Minimization-Maximization Problems: Applications (in Communication), Challenges and Algorithms

Presenter: Mingyi Hong

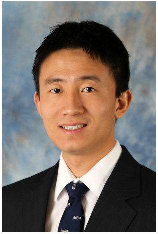
University of Minnesota
ECE Department

May 31, 2019

Collaborators

Songtao Lu

Ioannis Tsaknakis



Mingyi Hong

Yonxin Chen

Outline

1. Min-max problems and motivation
2. The proposed solutions
3. Theoretical guarantees
4. Numerical Results

Mini-max problems

This talk mainly focuses on the following optimization problem

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} \quad f(x; y)$$

f is some (possibly complicated) function over x, y

- x is the usual opt variable, power, precoder design, etc;
- $\min_x f(x, \cdot)$ is the usual cost minimization, i.e.,

$$\text{cost} = -\text{throughput}, \text{ or delay}, etc.$$

- y is used to model provisioning of fairness, robustness, resilience, etc.

Mini-Max Problems

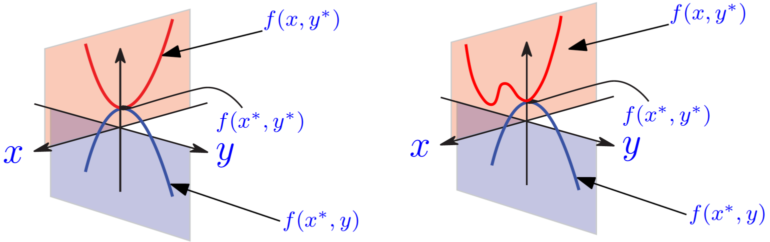


Figure: Left convex/concave min-max problem; Right: Non-convex/Concave min-max problem

Motivation from SP/Comm. perspective

- Question: Why Mini-Max problems?
- The "min" optimizes system level performance; while the "max" provides support such as fairness, robustness, resilience
- "min-max" together helps understand the scenario where some "adversary" (jammer) exists
- Lots of recent interests in this problem, applications in wireless transceiver design, adversarial (GAN)/robust learning, etc.

Example 1: Max-min Fair Beamformer Design

- Setting: MIMO interference channel with K users
- Goal: Design beamformers to maximize the min-rate utility under power and outage constraints

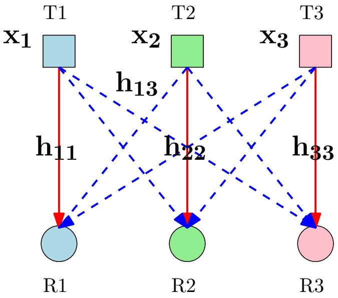


Figure: A set of transmitter-receiver pairs over an interference channel.

Example 1: Max-min Fair Beamformer Design

- The problem is given by ($\mathbf{x}_i$ is user i's transmitter/receiver)

$$\max_{\mathbf{x}} \min_{i \in [K]} R_i(\mathbf{x}_1, \cdots, \mathbf{x}_K)$$

- $R_i(\mathbf{x})$ can be highly non-convex: $\mathbf{X}_i$: the transmit covariance matrix

$$R_i(\mathbf{X}) = \log\det\left(\mathbf{H}_{ii}\mathbf{X}_i\mathbf{H}_{ii}^H\left(\mathbf{I}_{N_r} + \sum_{l \neq i}\mathbf{H}_{li}\mathbf{X}_l\mathbf{H}_{li}^H\right)^{-1} + \mathbf{I}_{N_r}\right)$$

Example 1: Max-min Fair Beamformer Design

- The problem is given by ($\mathbf{x}_i$ is user i's transmitter/receiver)

$$\max_{\mathbf{x}} \min_{i\in[K]} R_i(\mathbf{x}_1, \cdots, \mathbf{x}_K)$$

- $R_i(\mathbf{x})$ can be highly non-convex: $\mathbf{X}_i$: the transmit covariance matrix

$$R_i(\mathbf{X}) = \log\det\left(\mathbf{H}_{ii}\mathbf{X}_i\mathbf{H}_{ii}^H\left(\mathbf{I}_{N_r} + \sum_{l\neq i}\mathbf{H}_{li}\mathbf{X}_l\mathbf{H}_{li}^H\right)^{-1} + \mathbf{I}_{N_r}\right)$$

- Connection to mini-max problem we just need to:
  - Flip the sign: $\min_{\mathbf{x}} \max_{i\in[K]}(-R_i(\mathbf{x}))$
  - Add a variable $\mathbf{y}$ lives in a simplex, and equivalent formulation

$$\min_{\mathbf{x}} \max_{\mathbf{y}} \quad -\sum_{i=1}^{N} R_i(\mathbf{x})y_i := -\mathbf{y}^T\mathbf{R}(\mathbf{x})$$

$$\text{s.t.} \quad \sum_i y_i = 1, \quad y_i \geq 0,\ \forall\ i$$

Example 1: Max-min Fair Beamformer Design

- Problems with known global optimality
  - Max-Min SNR optimization [Zander, 1992] [Foschini and Gans, 1998]
  - MISO BF [Wiesel et al., 2005][Bengtsson and Ottersten, 1999]
  - Joint downlink BS association and power control [Sun-Hong-Luo 14]
  - Many more

Example 1: Max-min Fair Beamformer Design

- Problems with known global optimality
  - Max-Min SNR optimization [Zander, 1992] [Foschini and Gans, 1998]
  - MISO BF [Wiesel et al., 2005][Bengtsson and Ottersten, 1999]
  - Joint downlink BS association and power control [Sun-Hong-Luo 14]
  - Many more
- More recent works involving non-convexity (so, no global min)
  - MISO coordinated beamforming with outrage constraints [Li et al., 2015]
  - MIMO coordinate transceiver design [Liu et al., 2011] (single stream) [Razaviyayn et al., 2011] (multi-stream)
  - MIMO constant envelop transceiver design [Shao et al 19]
  - Many more...

Example 1: Max-min Fair Beamformer Design

- For problems that are global solvable, the standard approach: solve (a sequence of) convex problems like SDP/SOCP
- For non-convex problems, two popular ways in literature
  1. Approximate the mini-max objective
  2. Translate to an "envelop form"

Popular Approximation for Non-Convex Min-Max

- One way to simplify is to use the log-sum approximation

$$\max_i -R_i \approx \frac{1}{\gamma}\log_2\left(\sum_{i=1}^{K} 2^{-\gamma R_i}\right)$$

- A smooth approximation, large γ, good approximation
- Performance degradation

Popular Approximation for Non-Convex Min-Max

- The other way is to introduce an equivalent "envelop form" [Razaviyayn et al., 2011]

$$\min_{\lambda,\mathbf{x}} \lambda, \quad \text{s.t.} \quad -R_i(\mathbf{x}) \leq \lambda, \; \forall\, i, \; \mathbf{x} \in X \tag{1}$$

- Reduces to a minimization form
- But still challenging, involving multiple non-convex constraints

Example 2: Communication in the Presence of Adversary

- How to understand system performance/dynamics of communication systems with adversary?
- Regular user: minimize the cost; Jammer: maximize the cost

Example 2: Communication in the Presence of Adversary

- How to understand system performance/dynamics of communication systems with adversary?
- Regular user: minimize the cost; Jammer: maximize the cost
- An example [Gohary et al., 2009]: Interfering channel, N parallel tones, K users, optimization variables x_k^n's (power allocation of users over the tones), 1 jammer, optimizes y^n

$$\min_{x\in\mathcal{X}}\max_{y\in\mathcal{Y}}\sum_{(k,n)}-\log\left(1+\frac{h_{kk}^n x_k^n}{\sigma^2+\sum_{j=1,j\neq k}^K h_{jk}^n x_j^n + h_{0k}^n y^n}\right)$$

Example 3: Robust Learning (over multiple domains)

- Empirical (non-convex) risk minimization (for training ML models)

$$\min_{\mathbf{x}} \frac{1}{N}\sum_{i=1}^{N} f_i(\mathbf{x})$$

- Treating all data equally/similarly
- In practice, the same model $\mathbf{x}$ is used for multiple domains
- A model for digits recognition can be used in
  1. identify the handwritten digits
  2. recognize the printed digits (e.g., house number)
- A "robust" model has to deal with both of domains [Qian et al., 2018]

Example 3: Robust Learning (over multiple domains)

- Let the data draw from qth domain has lost function $\{f_i^q(\mathbf{x})\}_{i=1}^N$
- When there are Q sets of data drawn from different domain (to describe the same phenomenon) [Qian et al., 2018]

$$\min_{\mathbf{x}} \max_{\mathbf{y}} \sum_{q=1}^{Q} \left(\sum_{i=1}^{N} f_i^q(\mathbf{x}) \right) y_q, \quad \sum_{q=1}^{Q} y_q = 1,\ y_i \geq 0$$

- Here $\mathbf{y}$ can be interpreted as an adversarial distribution
- Identifying the "importance" of different data sets

Example 4: Distributed Learning

- Again the training problem, but with K distributed agents

$$\min_{\mathbf{x}} \frac{1}{K}\sum_{i=1}^{K} f_i(\mathbf{x}_i), \text{ s.t. } \quad \mathbf{x}_i = \mathbf{x}_j, \text{ if } (i,j) \text{ neighbors}$$

Example 4: Distributed Learning

- Again the training problem, but with K distributed agents

$$\min_{\mathbf{x}} \frac{1}{K}\sum_{i=1}^{K} f_i(\mathbf{x}_i), \text{ s.t. } \quad \mathbf{x}_i = \mathbf{x}_j, \text{ if } (i,j) \text{ neighbors}$$

- Stacking all variables $\mathbf{x} := [\mathbf{x}_i, \cdots, \mathbf{x}_K]$, re-write the above problem

$$\min_{\mathbf{x}} \; f(\mathbf{x}), \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = 0$$

where $\mathbf{A}$ is the network incidence matrix (neighboring relations)

Example 4: Distributed Learning

- Again the training problem, but with K distributed agents

$$\min_{\mathbf{x}} \frac{1}{K}\sum_{i=1}^{K} f_i(\mathbf{x}_i), \text{ s.t. } \quad \mathbf{x}_i = \mathbf{x}_j, \text{ if } (i,j) \text{ neighbors}$$

- Stacking all variables $\mathbf{x} := [\mathbf{x}_i, \cdots, \mathbf{x}_K]$, re-write the above problem

$$\min_{\mathbf{x}} \; f(\mathbf{x}), \quad \text{s.t.} \quad \mathbf{A}\mathbf{x} = 0$$

where $\mathbf{A}$ is the network incidence matrix (neighboring relations)

- Introducing the dual variable $\mathbf{y}$, we have

$$\min_{\mathbf{x}} \max_{\mathbf{y}} \qquad f(\mathbf{x}) + \mathbf{y}^T \mathbf{A}\mathbf{x}$$

Challenges

Compared with the min-only problem, mini-max problem is challenging:

1. Competing objectives, how to measure the progress of algorithm?
2. How to characterize the solution quality (when involving non-convexity)?
3. In communication/signal processing applications, we also want computationally efficient algorithms
4. Can we extend the existing algorithms for minimization (like gradient descent) problem to this setting?

Challenges

1. To illustrate some of the challenges, consider the simple setting

$$\min_{\mathbf{x}} \max_{\mathbf{y}} \; \mathbf{y}^T R(\mathbf{x}), \text{ s.t. } \mathbf{x} \in X, \; \mathbf{y} \in \Delta$$

2. Suppose that a good algorithm for the min-problem available (e.g., WMMSE [Shi et al., 2011], FP [Shen-Yu 18], Pricing [Shi et al 08])

$$\min_{\mathbf{x}} \; \mathbf{y}^T R(\mathbf{x}), \text{ s.t. } \mathbf{x} \in X$$

3. A natural approach: Alternatingly perform

$$\mathbf{x}^{r+1} = \text{Algorithm-Step}(\mathbf{x}^r, \mathbf{y}^r), \qquad \mathbf{y}^{r+1} = \max_{\mathbf{y} \in \Delta} \; \mathbf{y}^T R(\mathbf{x}^{r+1})$$

Challenges

- Unfortunately, does not work
- The max step only selects a single user, sets everyone else to zeros
- How about making $\mathbf{y}$ step less greedy (perform one gradient ascent)?

$$\mathbf{x}^{r+1} = \text{Algorithm-Step}(\mathbf{x}^r, \mathbf{y}^r), \quad \mathbf{y}^{r+1} = \left[\mathbf{y}^r + \gamma R(\mathbf{x}^{r+1})\right]^+$$

Challenges

- Unfortunately, does not work
- The max step only selects a single user, sets everyone else to zeros
- How about making $\mathbf{y}$ step less greedy (perform one gradient ascent)?

$$\mathbf{x}^{r+1} = \text{Algorithm-Step}(\mathbf{x}^r, \mathbf{y}^r), \quad \mathbf{y}^{r+1} = \left[\mathbf{y}^r + \gamma R(\mathbf{x}^{r+1})\right]^+$$

- Still does not work; even for bi-linear obj [Daskalakis et al., 2017]

Challenges

- Unfortunately, does not work
- The max step only selects a single user, sets everyone else to zeros
- How about making $\mathbf{y}$ step less greedy (perform one gradient ascent)?

$$\mathbf{x}^{r+1} = \text{Algorithm-Step}(\mathbf{x}^r, \mathbf{y}^r), \quad \mathbf{y}^{r+1} = \left[\mathbf{y}^r + \gamma R(\mathbf{x}^{r+1})\right]^+$$

- Still does not work; even for bi-linear obj [Daskalakis et al., 2017]

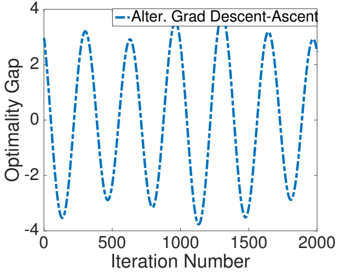


Figure: Gradient Descent/Ascent dynamics exhibit oscillations (for any stepsize)

Challenges

The classical literature in optimization has been concentrated in the case where f is convex/concave

- Extragradient method for finding saddle points in control [P. Vasilyev et al., 2010]
- Subgradient method for saddle-point problems [Nedić and Ozdaglar, 2009]
- Prox-method for smooth convex/concave problems [Nemirovski, 2005]
- Optimistic Gradient Descent Ascent [Daskalakis et al., 2017]
  - f bilinear, i.e $f(x,y) = x^T Ay$,
  - Gradient descent/ascent-type alg.
    $x^{r+1} = x^r - 2\alpha\nabla_x f(x^r, y^r) + \alpha\nabla_x f(x^{r-1}, y^{r-1})$
    $y^{r+1} = x^r + 2\alpha\nabla_y f(x^r, y^r) - \alpha\nabla_y f(x^{r-1}, y^{r-1})$
- A few recent related works on non-convex/concave setting [Nouiehed et al., 2019], [Rafique et al., 2018], [Sanjabi et al., 2018a], [Sanjabi et al., 2018b]

Outline

Our approaches: Main Idea

- Intuitively, the "min" and "max" problems are not created equal

Our approaches: Main Idea

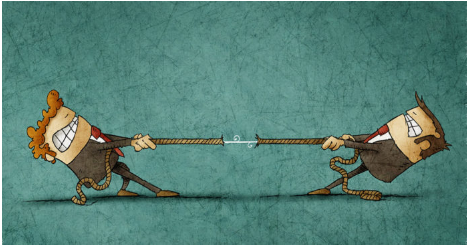
- Intuitively, the "min" and "max" problems are not created equal
- Min outside, Max inside, we should allow the Max be solve relatively well before performing Min

Our approaches: Main Idea

- Intuitively, the "min" and "max" problems are not created equal
- Min outside, Max inside, we should allow the Max be solve relatively well before performing Min
- The alternating GD fails because the two sides are "equally powerful"

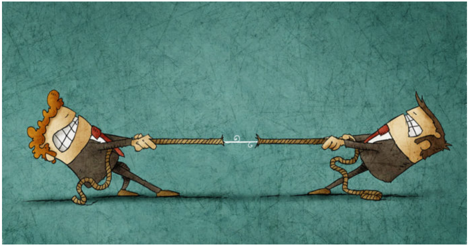
Our approaches: Main Idea

- Intuitively, the “min” and “max” problems are not created equal
- Min outside, Max inside, we should allow the Max be solve relatively well before performing Min
- The alternating GD fails because the two sides are “equally powerful”



Our approaches: Main Idea

- Intuitively, the "min" and "max" problems are not created equal
- Min outside, Max inside, we should allow the Max be solve relatively well before performing Min
- The alternating GD fails because the two sides are "equally powerful"
- The "exact-max" fails because one side that is "too powerful"



Our approaches: Main Idea

- A good algorithm should carefully balance between the two problems
- The min problem has to be slower than the max problem
- The max problem cannot be too aggressive
- Gradually adding regularizers to the min and max problems to control the speed of the two steps?

Our approaches: Main Idea

- Solution concept? First-Order optimality

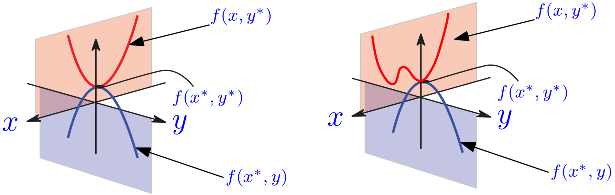
$$\nabla_{\mathbf{x}} f(\mathbf{x}^*, \mathbf{y}^*) = 0,\ \nabla_{\mathbf{y}} f(\mathbf{x}^*, \mathbf{y}^*) = 0$$

or similar concepts to deal with constraints/non-smooth regularizers

- Call ϵ-stationary solution if

$$\|\nabla_{\mathbf{x}} f(\mathbf{x}^*, \mathbf{y}^*)\| \le \epsilon,\ \|\nabla_{\mathbf{y}} f(\mathbf{x}^*, \mathbf{y}^*)\| \le \epsilon$$

- Second-Order optimality: $\mathbf{x}^*$ a "local min" and $\mathbf{y}^*$ a "local max"



Min-max optimization problems

We consider a slightly more general min-max problem that involves

1. multiple blocks for the min problem
2. better modeling scenarios e.g., precoder design for K users; each problem has simpler structure

$$\min_{\{x_i \in \mathcal{X}_i\}} \max_{y \in \mathcal{Y}} \quad f(x_1, x_2, \cdots, x_K; y)$$

- assume that f smooth; non-convex w.r.t. x - concave w.r.t. y

Hybrid Block Successive Approximation Alg. (HiBSA)

- A Hybrid block successive approximation algorithm
- Each time pick one block variable to perform descent/ascent
- Balance two regularization terms with coefficients β^r and γ^r
- A simplified version to illustrate ideas

1 Perform K proximal gradient steps (*Descent steps*)

$$\arg\min_{x_i \in \mathcal{X}_i} \langle \nabla_{x_i} f(x_i^r, w_i^{r+1}, y^r), x_i - x_i^r \rangle + \frac{1}{\beta^r} \|x_i - x_i^r\|^2$$

or equivalently

$$x_i^{r+1} = \text{proj}_{X_i} \left[x_i^r - \beta^r \nabla_{x_i} f(x_i^r, w_i^{r+1}, y^r) \right]$$

2 Perform regularized ascent step (*Ascent step*)

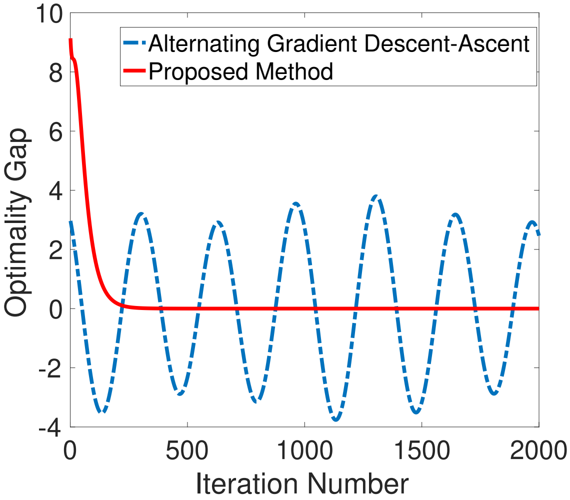
$$y^{r+1} = \arg\max_{y \in \mathcal{Y}} f(x^{r+1}, y) - \gamma^r \|y\|^2 - \frac{1}{2\rho} \|y - y^r\|^2$$

Hybrid Block Successive Approximation Alg. (HiBSA)

- Parameter choices β^r and γ^r are both diminishing sequences
- Intuition, the min steps slows down, allowing max problem to performs more steps per min update
- The max problem has some large regularization at the beginning, avoiding being too greedy at the beginning

Hybrid Block Successive Approximation Alg. (HiBSA)

Now if we apply to the previous problem $\min \max \mathbf{x}^T \mathbf{y}$



HiBSA - Extensions

- Can perform multiple ascent type steps in the maximization problem
- Do not need to perform gradient steps, but can solve some approximated minimization/maximization problems
- These extensions allows for flexible algorithm design, can plug and play existing minimization algorithms for the min-step
- For example, for min fair rate optimization algorithms, allow interlacing between WMMSE [Shi et al., 2011], or FP [Shen-Yu-18] steps, with ascent steps

Outline

1. Min-max problems and motivation
2. The proposed solutions
3. Theoretical guarantees
4. Numerical Results

Convergence Guarantees of HiBSA

Assumptions

(1) For $\{\gamma^r\}$ a diminishing sequence:

$$\gamma^r \to 0, \quad \sum_{r=1}^{\infty} (\gamma^r)^2 = \infty$$

(2) For $\{\beta^r\}$ also a diminishing sequence, but diminishes faster then γ:

$$\beta^r = \mathcal{O}\left((\gamma^r)^2\right)$$

A typical choice: $\gamma^r = \frac{1}{r^{1/4}}$, $\beta^r = 1/\sqrt{r}$

Convergence Guarantees of HiBSA

Theorem (Convergence of HiBSA - Concave case)

For a given $\epsilon > 0$ let $T(\epsilon)$ be the minimum number of iterations needed to reach an ϵ-stationary solution. Then we have

$$\epsilon = \mathcal{O}\left(\frac{\log(T(\epsilon))}{\sqrt{T(\epsilon)}}\right)$$

Section 4

Numerical Results

Max-min fairness coordinated beamforming design

- Setting : MISO interference channel [Li et al., 2015]
- Goal : Design beamformers in order to maximize the min-rate utility under power and outage constraints

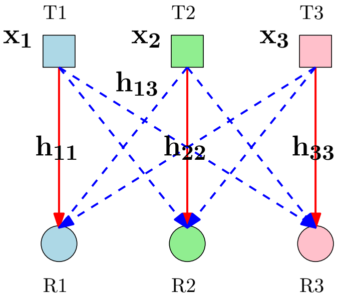


Figure: A set of transmitter-receiver pairs over an interference channel.

Max-min fairness coordinated beamforming design

- Formulation :

$$\max_{x_i \in \mathbb{C}^{N_t}, \forall i} \min_i R_i(\{x_k\}) \quad \text{s.t. } \|x_i\|^2 \le \bar{p}_i, \; \forall i \;\; + \text{ outage prob. constr.}$$

  - $x_i \in \mathbb{C}^{N_t}$ beamforming vectors, N_t no of transmitter antennas
  - $\bar{p}_i$ power constraints
- Solution [Li et al., 2015] :
  - Adopt a suitable surrogate function for the utility function
  - Substitute the inner min problem with the log-sum-exp approximation, i.e $\min_i R_i \approx -\frac{1}{\gamma} log_2 \left(\sum_{i=1}^{K} 2^{-\gamma R_i} \right)$
  - Solve the resulting problem exactly using CVX.

Max-min fairness coordinated beamforming design

1. Approximate problem; solve exactly with CVX [Li et al., 2015]
2. HiBSA [Proposed method]

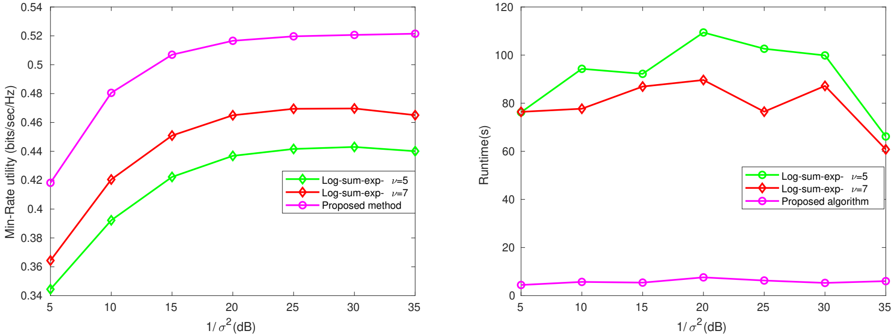


Figure: Min-rate utility and runtime w.r.t noise level

Proposed method achieves higher min-rate utility and is significantly faster.

Max-min fairness linear transceiver design

- Setting : MIMO interfering broadcast channel in a multicell cellular network [Razaviyayn et al., 2011]
- Goal : Design beamformers in order to maximize the min-rate utility under power constraints

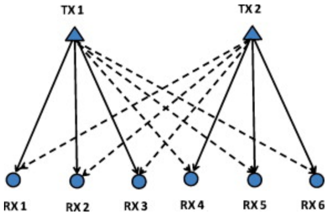


Figure: The interfering broadcast channel model

Max-min fairness linear transceiver design

- Formulation (envelop formulation) :

$$\min_{U,V,W,\lambda} \lambda$$
$$\text{s.t. } Tr[W_{i_k} E_{i_k}] - \log\ \det(W_{i_k}) - d_{i_k} \leq -\lambda,\ \forall i_k \in \mathcal{K}$$
$$\sum_{i \in \mathcal{I}_k} Tr[V_{i_k} V_{i_k}^H] \leq P_k, \forall k \in \mathcal{K}$$

  - i_k : ith user in cell k; $\mathcal{K}$: set of all cells; $\mathcal{I}_k$: set of users in cell k
  - E_{i_k} : MSE for user i_k; d_{i_k} : no of data streams to i_k; P_k : power const.
  - V, U : transmit/receive beamformers; W, λ : auxillary variables
  - Rate of user $i_k = \max_{U_{i_k}, W_{i_k}} \log\ \det(W_{i_k}) - Tr[W_{i_k} E_{i_k}] + d_{i_k}$
- Solution [Razaviyayn et al., 2011] :
  - Solve the U, W subproblems exactly utilizing closed-form solutions
  - Solve the V subproblem exactly using CVX.

Note that previously we resorted to an approximation, whereas here we

Max-min fairness linear transceiver design - Results

1. Envelop based solution [Razaviyayn et al., 2011]
2. HiBSA [Proposed method]

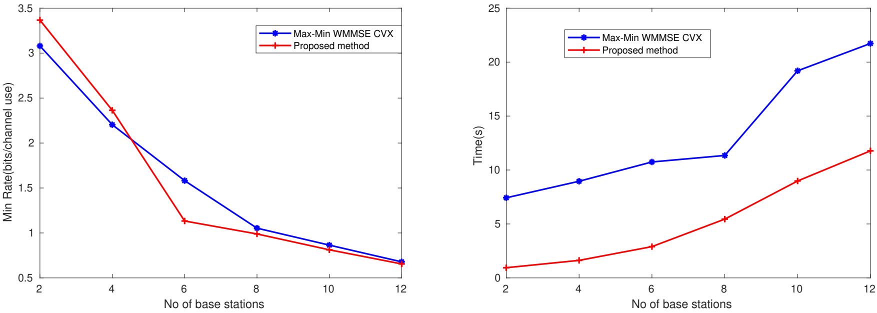


Figure: Min-rate and runtime w.r.t number of cells/base stations

Proposed method achieves comparable utility and is clearly faster.

Power Control in the Presence of Jammer

- Setting : Parallel interference channel model [Gohary et al., 2009]
- Goal :
    - Users : Maximize their individual rates
    - Jammer : Reduce the total sum-rate of the other users
- Methods :
    - Interference pricing method (setting w/o jammer)
    - WMMSE algorithm (setting w/o jammer)
    - Regularized gradient descent/ascent (with jammer) [Proposed method]

Power Control in the Presence of Jammer-Results

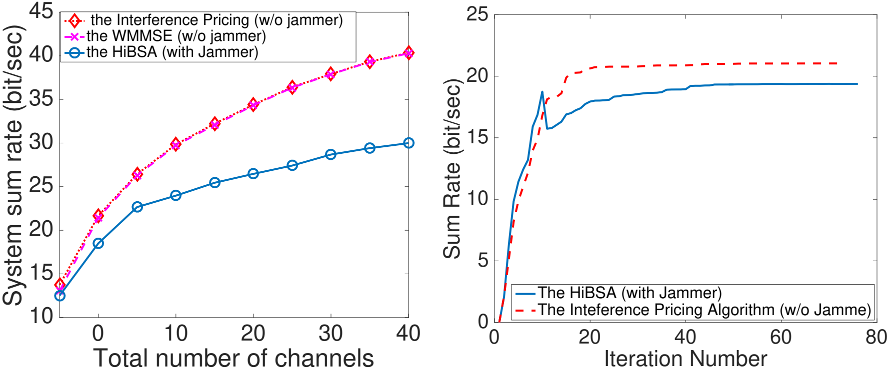


Figure: Sum-rate w.r.t number of channels (left) and number of iterations (right).

Conclusion

- Mini-Max is an interesting optimization problem arises in many contemporary applications
- SP/Comm problems, learning problems, GAN
- More challenging to analyze than the min-only problems
- Preliminary step towards understanding efficient algorithms; other related recent works [Nouiehed et al., 2019], [Rafique et al., 2018], [Sanjabi et al., 2018a], [Sanjabi et al., 2018b]
- Many open problems – both sides non-convex? how to characterize solution quality, etc.
- Our paper can be found online arXiv preprint arXiv:1902.08294

Thank You!

References I

Bengtsson, M. and Ottersten, B. (1999).
Optimal downlink beamforming using semidefinite optimization.
In *Proceedings of the 37th Annual Allerton Conference*.

Daskalakis, C., Ilyas, A., Syrgkanis, V., and Zeng, H. (2017).
Training GANs with optimism.
arXiv preprint arXiv:1711.00141.

Foschini, G. J. and Gans, M. (1998).
On limits of wireless communications in a fading environment when using multiple antennas.
Wireless Personal Communications, 6(3):311–335.

Gohary, R. H., Huang, Y., Luo, Z.-Q., and Pang, J.-S. (2009).
A generalized iterative water-filling algorithm for distributed power control in the presence of a jammer.
IEEE Transactions on Signal Processing, 57(7):2660–2674.

Li, W.-C., Chang, T.-H., and Chi, C.-Y. (2015).
Multicell coordinated beamforming with rate outage constraintpart ii: Efficient approximation algorithms.
IEEE Transactions on Signal Processing, 63(11):2763–2778.

Liu, Y.-F., Dai, Y.-H., and Luo, Z.-Q. (2011).
Max-min fairness linear transceiver design for a multi-user MIMO interference channel.
In *the Proceedings of the international conference on communicaitons*.

Nedić, A. and Ozdaglar, A. (2009).
Subgradient methods for saddle-point problems.
Journal of optimization theory and applications, 142(1):205–228.

References II

Nemirovski, A. (2005).
Prox-method with rate of convergence o(1/t) for variational inequalities with lipschitz continuous monotone operators and smooth convex-concave saddle point problems.
SIAM J. on Optimization, 15(1):229–251.

Nouiehed, M., Sanjabi, M., Lee, J. D., and Razaviyayn, M. (2019).
Solving a class of non-convex min-max games using iterative first order methods.
arXiv preprint arXiv:1902.08297.

P. Vasilyev, F., V. Khoroshilova, E., and S. Antipin, A. (2010).
An extragradient method for finding the saddle point in an optimal control problem.
Moscow University Computational Mathematics and Cybernetics, 34:113–118.

Qian, Q., Zhu, S., Tang, J., Jin, R., Sun, B., and Li, H. (2018).
Robust optimization over multiple domains.
arXiv preprint arXiv:1805.07588.

Rafique, H., Liu, M., Lin, Q., and Yang, T. (2018).
Non-convex min-max optimization: Provable algorithms and applications in machine learning.
arXiv preprint arXiv:1810.02060.

Razaviyayn, M., Hong, M., and Luo, Z. (2011).
Linear transceiver design for a mimo interfering broadcast channel achieving max-min fairness.
In *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pages 1309–1313.

References III

Sanjabi, M., Jimmy, B., Razaviyayn, M., and Lee, J. D. (2018a).
On the convergence and robustness of training GANs with regularized optimal transport.
In *Proceedings of Advances in Neural Information Processing Systems*, pages 7088–7098.

Sanjabi, M., Razaviyayn, M., and Lee, J. D. (2018b).
Solving non-convex non-concave min-max games under polyak-lojasiewicz condition.
arXiv preprint arXiv:1812.02878.

Shi, Q., Razaviyayn, M., Luo, Z.-Q., and He, C. (2011).
An iteratively weighted mmse approach to distributed sum-utility maximization for a mimo interfering broadcast channel.
IEEE Transactions on Signal Processing, 59(9):4331–4340.

Wiesel, A., Eldar, Y. C., and Shamai, S. (2005).
Linear precoding via conic optimization for fixed mimo receivers.
IEEE Transactions on signal processing, 54(1):161–176.

Zander, J. (1992).
Performance of optimal transmitter power control in cellular system.
IEEE Transactions on Vehicular Technology, 41(1):57 –63.