequality)

A convex restriction approximation of problem (49):

$$\min_{\substack{\boldsymbol{S}_i \in \mathbb{H}^{N_t}, x_i, y_i \in \mathbb{R}, \\ i=1,\ldots,K}} \sum_{i=1}^{K} \mathrm{Tr}(\boldsymbol{S}_i) \tag{57a}$$

$$\text{s.t. } \mathrm{Tr}(\boldsymbol{Q}_i) - \sqrt{2\ln(1/\rho_i)} \cdot x_i + \ln(\rho_i) \cdot y_i + s_i \geq 0, \ \forall i, \tag{57b}$$

$$\left\| \begin{bmatrix} \mathbf{vec}(\boldsymbol{Q}_i) \\ \sqrt{2}\boldsymbol{r}_i \end{bmatrix} \right\|_2 \leq x_i, \ i = 1, \ldots, K, \tag{57c}$$

$$y_i \boldsymbol{I}_{N_t} + \boldsymbol{Q}_i \succeq \mathbf{0}, \ i = 1, \ldots, K, \tag{57d}$$

$$y_1, \ldots, y_K \geq 0, \ \boldsymbol{S}_1, \ldots, \boldsymbol{S}_K \succeq \mathbf{0}, \tag{57e}$$

where $\boldsymbol{Q}_i$, $\boldsymbol{r}_i$ and s_i are defined in (51), $i = 1, \ldots, K$.

Outline (Subsection I)

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

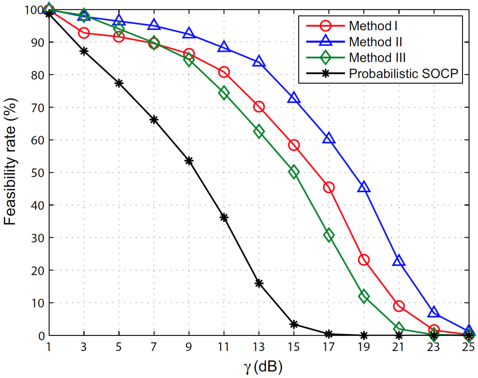
- $N_t = K = 3$;
- Users' noise powers: $\sigma_1^2 = \cdots = \sigma_K^2 = 0.1$;
- Preset outage probabilities: $\rho_1 = \cdots = \rho_K = 0.1$;
- SINR requirements: $\gamma_1 = \cdots = \gamma_K \triangleq \gamma$ (recall that $\gamma_i = 2^{r_i} - 1$);
- Spatially i.i.d. Gaussian CSI errors $\boldsymbol{C}_1 = \cdots = \boldsymbol{C}_K = 0.002\boldsymbol{I}_{N_t}$;
- In each simulation, 500 sets of the presumed channels $\{\bar{\boldsymbol{h}}_i\}_{i=1}^K$ are randomly and independently generated with $\bar{\boldsymbol{h}}_i \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{N_t})$;
- Performance comparisons with probabilistic SOCP [Shenouda08].

[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.

Feasibility rate of the various methods, where Method II is Bernstein inequality based method.

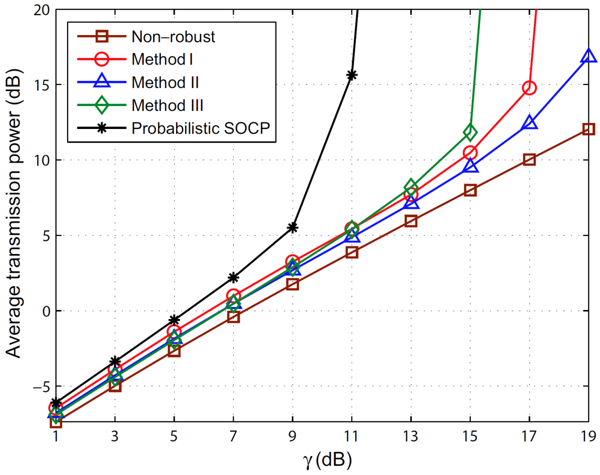
Feasibility rate $\triangleq \frac{\text{no. of feasible channels}}{\text{no. of tested channels}} \times 100\%$;

The lower the feasibility rate, the more conservative *(implying higher transmit power in general)* for the method under test.



3. Simulation Results

Transmit power performance of the various methods.



3. Simulation Results

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

Table 1: Ratios of rank-one solutions.

ρ	0.1			
γ (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

ρ	0.01			
γ (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

Outline (Subsection I)

1. System Model and Problem Statement
2. Convex Restriction Methods to the Rate Outage Constrained Problem: Bernstein-Type Inequality
3. Simulation Results
4. Conclusions

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.

[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:* **114** *by Google Scholar*)

Outline

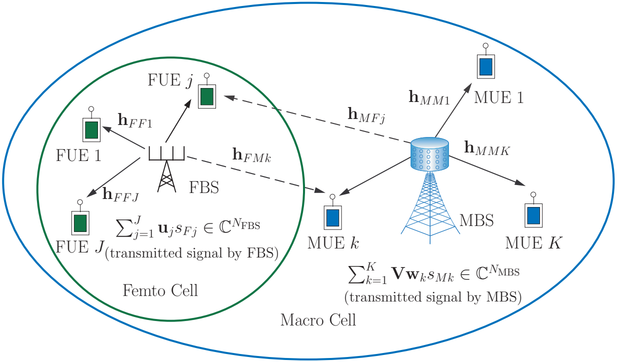
1. **Part I: Fundamentals of Convex Optimization**
2. **Part II: Application in Hyperspectral Image Analysis: (Big Data Analysis and Machine Learning)**
3. **Part III: Application in Wireless Communications (5G Systems)**
   - **Subsection I:** Outage Constrained Robust Transmit Optimization for Multiuser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

Outline (Subsection II)

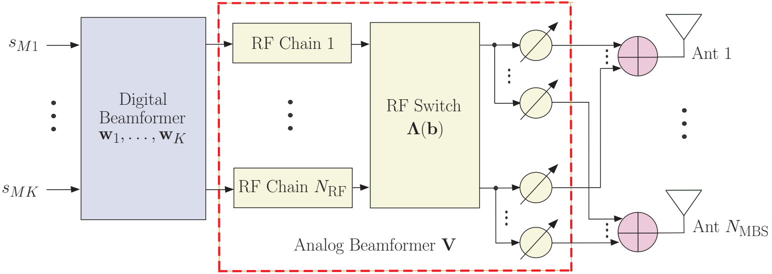
1. System Model and Problem Formulation
2. Proposed Outage Constrained HyCoBF Design
    1. Analog Beamforming Algorithm
    2. Digital Robust CoBF Design
3. Simulation Results
4. Conclusions

Outline (Subsection II)

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



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



[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.

1. System Model and Problem Formulation

- Let $\mathbf{F} \in \mathbb{C}^{N_{\text{MBS}} \times N_{\text{MBS}}}$ be the N_{MBS}-point discrete Fourier transform (DFT) matrix (codebook), and

$$\mathcal{B} \triangleq \{\mathbf{b} \in \{1, \ldots, N_{\text{MBS}}\}^{N_{\text{RF}}} \mid b_i \neq b_j,\ \forall i \neq j\}, \tag{58}$$

collecting all the $\binom{N_{\text{MBS}}}{N_{\text{RF}}}$ different combinations.

- The selected analog beamforming matrix $\mathbf{V}$, consisting of N_{RF} distinct columns of $\mathbf{F}$, and the RF switch matrix $\mathbf{\Lambda}(\mathbf{b})$ can be expressed as

$$\begin{aligned} &\mathbf{\Lambda}(\mathbf{b}) = [\boldsymbol{e}(b_1), \ldots, \boldsymbol{e}(b_{N_{\text{RF}}})] \in \mathbb{R}^{N_{\text{MBS}} \times N_{\text{RF}}},\ \mathbf{b} \in \mathcal{B}, \\ &\mathbf{V} = \mathbf{F}\mathbf{\Lambda}(\mathbf{b}) \in \mathbb{C}^{N_{\text{MBS}} \times N_{\text{RF}}}, \end{aligned} \tag{59}$$

where $\boldsymbol{e}(b_i)$ denotes the b_i-th unit column vector.

- Received signal of MUE k:

$$y_{Mk} = \mathbf{h}_{MMk}^H \mathbf{V}\mathbf{w}_k s_{Mk} + \sum_{l=1, l\neq k}^{K} \mathbf{h}_{MMk}^H \mathbf{V}\mathbf{w}_l s_{Ml} + \sum_{j=1}^{J} \mathbf{h}_{FMk}^H \mathbf{u}_j s_{Fj} + n_{Mk}. \quad (60)$$

  - s_{Mk}, s_{Fj} : transmit signals with unit power intended for MUE k and FUE j, respectively; $n_{Mk} \sim \mathcal{CN}(0, \sigma_{Mk}^2)$ (AWGN);
  - $\mathbf{h}_{MMk} \in \mathbb{C}^{N_{\mathrm{MBS}}}$, $\mathbf{h}_{FMk} \in \mathbb{C}^{N_{\mathrm{FBS}}}$: channels from MBS and FBS to MUE k, respectively;
  - $\mathbf{w}_k \in \mathbb{C}^{N_{\mathrm{RF}}}$, $\mathbf{u}_j \in \mathbb{C}^{N_{\mathrm{FBS}}}$: baseband beamforming vectors for MUE k and FUE j, respectively.

- *Signal-to-interference-plus-noise ratio* (SINR) of MUE k:

$$\mathrm{SINR}_{Mk} = \frac{|\mathbf{h}_{MMk}^H \mathbf{V}\mathbf{w}_k|^2}{\sum_{l=1, l\neq k}^{K} |\mathbf{h}_{MMk}^H \mathbf{V}\mathbf{w}_l|^2 + \sum_{j=1}^{J} |\mathbf{h}_{FMk}^H \mathbf{u}_j|^2 + \sigma_{Mk}^2}. \quad (61)$$

- Received signal of FUE j:

$$y_{Fj} = \mathbf{h}_{FFj}^H \mathbf{u}_j s_{Fj} + \sum_{m=1, m \neq j}^{J} \mathbf{h}_{FFj}^H \mathbf{u}_m s_{Fm} + \sum_{k=1}^{K} \mathbf{h}_{MFj}^H \mathbf{V} \mathbf{w}_k s_{Mk} + n_{Fj}. \quad (62)$$

  - $\mathbf{h}_{MFj} \in \mathbb{C}^{N_{\text{MBS}}}$, $\mathbf{h}_{FFj} \in \mathbb{C}^{N_{\text{FBS}}}$: channels from MBS and FBS to FUE j, respectively; $n_{Fj} \sim \mathcal{CN}(0, \sigma_{Fj}^2)$.

- SINR of FUE j:

$$\text{SINR}_{Fj} = \frac{|\mathbf{h}_{FFj}^H \mathbf{u}_j|^2}{\sum_{m=1, m \neq j}^{J} |\mathbf{h}_{FFj}^H \mathbf{u}_m|^2 + \sum_{k=1}^{K} |\mathbf{h}_{MFj}^H \mathbf{V} \mathbf{w}_k|^2 + \sigma_{Fj}^2}. \quad (63)$$

- Imperfect CSI model [Wang'14]:

$$\mathbf{h}_{MMk} = \hat{\mathbf{h}}_{MMk} + \mathbf{e}_{MMk}, \ \mathbf{h}_{MFj} = \hat{\mathbf{h}}_{MFj} + \mathbf{e}_{MFj}, \tag{64a}$$

$$\mathbf{h}_{FFj} = \hat{\mathbf{h}}_{FFj} + \mathbf{e}_{FFj}, \ \mathbf{h}_{FMk} = \hat{\mathbf{h}}_{FMk} + \mathbf{e}_{FMk}. \tag{64b}$$

  - $\hat{\mathbf{h}}_{MMk}, \hat{\mathbf{h}}_{MFj} \in \mathbb{C}^{N_{\mathrm{MBS}}}$, and $\hat{\mathbf{h}}_{FFj}, \hat{\mathbf{h}}_{FMk} \in \mathbb{C}^{N_{\mathrm{FBS}}}$: channel estimates that are known to MBS and FBS;
  - $\mathbf{e}_{MMk}, \mathbf{e}_{MFj} \in \mathbb{C}^{N_{\mathrm{MBS}}}$, $\mathbf{e}_{FFj}, \mathbf{e}_{FMk} \in \mathbb{C}^{N_{\mathrm{FBS}}}$: Gaussian CSI errors

$$\mathbf{e}_{MMk} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{C}_{MMk}\right), \ \mathbf{e}_{MFj} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{C}_{MFj}\right), \tag{65a}$$

$$\mathbf{e}_{FFj} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{C}_{FFj}\right), \ \mathbf{e}_{FMk} \sim \mathcal{CN}\left(\mathbf{0}, \mathbf{C}_{FMk}\right), \tag{65b}$$

where $\mathbf{C}_{MMk} \succeq \mathbf{0}$, $\mathbf{C}_{MFj} \succeq \mathbf{0}$ and $\mathbf{C}_{FFj} \succeq \mathbf{0}$, $\mathbf{C}_{FMk} \succeq \mathbf{0}$.

[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, November 2014.

Outage Constrained Robust HyCoBF Problem

Given γ_{Mk} and γ_{Fj} for the target SINRs of MUEs and FUEs, respectively, and the associated outage probabilities ρ_{Mk} and ρ_{Fj},

$$\min_{\mathbf{V},\{\mathbf{w}_k\},\{\mathbf{u}_j\}} \sum_{k\in\mathcal{I}_K} \|\mathbf{V}\mathbf{w}_k\|^2 + \sum_{j\in\mathcal{I}_J} \|\mathbf{u}_j\|^2 \tag{66a}$$

$$\text{s.t.} \quad \Pr(\text{SINR}_{Mk} \geq \gamma_{Mk}) \geq 1-\rho_{Mk},\ \forall k \in \mathcal{I}_K = \{1,\ldots,K\} \tag{66b}$$

$$\Pr(\text{SINR}_{Fj} \geq \gamma_{Fj}) \geq 1-\rho_{Fj},\ \forall j \in \mathcal{I}_J = \{1,\ldots,J\} \tag{66c}$$

$$\mathbf{V} \in \{\mathbf{F}\boldsymbol{\Lambda}(\mathbf{b}) \mid \mathbf{b} \in \mathcal{B}\}\ (\text{cf. (2)}). \tag{66d}$$

- Solving the robust HyCoBF design problem (66) is hard, due to nonconvex probabilistic constraints (66b), (66c) [Nemirovski'06] and analog beam selection constraint (66d).

[Nemirovski'06] A. Nemirovski and A. Shapiro, "Convex approximations of chance constrained programs," *SIAM J. Optim.*, vol. 17, no. 4, pp. 969-996, 2006.

Outline (Subsection II)

- A new beam selection criterion named power ratio maximization (PRM), by maximizing the ratio of the total channel power of MUEs to the total intercell interference channel power from MBS to FUEs:

$$\mathbf{V}^\star = \mathbf{F}\boldsymbol{\Lambda}(\mathbf{b}^\star) \quad \text{(cf. (59))}, \tag{67}$$

$$\mathbf{b}^\star = \arg\max_{\mathbf{b}\in\mathcal{B}} \left\{ \mathcal{J}(\mathbf{b}) \triangleq \frac{\|\widehat{\mathbf{H}}_{MM}^H \mathbf{F}\boldsymbol{\Lambda}(\mathbf{b})\|_F^2}{\|\widehat{\mathbf{H}}_{MF}^H \mathbf{F}\boldsymbol{\Lambda}(\mathbf{b})\|_F^2} \right\}, \quad \text{(PRM criterion)} \tag{68}$$

  where $\mathcal{B}$ was defined in (58), $\widehat{\mathbf{H}}_{MM} \triangleq [\hat{\mathbf{h}}_{MM1}, \ldots, \hat{\mathbf{h}}_{MMK}]$ and $\widehat{\mathbf{H}}_{MF} \triangleq [\hat{\mathbf{h}}_{MF1}, \ldots, \hat{\mathbf{h}}_{MFJ}]$.
- A low-complexity beam selection algorithm for solving (68) is proposed, motivated by the idea of single most likely replacement (SMLR) detector used in seismic deconvolution for detecting a Bernoulli-Gaussian signal with nonzero magnitudes [Mendel'83].

[Mendel'83] J. M. Mendel, "Optimal seismic deconvolution: An estimated-based approach," *Oval Road, London, UK: Academic Press, Inc.*, 1983.

2.1. Analog Beamforming Algorithm

- Low complexity and guaranteed convergence: Computing $\mathcal{J}(\mathbf{b})$ by (68) $N_{\mathrm{RF}} \times (N_{\mathrm{MBS}} - N_{\mathrm{RF}})$ times per iteration.

Beam Selection Algorithm-PRM

1. Given $\widehat{\mathbf{H}}_{MM}$, $\widehat{\mathbf{H}}_{MF}$, $\mathcal{I}_{N_{\mathrm{MBS}}} = \{1, \ldots, N_{\mathrm{MBS}}\}$;
2. Initialize $\mathbf{b} = (b_1, \ldots, b_{N_{\mathrm{RF}}}) \in \mathcal{B}$, and $\mathcal{S} = \mathcal{I}_{N_{\mathrm{MBS}}} \setminus \{b_1, \ldots, b_{N_{\mathrm{RF}}}\}$;
3. Compute $\mathcal{J}(\mathbf{b})$ by (68);
4. **repeat**
5. Obtain $j_\ell = \arg\max_{j\in\mathcal{S}} \mathcal{J}(\mathbf{b}_\ell(j))$, $\ell = 1, \ldots, N_{\mathsf{RF}}$, where $\mathbf{b}_\ell(j) \triangleq (b_1, \ldots, b_{\ell-1}, j, b_{\ell+1}, \ldots, b_{N_{\mathrm{RF}}})$;
6. Obtain $l = \arg\max\{\mathcal{J}(\mathbf{b}_\ell(j_\ell)),\ \ell = 1, \ldots, N_{\mathsf{RF}}\}$;
7. If $\mathcal{J}(\mathbf{b}_l(j_l)) > \mathcal{J}(\mathbf{b})$, update $\mathbf{b} := \mathbf{b}_l(j_l)$, $\mathcal{S} := \mathcal{I}_{N_{\mathrm{MBS}}} \setminus \{b_1, \ldots, b_{N_{\mathrm{RF}}}\}$, $\mathcal{J}(\mathbf{b}) := \mathcal{J}(\mathbf{b}_l(j_l))$;
8. **until** $\mathcal{J}(\mathbf{b}_l(j_l)) \leq \mathcal{J}(\mathbf{b})$.
9. Output selected analog beamforming matrix $\mathbf{V}^\star = \mathbf{F}\mathbf{\Lambda}(\mathbf{b})$.

Outline (Subsection II)

1. System Model and Problem Formulation
2. Proposed Outage Constrained HyCoBF Design
    1. Analog Beamforming Algorithm
    2. Digital Robust CoBF Design
3. Simulation Results
4. Conclusions

2.2. Conservative Approximate CoBF Solution

- Let $\mathbf{W}_k = \mathbf{w}_k \mathbf{w}_k^H$ and $\mathbf{U}_j = \mathbf{u}_j \mathbf{u}_j^H$, and define

$$\mathbf{B}_k \triangleq \gamma_{Mk}^{-1} \mathbf{V}^\star \mathbf{W}_k (\mathbf{V}^\star)^H - \sum_{l \neq k}^{K} \mathbf{V}^\star \mathbf{W}_l (\mathbf{V}^\star)^H; \;\; \mathbf{D} \triangleq -\sum_{j=1}^{J} \mathbf{U}_j; \tag{69a}$$

$$\mathbf{F}_j \triangleq \gamma_{Fj}^{-1} \mathbf{U}_j - \sum_{l \neq j}^{J} \mathbf{U}_l; \;\; \mathbf{G} \triangleq -\sum_{k=1}^{K} \mathbf{V}^\star \mathbf{W}_k (\mathbf{V}^\star)^H; \tag{69b}$$

$$\mathbf{Q}_{MMk} \triangleq \mathbf{C}_{MMk}^{1/2} \mathbf{B}_k \mathbf{C}_{MMk}^{1/2}; \;\; \mathbf{Q}_{FMk} \triangleq \mathbf{C}_{FMk}^{1/2} \mathbf{D} \mathbf{C}_{FMk}^{1/2}; \tag{69c}$$

$$\mathbf{Q}_{FFj} \triangleq \mathbf{C}_{FFj}^{1/2} \mathbf{F}_j \mathbf{C}_{FFj}^{1/2}; \;\; \mathbf{Q}_{MFj} \triangleq \mathbf{C}_{MFj}^{1/2} \mathbf{G} \mathbf{C}_{MFj}^{1/2}; \tag{69d}$$

$$\mathbf{r}_{MMk} \triangleq \mathbf{C}_{MMk}^{1/2} \mathbf{B}_k \hat{\mathbf{h}}_{MMk}; \;\; \mathbf{r}_{FMk} \triangleq \mathbf{C}_{FMk}^{1/2} \mathbf{D} \hat{\mathbf{h}}_{FMk}; \tag{69e}$$

$$\mathbf{r}_{FFj} \triangleq \mathbf{C}_{FFj}^{1/2} \mathbf{F}_j \hat{\mathbf{h}}_{FFj}; \;\; \mathbf{r}_{MFj} \triangleq \mathbf{C}_{MFj}^{1/2} \mathbf{G} \hat{\mathbf{h}}_{MFj}; \tag{69f}$$

$$c_{Mk} \triangleq \hat{\mathbf{h}}_{MMk}^H \mathbf{B}_k \hat{\mathbf{h}}_{MMk} + \hat{\mathbf{h}}_{FMk}^H \mathbf{D} \hat{\mathbf{h}}_{FMk} - \sigma_{Mk}^2; \tag{69g}$$

$$c_{Fj} \triangleq \hat{\mathbf{h}}_{FFj}^H \mathbf{F}_j \hat{\mathbf{h}}_{FFj} + \hat{\mathbf{h}}_{MFj}^H \mathbf{G} \hat{\mathbf{h}}_{MFj} - \sigma_{Fj}^2. \tag{69h}$$

Outage Constrained Digital CoBF: Conservative Approximation

Applying semidefinite relaxation (SDR) (i.e., replacing $\mathbf{w}_k\mathbf{w}_k^H$ by $\mathbf{W}_k \succeq \mathbf{0}$ and $\mathbf{u}_j\mathbf{u}_j^H$ by $\mathbf{U}_j \succeq \mathbf{0}$) to problem (66) yields

$$\min_{\{\mathbf{W}_k\},\{\mathbf{U}_j\}} \sum_{k\in\mathcal{I}_K} \mathrm{Tr}\left(\mathbf{V}^\star\mathbf{W}_k(\mathbf{V}^\star)^H\right) + \sum_{j\in\mathcal{I}_J} \mathrm{Tr}\left(\mathbf{U}_j\right) \tag{70a}$$

$$\text{s.t.}\ \Pr\Big\{\boldsymbol{\delta}_1^H\mathbf{Q}_{MMk}\boldsymbol{\delta}_1 + \boldsymbol{\delta}_2^H\mathbf{Q}_{FMk}\boldsymbol{\delta}_2 + 2\mathrm{Re}\{\boldsymbol{\delta}_1^H\mathbf{r}_{MMk}\} + 2\mathrm{Re}\{\boldsymbol{\delta}_2^H\mathbf{r}_{FMk}\} + c_{Mk} \geq 0\Big\} \geq 1-\rho_{Mk},\ \ \forall k \in \mathcal{I}_K, \tag{70b}$$

$$\Pr\Big\{\boldsymbol{\delta}_1^H\mathbf{Q}_{MFj}\boldsymbol{\delta}_1 + \boldsymbol{\delta}_2^H\mathbf{Q}_{FFj}\boldsymbol{\delta}_2 + 2\mathrm{Re}\{\boldsymbol{\delta}_1^H\mathbf{r}_{MFj}\} + 2\mathrm{Re}\{\boldsymbol{\delta}_2^H\mathbf{r}_{FFj}\} + c_{Fj} \geq 0\Big\} \geq 1-\rho_{Fj},\ \ \forall j \in \mathcal{I}_J, \tag{70c}$$

$$\mathbf{W}_k \succeq \mathbf{0},\ \forall k \in \mathcal{I}_K,\ \mathbf{U}_j \succeq \mathbf{0},\ \forall j \in \mathcal{I}_J. \tag{70d}$$

where $\boldsymbol{\delta}_1 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_{\mathrm{MBS}}})$, $\boldsymbol{\delta}_2 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_{\mathrm{FBS}}})$.

- Define

$$g_1(\boldsymbol{\delta}_1, \mathbf{Q}_{MMk}, \mathbf{r}_{MMk}) \triangleq \boldsymbol{\delta}_1^H \mathbf{Q}_{MMk} \boldsymbol{\delta}_1 + 2\mathrm{Re}\{\mathbf{r}_{MMk}^H \boldsymbol{\delta}_1\}, \quad (71)$$

$$g_2(\boldsymbol{\delta}_2, \mathbf{Q}_{FMk}, \mathbf{r}_{FMk}) \triangleq \boldsymbol{\delta}_2^H \mathbf{Q}_{FMk} \boldsymbol{\delta}_2 + 2\mathrm{Re}\{\mathbf{r}_{FMk}^H \boldsymbol{\delta}_2\}. \quad (72)$$

- Alternative expression for the outage probability constraint in (70b):

$$\Pr\Big\{g_1(\boldsymbol{\delta}_1, \mathbf{Q}_{MMk}, \mathbf{r}_{MMk}) + g_2(\boldsymbol{\delta}_2, \mathbf{Q}_{FMk}, \mathbf{r}_{FMk}) + c_{Mk} \geq 0\Big\} \geq 1 - \rho_{Mk}. \quad (73)$$

- (70c) can also be equivalently re-expressed as the same form as (73).
- An extension form of the Bernstein-type inequality [Bechar'09] is presented in Lemma 1 below for finding a conservative convex approximation to (73).

[Bechar'09] I. Bechar, "A Bernstein-type inequality for stochastic processes of quadratic forms of Gaussian variables," http://arxiv.org/abs/0909.3595, Apr. 2009.

2.2. Conservative Approximate CoBF Solution

Lemma 1: Extension Form of the Bernstein-type Inequality

Given $\boldsymbol{\delta}_1 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_{\text{MBS}}})$, $\boldsymbol{\delta}_2 \sim \mathcal{CN}(\mathbf{0}, \mathbf{I}_{N_{\text{FBS}}})$, $\mathbf{Q}_{MMk} \in \mathbb{H}^{N_{\text{MBS}}}$, $\mathbf{Q}_{FMk} \in \mathbb{H}^{N_{\text{FBS}}}$, $\mathbf{r}_{FMk} \in \mathbb{C}^{N_{\text{FBS}}}$ Then, the following inequality holds:

$$\Pr\Big\{g_1(\boldsymbol{\delta}_1, \mathbf{Q}_{MMk}, \mathbf{r}_{MMk}) + g_2(\boldsymbol{\delta}_2, \mathbf{Q}_{FMk}, \mathbf{r}_{FMk}) \geq \Upsilon(\ln(1/\rho_{Mk}) \mid \mathbf{Q}_{MMk}, \mathbf{r}_{MMk}, \mathbf{Q}_{FMk}, \mathbf{r}_{FMk})\Big\} \geq 1 - \rho_{Mk}, \ \forall k \in \mathcal{I}_K, \quad (74)$$

where $\Upsilon : \mathbb{R}_{++} \to \mathbb{R}$ is defined as

$$\Upsilon(\ln(1/\rho_{Mk}) \mid \mathbf{Q}_{MMk}, \mathbf{r}_{MMk}, \mathbf{Q}_{FMk}, \mathbf{r}_{FMk}) \triangleq \mathrm{Tr}(\mathbf{Q}_{MMk}) + \mathrm{Tr}(\mathbf{Q}_{FMk}) - \ln(1/\rho_{Mk}) \cdot \lambda^{+}(\mathbf{Q}_{MMk}, \mathbf{Q}_{FMk}) - \alpha_{Mk}\sqrt{\|\mathbf{Q}_{MMk}\|_F^2 + 2\|\mathbf{r}_{MMk}\|^2 + \|\mathbf{Q}_{FMk}\|_F^2 + 2\|\mathbf{r}_{FMk}\|^2}, \quad (75)$$

in which $\lambda^{+}(\mathbf{Q}_{MMk}, \mathbf{Q}_{FMk}) \triangleq \max\{\lambda_{\max}(-\mathbf{Q}_{MMk}), \lambda_{\max}(-\mathbf{Q}_{FMk}), 0\}$ and $\alpha_{Mk} = \sqrt{2\ln(1/\rho_{Mk})}$.

By Lemma 1, $\Upsilon(\ln(1/\rho_{Mk}) \mid \mathbf{Q}_{MMk}, \mathbf{r}_{MMk}, \mathbf{Q}_{FMk}, \mathbf{r}_{FMk}) + c_{Mk} \geq 0$ is an *approximate deterministic conservative convex constraint* to (70b).

Robust Digital CoBF Problem (Centralized Solution)

By Lemma 1, problem (70) can be approximated as a convex semidefinite programming (SDP) [Luo'10]:

$$\min_{\substack{\{\mathbf{W}_k\},\{\mathbf{U}_j\},\\ \mathbf{t}_M,\mathbf{t}_F}} \sum_{k\in\mathcal{I}_K} \mathrm{Tr}\left(\mathbf{V}^\star\mathbf{W}_k(\mathbf{V}^\star)^H\right) + \sum_{j\in\mathcal{I}_J} \mathrm{Tr}\left(\mathbf{U}_j\right) \tag{76a}$$

$$\text{s.t.} \quad (\{\mathbf{W}_k\},\{\mathbf{U}_j\},\mathbf{t}_M) \in \mathcal{C}_M, \tag{76b}$$

$$(\{\mathbf{W}_k\},\{\mathbf{U}_j\},\mathbf{t}_F) \in \mathcal{C}_F, \tag{76c}$$

$$\mathbf{W}_k \succeq \mathbf{0},\ \forall k\in\mathcal{I}_K,\ \mathbf{U}_j \succeq \mathbf{0},\ \forall j\in\mathcal{I}_J, \tag{76d}$$

- $\mathcal{C}_M$ and $\mathcal{C}_F$ are approximate conservative convex constraint sets to (70b) and (70c) (cf. (78) and (79) in Appendix A), respectively;
- $\mathbf{t}_M \in \mathbb{R}^{3K}$, $\mathbf{t}_F \in \mathbb{R}^{3J}$ are auxiliary variables.

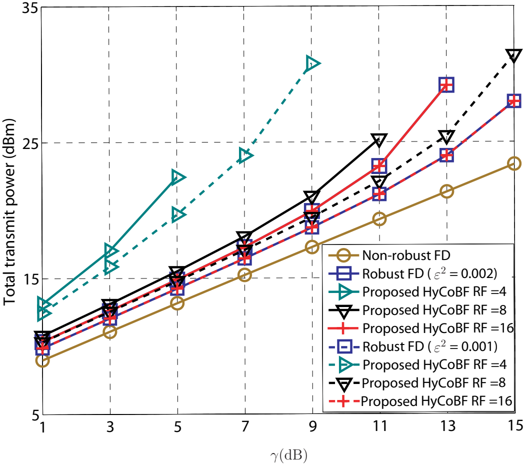
[Luo'10] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, "Semidefinite relaxation of quadratic optimization problems," *IEEE Signal Process. Mag.*, vol. 27, pp. 20-34, May 2010.

Outline (Subsection II)

Simulation Setting:

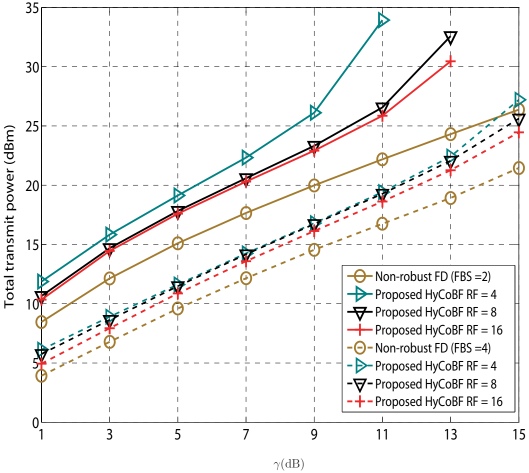
- Users' AWGN powers: $\sigma_k^2 = \sigma_j^2 = \sigma^2, \; \forall k \in \mathcal{I}_K, \; \forall j \in \mathcal{I}_J$;
- Target SINRs: $\gamma_{Mk} = \gamma_{Fj} = \gamma, \forall k, j$;
- SINR outage probabilities: $\rho_{Mk} = \rho_{Fj} = \rho$;
- CSI error covariance matrices: $\mathbf{C}_{MMk} = \mathbf{C}_{MFj} = \varepsilon^2 \mathbf{I}_{N_{\text{MBS}}}$, $\mathbf{C}_{FFj} = \mathbf{C}_{FMk} = \varepsilon^2 \mathbf{I}_{N_{\text{FBS}}}$;
- 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 $N_{\text{MBS}} = N_{\text{RF}}$.

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



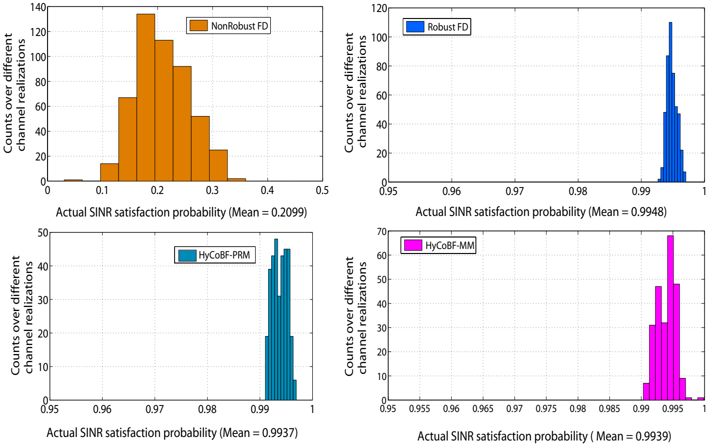
3. Simulation Results (Power Performance)

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



3. Simulation Results (Satisfaction Probability)

- $N_{\mathrm{MBS}} = 16$, $N_{\mathrm{RF}} = K = 4$; $N_{\mathrm{FBS}} = J = 2$; $\rho = 0.1$, $\gamma = 9$ dB and $\varepsilon^2 = 0.002$.



- In simulation, a "rank-one solution" for $\mathbf{W}_k^\star$ and $\mathbf{U}_j^\star$ is obtained if the following conditions hold:

$$\frac{\lambda_{\max}\left(\mathbf{W}_k^\star\right)}{\mathrm{Tr}\left(\mathbf{W}_k^\star\right)} \geq 0.9999, \frac{\lambda_{\max}\left(\mathbf{U}_j^\star\right)}{\mathrm{Tr}\left(\mathbf{U}_j^\star\right)} \geq 0.9999, k \in \mathcal{I}_{\mathcal{K}}, j \in \mathcal{I}_{\mathcal{J}}. \quad (77)$$

- 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.

ε^2	0.01			
γ (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)

ε^2	0.002			
γ (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)

Outline (Subsection II)

4. Conclusions

- We have introduced an *outage probability constrained robust HyCoBF design* for a massive MIMO enabled HetNet in the presence of Gaussian CSI errors, by solving a nonconvex total transmit power minimization problem.
- The robust HyCoBF design consists of:
  - The analog beamforming at MBS is a *newly devised low-complexity beam selection scheme by maximizing the ratio of MUEs' channel power to FUEs' interference channel power*, given a DFT matrix codebook.
  - The digital beamforming at MBS and FBS is a *conservative approximate CoBF solution* (by SDR technique and an extended Bernstein-type inequality).

- 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].

[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," *accepted as a regular paper by IEEE Access*.

Appdendix A

The convex constraint by using Lemma 1 for (70b) can be expressed in the following form:

$$\mathcal{C}_M \triangleq \Big\{ \big(\{\mathbf{W}_k\}, \{\mathbf{U}_j\}, \mathbf{t}_M\big) \Big| \\
\mathrm{Tr}(\mathbf{Q}_{MMk}) + \mathrm{Tr}(\mathbf{Q}_{FMk}) + \ln(\rho_{Mk}) y_{Mk} + c_{Mk} - \\
\alpha_{Mk} \left\| [x_{MMk}, x_{FMk}]^T \right\| \geq 0, \forall k \in \mathcal{I}_K \\
\left\| \begin{bmatrix} \mathrm{vec}(\mathbf{Q}_{MMk}) \\ \sqrt{2}\mathbf{r}_{MMk} \end{bmatrix} \right\| \leq x_{MMk}, \forall k \in \mathcal{I}_K \\
\left\| \begin{bmatrix} \mathrm{vec}(\mathbf{Q}_{FMk}) \\ \sqrt{2}\mathbf{r}_{FMk} \end{bmatrix} \right\| \leq x_{FMk}, \forall k \in \mathcal{I}_K \\
y_{Mk} \mathbf{I}_{N_{\mathrm{MBS}}} + \mathbf{Q}_{MMk} \succeq \mathbf{0}, \\
y_{Mk} \mathbf{I}_{N_{\mathrm{FBS}}} + \mathbf{Q}_{FMk} \succeq \mathbf{0}, y_{Mk} \geq 0, \ \forall k \in \mathcal{I}_K \Big\} \tag{78}$$

where $\mathbf{t}_M \triangleq [x_{MM1}, \ldots, x_{MMK}, x_{FM1}, \ldots, x_{FMK}, y_{M1}, \ldots, y_{MK}]^T \in \mathbb{R}^{3K}$ are the introduced auxiliary variables.

Similarly, (70c) can be expressed in the following form:

$$
\begin{aligned}
\mathcal{C}_F \triangleq \Big\{ & (\{\mathbf{W}_k\}, \{\mathbf{U}_j\}, \mathbf{t}_F) \Big| \\
& \mathrm{Tr}(\mathbf{Q}_{FFj}) + \mathrm{Tr}(\mathbf{Q}_{MFj}) + \ln(\rho_{Fj}) y_{Fj} + c_{Fj} - \\
& \beta_{Fj} \left\| [x_{FFj}, x_{MFj}]^T \right\| \geq 0, \forall j \in \mathcal{I}_J \\
& \left\| \begin{bmatrix} \mathrm{vec}(\mathbf{Q}_{FFj}) \\ \sqrt{2}\mathbf{r}_{FFj} \end{bmatrix} \right\| \leq x_{FFj}, \forall j \in \mathcal{I}_J \\
& \left\| \begin{bmatrix} \mathrm{vec}(\mathbf{Q}_{MFj}) \\ \sqrt{2}\mathbf{r}_{MFj} \end{bmatrix} \right\| \leq x_{MFj}, \forall j \in \mathcal{I}_J \\
& y_{Fj} \mathbf{I}_{N_{\mathrm{MBS}}} + \mathbf{Q}_{MFj} \succeq \mathbf{0}, \\
& y_{Fj} \mathbf{I}_{N_{\mathrm{FBS}}} + \mathbf{Q}_{FFj} \succeq \mathbf{0}, y_{Fj} \geq 0, \ \forall j \in \mathcal{I}_J \Big\}
\end{aligned}
\tag{79}
$$

where $\mathbf{t}_F \triangleq [x_{FF1}, \ldots, x_{FFJ}, x_{MF1}, \ldots, x_{MFJ}, y_{F1}, \ldots, y_{FJ}]^T \in \mathbb{R}^{3J}$ collects all the auxiliary variables in the derivation of (79).

Conclusion: A new book

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

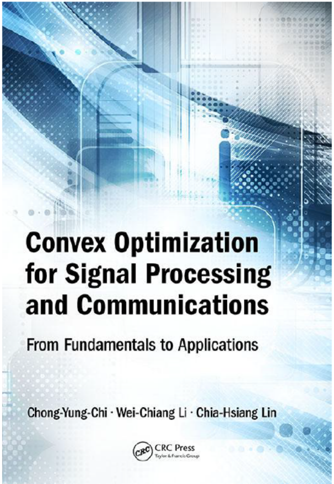
Chong-Yung Chi, Wei-Chiang Li, Chia-Hsiang Lin

(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 8 major universities *(Shandong Univ, Tsinghua Univ, Tianjin Univ, BJTU, Xiamen Univ., UESTC, SYSU, BUPT)* in Mainland China more than 12 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 7 years.