Mobile Network Energy Efficiency Optimization in MIMO Multi-Cell System

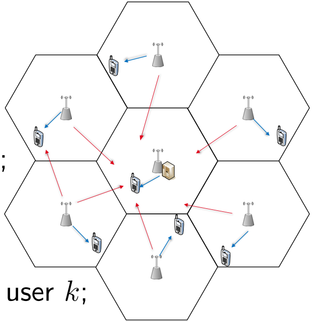
Yang Yang and Dario Sabella

Intel Deutschland GmbH

Acknowledgement

Financial support from EU H2020 projects Fantastic5G and Flex5Gware.

Background: problem model



- MIMO multi-cell network with K cells.
- Each cell serves a single user;
- $\mathbf{Q}_j$: transmit covariance matrix basestation j;
- σ_k^2: noise covariance at user k;
- $\{\mathbf{H}_{kj}\}_{k,j}$: channel from BS j to user k;
- $\sum_{j\neq k} \mathbf{H}_{kj}\mathbf{Q}_j\mathbf{H}_{kj}^H$: interference covariance at user k;
- Downlink achievable rate for user k:

$$r_k(\mathbf{Q}) \triangleq \log\det\left(\mathbf{I} + \left(\sigma_k^2\mathbf{I} + \sum_{j\neq k}\mathbf{H}_{kj}\mathbf{Q}_j\mathbf{H}_{kj}^H\right)^{-1}\mathbf{H}_{kk}\mathbf{Q}_k\mathbf{H}_{kk}^H\right)$$

Background: problem model

- Global Energy efficiency (GEE): Ratio of sum transmission rate divided by sum power consumption.

$$\begin{aligned} \underset{\mathbf{Q}}{\text{maximize}} \quad & \frac{\sum_{k=1}^{K} r_k(\mathbf{Q})}{\sum_{k=1}^{K}(P_{0,k} + \rho_k \text{tr}(\mathbf{Q}_k))} \\ \text{subject to} \quad & \mathbf{Q}_k \succeq \mathbf{0}, \text{tr}(\mathbf{Q}_k) \leq P_k, \forall k. \end{aligned}$$

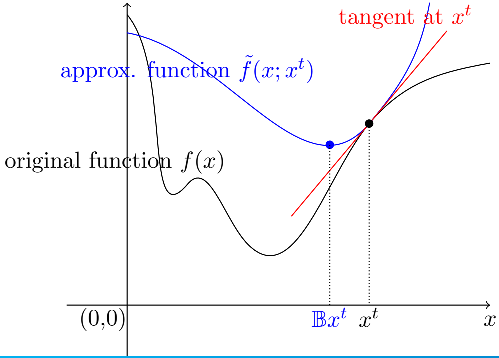
- $P_{0,k}$: the power consumption at the zero RF output power.
- ρ_k: the slope of the load dependent power consumption.
- P_k: the total power constraint.
- The GEE maximization problem is nonconvex and generally difficult to solve (NP-hard).
- Contribution: An efficient iterative algorithm to maximize the energy efficiency.

Background: iterative algorithms

- Consider the following optimization problem:

$$(P): \quad \begin{array}{ll} \underset{\mathbf{x}}{\text{minimize}} & f(\mathbf{x}) \\ \text{subject to} & \mathbf{x} \in \mathcal{X}. \end{array}$$

- Idea: The original problem (P) is solved by solving a sequence of successively refined and easily solvable approximate problems.



Background: iterative algorithms

- A vector $\mathbb{B}\mathbf{x}^t - \mathbf{x}^t$ is a descent direction of $f(\mathbf{x})$ at $\mathbf{x}^t$ if

$$\nabla f(\mathbf{x}^t)^T(\mathbb{B}\mathbf{x}^t - \mathbf{x}^t) < 0.$$

- From the first-order Taylor expansion around $\mathbf{x}^t$:

$$f(\mathbf{x}^t + \gamma(\mathbb{B}\mathbf{x}^t - \mathbf{x}^t)) = f(\mathbf{x}^t) + \underbrace{\gamma \underbrace{\nabla f(\mathbf{x}^t)^T(\mathbb{B}\mathbf{x}^t - \mathbf{x}^t)}_{<0} + o(\gamma\|\mathbb{B}\mathbf{x}^t - \mathbf{x}^t\|)}_{<0 \text{ if } \gamma \text{ is sufficiently small}}.$$

- If $\mathbb{B}\mathbf{x}^t - \mathbf{x}^t$ is a descent direction, then there exists a $\bar{\gamma}^t > 0$ such that

$$f(\mathbf{x}^t + \gamma^t(\mathbb{B}\mathbf{x}^t - \mathbf{x}^t)) < f(\mathbf{x}^t), \forall \gamma^t \in (0, \bar{\gamma}^t].$$

- A practical way to calculate the stepsize is the successive line search.

Background: iterative algorithms

- For example, in the gradient projection method:

$$\mathbb{B}\mathbf{x}^t = [\mathbf{x}^t - s\nabla f(\mathbf{x}^t)]_{\mathcal{X}} = \arg\min_{\mathbf{x}\in\mathcal{X}} \underbrace{\left\|\mathbf{x} - (\mathbf{x}^t - s\nabla f(\mathbf{x}^t))\right\|_2^2}_{\tilde{f}(\mathbf{x};\mathbf{x}^t)}$$

  where $s > 0$ and $[\bullet]_{\mathcal{X}}$ is the projection operator.

- From the first-order optimality condition (i.e., $\nabla f(\mathbf{x}^\star)^T(\mathbf{x} - \mathbf{x}^\star) \geq 0,\ \forall \mathbf{x} \in \mathcal{X}$):

$$(\mathbb{B}\mathbf{x}^t - (\mathbf{x}^t - s\nabla f(\mathbf{x}^t)))^T(\mathbf{x} - \mathbb{B}\mathbf{x}^t) \geq 0,\ \forall \mathbf{x} \in \mathcal{X}.$$

- Setting $\mathbf{x} = \mathbf{x}^t$ in the first-order optimality condition yields

$$(\mathbb{B}\mathbf{x}^t - (\mathbf{x}^t - s\nabla f(\mathbf{x}^t)))^T(\mathbf{x}^t - \mathbb{B}\mathbf{x}^t) \geq 0$$
$$\Downarrow$$
$$\nabla f(\mathbf{x}^t)^T(\mathbb{B}\mathbf{x}^t - \mathbf{x}^t) \leq -\frac{1}{s}\left\|\mathbb{B}\mathbf{x}^t - \mathbf{x}^t\right\|_2^2 < 0.$$

Background: iterative algorithms

Algorithm: $\mathbf{x}^0 \in \mathcal{X}$; repeat the following steps until convergence:

- **S1:** Compute $\mathbb{B}\mathbf{x}^t = \arg\min_{\mathbf{x}\in\mathcal{X}} \tilde{f}(\mathbf{x};\mathbf{x}^t)$.
- **S2:** Compute γ^t by successive (or exact) line search.
- **S3:** Update $\mathbf{x}$: $\mathbf{x}^{t+1} = \mathbf{x}^t + \gamma^t(\mathbb{B}\mathbf{x}^t - \mathbf{x}^t)$ and $t \leftarrow t+1$.

Background: iterative algorithms

Theorem

Assume the following assumptions are satisfied:

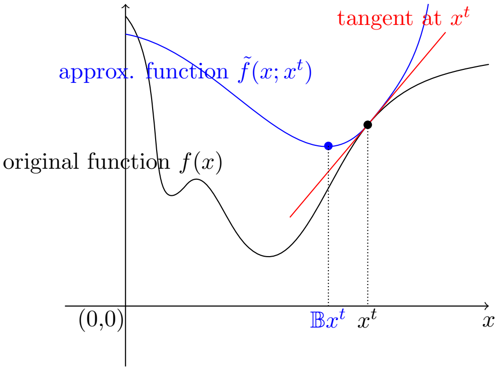
- *(A1):* $\tilde{f}(\mathbf{x};\mathbf{x}^t)$ *is pseudoconvex in* $\mathbf{x} \in \mathcal{X}$ *for any given* $\mathbf{x}^t \in \mathcal{X}$*;*
- *(A2):* $\tilde{f}(\mathbf{x};\mathbf{x}^t)$ *is continuously differentiable in* $\mathbf{x}$ *for a fixed* $\mathbf{x}^t$ *and* $\nabla\tilde{f}(\mathbf{x}^t;\mathbf{x}^t) = \nabla f(\mathbf{x}^t)$*;*
- *(A3):* $\tilde{f}(\mathbf{x};\mathbf{x}^t)$ *is continuous in* $\mathbf{x}^t$ *for a fixed* $\mathbf{x}$*;*

Then $\{f(\mathbf{x}^t)\}$ *is a decreasing sequence and any limit point of* $\{\mathbf{x}^t\}$ *is a stationary point of* (P)*.*

Proof in the following paper:
Y. Yang and M. Pesavento, "A unified successive pseudoconvex approximation framework," *IEEE Transactions on Signal Processing*, vol. 65, no. 13, Jul. 2017.

Background: iterative algorithms

A graphical illustration of the approximate function:



- Note that the approximate function $\tilde{f}(\mathbf{x};\mathbf{x}^t)$ does <u>not</u> have to be a global upper bound of the original function $f(\mathbf{x}^t)$, i.e., $\tilde{f}(\mathbf{x};\mathbf{x}^t) \ngeq f(\mathbf{x})$.

Background: convex functions

- A function $f(\mathbf{x})$ is said to be quasiconvex (strictly quasi-convex) if

$$f((1-\alpha)\mathbf{x}+\alpha\mathbf{y}) \leq (<) \max(f(\mathbf{x}), f(\mathbf{y})), \forall \alpha \in (0,1).$$

- A function $f(\mathbf{x})$ is said to be pseudoconvex if

$$f(\mathbf{y}) < f(\mathbf{x}) \Longrightarrow \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) < 0.$$

- A function is said to be convex (strictly convex) if

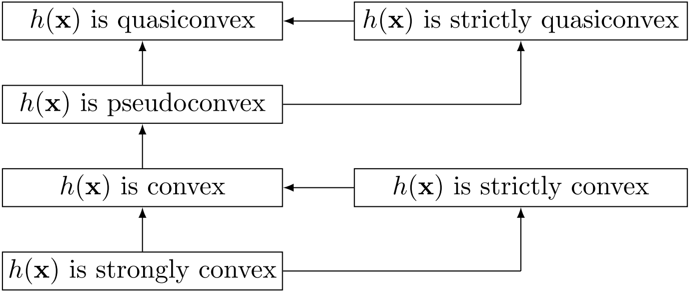
$$f(\mathbf{y}) - f(\mathbf{x}) \geq (>) \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}), \forall \mathbf{x}, \mathbf{y} \in \mathcal{X}.$$

- A function is said to be strongly convex with constant $a > 0$ if

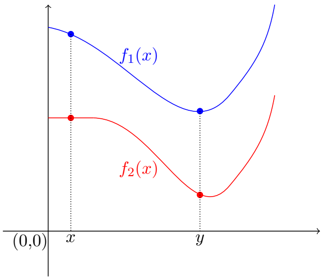
$$f(\mathbf{y}) - f(\mathbf{x}) \geq \nabla f(\mathbf{x})^T(\mathbf{y}-\mathbf{x}) + \frac{a}{2}\|\mathbf{x}-\mathbf{y}\|_2^2, \forall \mathbf{x}, \mathbf{y} \in \mathcal{X}.$$

Background: convex functions

Relationship of functions with different degree of convexity:



Background: convex functions



- Function $f_1(x)$ is pseudoconvex, because

$$f_1(y) < f_1(x) \text{ and } \underbrace{\nabla f_1(x)}_{<0} \underbrace{(y-x)}_{>0} < 0.$$

- Function $f_2(x)$ is quasiconvex but not pseudoconvex, because

$$f_1(y) < f_1(x) \text{ but } \underbrace{\nabla f_2(x)}_{=0} \underbrace{(y-x)}_{>0} = 0.$$

The proposed algorithm

- Energy efficiency (EE): sum rate divided by total power consumption:

$$\begin{array}{ll} \underset{\mathbf{Q}\triangleq(\mathbf{Q}_k)_{k=1}^K}{\text{maximize}} & \dfrac{\sum_{k=1}^K r_k(\mathbf{Q}) \leftarrow \text{nonconcave in } \mathbf{Q}}{\sum_{k=1}^K (P_{0,k} + \rho_k \text{tr}(\mathbf{Q}_k) \leftarrow \text{convex in } \mathbf{Q}} \leftarrow \text{nonconcave} \\ \text{subject to} & \mathbf{Q}_k \succeq \mathbf{0}, \text{tr}(\mathbf{Q}_k) \leq P_k,\ k = 1, \ldots, K, \end{array}$$

where

$$\underbrace{r_k(\mathbf{Q})}_{\text{nonconcave in } \mathbf{Q}} \triangleq \log\det\left(\mathbf{I} + \left(\sigma_k^2 \mathbf{I} + \sum_{j \neq k} \mathbf{H}_{kj} \mathbf{Q}_j \mathbf{H}_{kj}^H\right)^{-1} \mathbf{H}_{kk} \mathbf{Q}_k \mathbf{H}_{kk}^H\right).$$

- The EE maximization problem is nonconvex and generally difficult to solve (NP-hard).

The proposed algorithm: approximate function

- We propose the following approximate function at $\mathbf{Q}^t$:

$$\tilde{f}(\mathbf{Q};\mathbf{Q}^t) = \frac{\sum_{k=1}^{K} \tilde{r}_k(\mathbf{Q}_k;\mathbf{Q}^t)}{\sum_{k=1}^{K}(P_{0,k} + \rho_k \mathrm{tr}(\mathbf{Q}_k))}.$$

where

$$\tilde{r}_k(\mathbf{Q}_k;\mathbf{Q}^t) \triangleq \underbrace{\underbrace{r_k(\mathbf{Q}_k, \mathbf{Q}_{-k}^t)}_{\text{concave in } \mathbf{Q}_k} + \sum_{j \neq k} \underbrace{(\mathbf{Q}_k - \mathbf{Q}_k^t)\nabla_k r_j(\mathbf{Q}^t)}_{\text{linear in } \mathbf{Q}_k}}_{\text{concave in } \mathbf{Q}_k},$$

- Assumption (A1): $\tilde{f}(\mathbf{Q};\mathbf{Q}^t)$, the ratio between a concave function and a convex function, is pseudoconcave (but not concave).
- Assumption (A2): Equal gradient at $\mathbf{Q}^t$ (straightforward algebra):

$$\nabla_{\mathbf{Q}^\star} \tilde{f}(\mathbf{Q};\mathbf{Q}^t)\big|_{\mathbf{Q}=\mathbf{Q}^t} = \nabla_{\mathbf{Q}^\star} f(\mathbf{Q})\big|_{\mathbf{Q}=\mathbf{Q}^t}.$$

The proposed algorithm: approximate function

- We propose the following approximate function at $\mathbf{Q}^t$:

$$\tilde{f}(\mathbf{Q};\mathbf{Q}^t) = \frac{\sum_{k=1}^K \tilde{r}_k(\mathbf{Q}_k;\mathbf{Q}^t)}{\sum_{k=1}^K (P_{0,k} + \rho_k \mathrm{tr}(\mathbf{Q}_k))}.$$

where

$$\tilde{r}_k(\mathbf{Q}_k;\mathbf{Q}^t) \triangleq \underbrace{\underbrace{r_k(\mathbf{Q}_k, \mathbf{Q}_{-k}^t)}_{\text{concave in } \mathbf{Q}_k} + \sum_{j \neq k} \underbrace{(\mathbf{Q}_k - \mathbf{Q}_k^t)\nabla_k r_j(\mathbf{Q}^t)}_{\text{linear in } \mathbf{Q}_k}}_{\text{concave in } \mathbf{Q}_k},$$

- The partial concavity and fractional operater are preserved in the approximate function:
  - The nice problem structure is preserved as much as possible;
  - The convergence speed is (empirically) increased, as observed first in:
    - G. Scutari, F. Facchinei, P. Song, D. P. Palomar, and J.-S. Pang, "Decomposition by partial linearization: Parallel optimization of multi-agent systems," *IEEE Transactions on Signal Processing*, vol. 62, no. 3, pp. 641-656, Feb. 2014.
- Stepsize: successive line search.

The proposed algorithm: approximate problem

We define the approximate problem at $\mathbf{Q} = \mathbf{Q}^t$ as:

$$\mathbb{B}\mathbf{Q}^t = \underset{(\mathbf{Q}_k \succeq \mathbf{0}, \mathrm{tr}(\mathbf{Q}_k) \leq P_k)_{k=1}^K}{\arg\max} \frac{\sum_{k=1}^K \tilde{r}_k(\mathbf{Q}_k; \mathbf{Q}^t)}{\sum_{k=1}^K (P_{0,k} + \rho_k \mathrm{tr}(\mathbf{Q}_k))}.$$

- It represents a fractional programming problem and can be solved by the **Dinkelbach's algorithm**:

$$\begin{aligned} &\underset{(\mathbf{Q}_k \succeq \mathbf{0}, \mathrm{tr}(\mathbf{Q}_k) \leq P_k)_{k=1}^K}{\arg\max} && \sum_{k=1}^K \tilde{r}_k(\mathbf{Q}_k; \mathbf{Q}^t) - \lambda \sum_{k=1}^K (P_{0,k} + \rho_k \mathrm{tr}(\mathbf{Q}_k)) \\ & && \Updownarrow \text{ (parallel decomposition)} \\ &\underset{\mathbf{Q}_k \succeq \mathbf{0}, \mathrm{tr}(\mathbf{Q}_k) \leq P_k}{\arg\max} && \tilde{r}_k(\mathbf{Q}_k; \mathbf{Q}^t) - \lambda (P_{0,k} + \rho_k \mathrm{tr}(\mathbf{Q}_k)), \ k = 1, \ldots, K. \end{aligned}$$

  This solution can be computed in closed form (generalized waterfilling).
- Easy implementation: closed-form solution + parallel decomposition.

The proposed algorithm: approximate problem

at $\tau = 0$: $\lambda^{t,0} = 0$ (or other value).
Step 1: Given $\lambda^{t,\tau}$ at iteration $\tau + 1$ solve the following concave problem:

$$\mathbf{Q}_k(\lambda^{t,\tau}) \triangleq \underset{\mathbf{Q}_k \succeq \mathbf{0}, \text{tr}(\mathbf{Q}_k) \leq P_k}{\arg\max} \; \tilde{r}_k(\mathbf{Q}_k; \mathbf{Q}^t) - \lambda^{t,\tau}(P_{0,k} + \rho_k \text{tr}(\mathbf{Q}_k)).$$

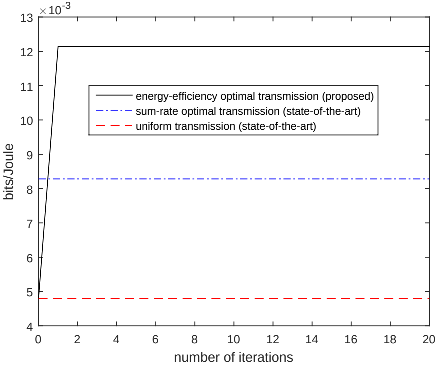
This solution can be computed in closed form (generalized waterfilling).
Step 2: The variable $\lambda^{t,\tau}$ is then updated in iteration $\tau + 1$ as

$$\lambda^{t,\tau+1} = \frac{\sum_{k=1}^{K} \tilde{r}_k(\mathbf{Q}_k(\lambda^{t,\tau}); \mathbf{Q}^t)}{\sum_{k=1}^{K} P_{0,k} + \rho_k \text{tr}(\mathbf{Q}_k(\lambda_k^{t,\tau}))}.$$

- The Dinkelbach's algorithm converges $\lim_{\tau \to \infty} \mathbf{Q}(\lambda^{t,\tau}) = \mathbb{B}\mathbf{Q}^t$ at superlinear rate.

The proposed algorithm: simulations

# of cells	7
# of Tx antennas	4
# of Rx antennas	4
$P_{0,k}$	16W
$P_k/M_{T,k}$	36dBm
ρ	2.6



- The benchmark EE:
  - The sum-rate maximizing scheme, i.e., $(\mathbf{Q}_k)_{k=1}^{K}$ maximizes $\sum_{k=1}^{K} r_k(\mathbf{Q})$ subject to the constraints: $\mathbf{Q}_k \succeq \mathbf{0}, \mathrm{tr}(\mathbf{Q}_k) \leq P_k$ for all $k = 1, \ldots, K$;
  - The uniform transmission scheme, i.e., $\mathbf{Q}_k = P_k/M_{Tx} \cdot \mathbf{I}$;
- Fast convergence and notable improvement are observed.

Questions? Thank you!