Transmit Beamforming for Minimum Outage via Stochastic Approximation

Yunmei Shi, Aritra Konar,

Nikos Sidiropoulos, Xing-Peng Mao & Yong-Tan Liu

Dept. of ECE, University of Virginia

Dept. of EIE, Harbin Institute of Technology

Dept. of ECE, University of Minnesota

CAMSAP 2017

University of Minnesota

Driven to Discover℠

Introduction

- Transmit Beamforming [Farrokhi *et al.* 1998, Bengtsson-Ottersten 2001, Lopez 2002, Sidiropoulos *et al.* 2006]
  - Exploit CSIT at base station (BS) for enhancing throughput in multi-antenna systems
  - Exact CSIT cannot be obtained in practice
  - Accurately estimating CSI incurs large system overhead
  - Alternative: Use robust beamformer design approach for dealing with channel uncertainty
- This talk:
  - Model: Equally applicable to both point-to-point MISO and single-group multicast beamforming scenarios
  - Adopt outage based design criterion
    - Downlink channels modeled as random vectors
    - Minimize probability of QoS dropping below a certain threshold s.t. power constraints

Prior Art

- **Worst-case design:** [Karipidis *et al.* 2008, Zheng *et al.* 2008, Tajer *et al.* 2011, Song *et al.* 2012, Huang *et al.* 2013, Ma *et al.* 2017]
  - Metric: worst-case QoS w.r.t. all channel perturbations
  - Can result in a very conservative design
- **Outage-based design:** [Xie *et al.* 2005, Vorobyov *et al.* 2008, Ntranos *et al.* 2009, Wang *et al.* 2014, He-Wu 2015, Sohrabi-Davidson 2016]
  - Metric: QoS satisfies threshold with high probability
  - Vary level of conservativeness by changing threshold
  - Majority of prior art:
    - (partial) knowledge of distribution required
    - Approximation algorithms – centralized, can be computationally demanding
  - This talk:
    - No explicit assumptions made on underlying distribution
    - Use stochastic approximation to develop simple online algorithms
    - Limited theoretical analysis, works well in simulations

Problem Statement

- Point-to-point scenario:
  - BS equipped with N transmit antennas
  - Received signal at user:

$$y = \mathbf{h}^H \mathbf{w} s + n$$

  - Model temporal variations of $\mathbf{h}$ as different realizations drawn from an underlying distribution
    - Example: Gaussian Mixture Model [Ntranos *et al.* 2009]
    - Interpretation: Each Gaussian kernel corresponds to a channel state
  - Formulation:

$$\boxed{\min_{\mathbf{w} \in \mathcal{W}} \left\{ F(\mathbf{w}) := \Pr\left(|\mathbf{w}^H \mathbf{h}|^2 \leq \gamma \right) \right\}}$$

$$\boxed{\mathcal{W} := \{\mathbf{w} \in \mathbb{C}^N \mid |w(n)|^2 \leq P_n, \forall\, n \in [N]\}} \Longrightarrow$$ Set of per-antenna power constraints

Problem Statement

- Challenges:
  - If distribution known apriori, no requirement for CSIT
  - However, optimization problem may not be easy to solve
    - NP-hard under GMM assumption [Ntranos *et al.* 2009]
  - Hard to design algorithms for approximate minimization
    - Evaluating cost function and its higher-order derivatives may require computing cumbersome integrals
  - Exact knowledge of distribution not available in practice
- Approach:
  - Reformulate as stochastic optimization problem

$$\boxed{\min_{\mathbf{w}\in\mathcal{W}} \Pr\left(\left|\mathbf{w}^H\mathbf{h}\right|^2 \leq \gamma\right) \Leftrightarrow \min_{\mathbf{w}\in\mathcal{W}} \mathbb{E}_{\mathbf{h}}\left[\mathbb{I}_{\{|\mathbf{w}^H\mathbf{h}|^2\leq\gamma\}}\right]}$$

  - Use stochastic approximation based algorithms for computing solutions

Stochastic Approximation

- Benefits:
  - Explicit knowledge of channel distribution not required
  - Amenable for online implementation
  - Naturally robust to intermittent/stale feedback from the user
    - All channel vectors are statistically equivalent
    - Feedback requirements are considerably relaxed
  - Overall, well suited for massive MIMO systems

- Major roadblock:
  - Indicator function is discontinuous, non-convex
    - Prevents direct application of stochastic approximation based algorithms
  - Our approach: Approximate indicator function via smooth surrogates

Constructing smooth surrogates

- Transformation to real domain:
  - Define $f(\tilde{\mathbf{w}};\tilde{\mathbf{h}}) := \mathbb{I}_{\{\|\tilde{\mathbf{H}}^T\tilde{\mathbf{w}}\|_2^2 \leq \gamma\}}$ where $\tilde{\mathbf{w}} := [\Re[\mathbf{w}]^T, \Im[\mathbf{w}]^T]^T \in \mathbb{R}^{2N}$ $\tilde{\mathbf{h}} := [\Re[\mathbf{h}]^T, \Im[\mathbf{h}]^T]^T \in \mathbb{R}^{2N}$ and $\tilde{\mathbf{H}} := \begin{bmatrix} \Re[\mathbf{h}] & \Im[\mathbf{h}] \\ \Im[\mathbf{h}] & -\Re[\mathbf{h}] \end{bmatrix} \in \mathbb{R}^{2N\times 2}$

- Sigmoid Approximation: $u(\tilde{\mathbf{w}};\tilde{\mathbf{h}}) := \dfrac{1}{1+\exp\left(\|\tilde{\mathbf{H}}^T\tilde{\mathbf{w}}\|_2^2 - \gamma\right)}$

- Smoothed Point-wise Max Approximation:

$$v(\tilde{\mathbf{w}};\tilde{\mathbf{h}}) := \max\left\{0, 1 - \frac{\|\tilde{\mathbf{H}}^T\tilde{\mathbf{w}}\|_2^2}{\gamma}\right\} = \max_{0\leq y\leq 1}\left\{y\left(1 - \frac{\|\tilde{\mathbf{H}}^T\tilde{\mathbf{w}}\|_2^2}{\gamma}\right)\right\}$$

Non-differentiable! Solution: Apply Nesterov Smoothing [Nesterov 2005]

$$v^{(\mu)}(\tilde{\mathbf{w}};\tilde{\mathbf{h}}) = \max_{0\leq y\leq 1}\left\{y\left(1 - \frac{\|\tilde{\mathbf{H}}^T\tilde{\mathbf{w}}\|_2^2}{\gamma}\right) - \frac{\mu}{2}y^2\right\} = \begin{cases} 0, & g(\tilde{\mathbf{w}};\tilde{\mathbf{h}}) < 0 \\ \frac{1}{2\mu}\left(g(\tilde{\mathbf{w}};\tilde{\mathbf{h}})\right)^2, & 0 \leq g(\tilde{\mathbf{w}};\tilde{\mathbf{h}}) \leq \mu \\ g(\tilde{\mathbf{w}};\tilde{\mathbf{h}}) - \frac{\mu}{2}, & g(\tilde{\mathbf{w}};\tilde{\mathbf{h}}) > \mu \end{cases}$$

$$\boxed{g(\tilde{\mathbf{w}};\tilde{\mathbf{h}}) := 1 - \frac{\|\tilde{\mathbf{H}}^T\tilde{\mathbf{w}}\|_2^2}{\gamma}}$$

Problem Formulation

- Modified Problem:
  - Replace indicator function by smooth surrogates to obtain

$$\min_{\tilde{\mathbf{w}} \in \tilde{\mathcal{W}}} \left\{ U(\tilde{\mathbf{w}}) := \mathbb{E}_{\tilde{\mathbf{h}}}[u(\tilde{\mathbf{w}}; \tilde{\mathbf{h}})] \right\}$$

$$\min_{\tilde{\mathbf{w}} \in \tilde{\mathcal{W}}} \left\{ V^{(\mu)}(\tilde{\mathbf{w}}) := \mathbb{E}_{\tilde{\mathbf{h}}}[v^{(\mu)}(\tilde{\mathbf{w}}; \tilde{\mathbf{h}})] \right\}$$

  - Represent both via the prototypical optimization problem

$$\min_{\mathbf{x} \in \mathcal{X}} \mathbb{E}_{\boldsymbol{\xi}}[f(\mathbf{x}; \boldsymbol{\xi})]$$

$\mathcal{X} \subset \mathbb{R}^d$: convex, compact and simple

$\boldsymbol{\xi}$: random vector drawn from unknown probability distribution with support set $\boldsymbol{\Xi} \subset \mathbb{R}^d$

$f : \mathcal{X} \times \boldsymbol{\Xi} \to \mathbb{R}$

$f(.; \boldsymbol{\xi})$: non-convex, twice differentiable

  - Minimize by sequentially processing stream of realizations $\{\boldsymbol{\xi}_t\}_{t=0}^{\infty}$

Algorithms

- Online Gradient Descent (OGD)
  - Given realization $\boldsymbol{\xi}_t$, define $f_t(\mathbf{x}) := f(\mathbf{x}; \boldsymbol{\xi}_t)$
  - Perform update

$$\mathbf{x}^{(t+1)} = \Pi_{\mathcal{X}}(\mathbf{x}^{(t)} - \alpha_t \nabla f_t(\mathbf{x}^{(t)})), \forall\, t \in \mathbb{N}$$

- Online Variance Reduced Gradient (OVRG) [Frostig *et al.* 2015]
  - Streaming variant of SVRG [Johnson-Zhang 2013]
  - Epoch based algorithm
  - At each stage $s \in [S]$, define centering variable $\mathbf{y}_s$
  - Gradient $\mathbb{E}_{\boldsymbol{\xi}}[\nabla f(\mathbf{y}_s; \boldsymbol{\xi})]$ is unavailable, so form surrogate via mini-batching

$$\hat{\mathbf{g}}_s := \frac{1}{k_s} \sum_{i \in [k_s]} \nabla f_i(\mathbf{y}_s)$$

    - Perform update

$$\mathbf{x}_s^{(t+1)} = \Pi_{\mathcal{X}}(\mathbf{x}_s^{(t)} - \alpha_s^{(t)}(\nabla f_t(\mathbf{x}_s^{(t)}) - \nabla f_t(\mathbf{y}_s) + \hat{\mathbf{g}}_s)), \forall\, t \in [T]$$

Baseline for comparison

Alternative approach

$$\min_{\mathbf{w}\in\mathcal{W}} \Pr[|\mathbf{w}^H\mathbf{h}|^2 \le \gamma] \Longleftrightarrow \max_{\mathbf{w}\in\mathcal{W}} \Pr[|\mathbf{w}^H\mathbf{h}|^2 \ge \gamma]$$

- Maximize lower bound of objective function
- NP-hard to compute [Ntranos *et al.* 2009]
- Construct lower bound using moment information [He-Wu 2015]
  - Entails solving non-trivial, non-convex problem
  - Not suitable for online approximation
- Use Markov's inequality to maximize upper bound [Ntranos *et al.* 2009]

$$\Pr[|\mathbf{w}^H\mathbf{h}|^2 \ge \gamma] \le \gamma^{-1}\mathbf{w}^H\mathbf{R}\mathbf{w}, \forall\ \mathbf{w} \in \mathcal{W} \qquad \boxed{\mathbf{R} := \mathbb{E}[\mathbf{h}\mathbf{h}^H]}$$

  - Problem formulation: $\boxed{\max_{\mathbf{w}\in\mathcal{W}} \mathbf{w}^H\mathbf{R}\mathbf{w}}$
  - Approximately maximize in online setting using framework of Stochastic SUM [Sanjabi-Razaviyayn-Luo 2016]

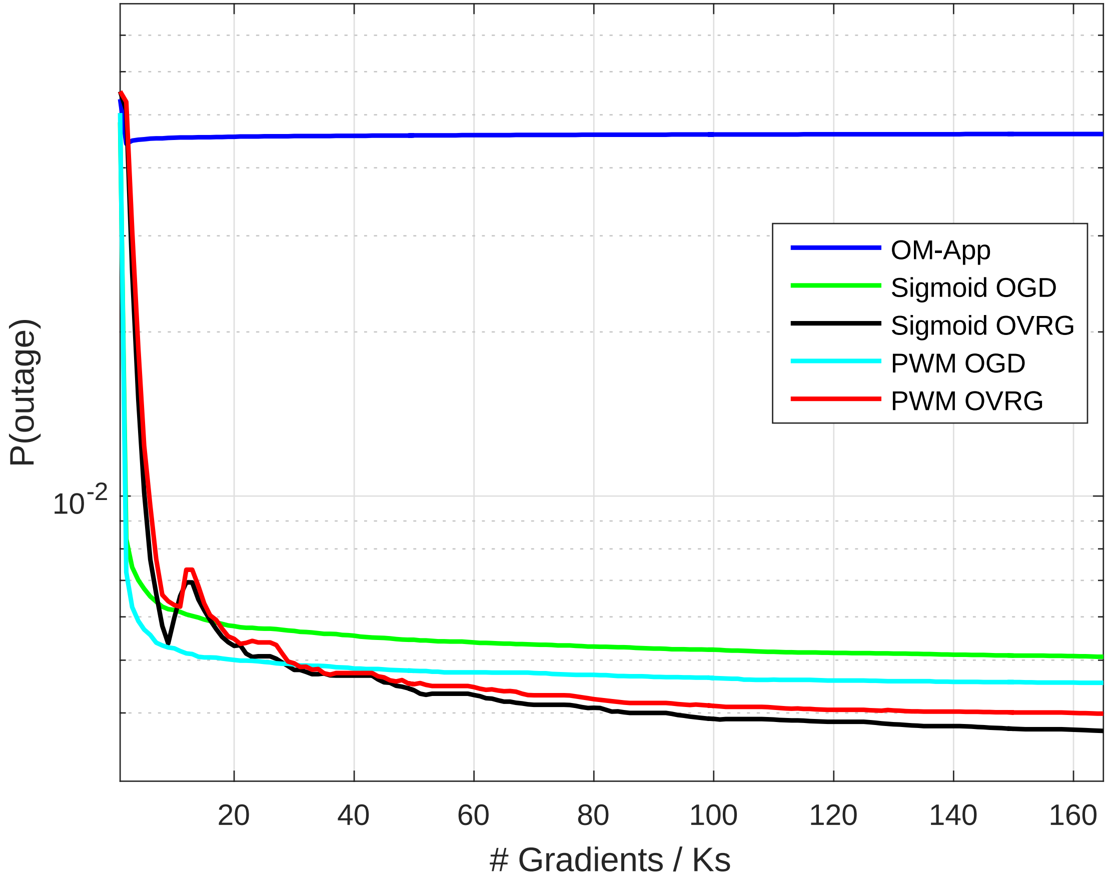
Experiments

Setup

- Algorithms: Sigmoid OGD & OVRG, PWM OGD & OVRG, Online Markov Approximation (OM-App)
- Set smoothing parameter $\mu = 10^{-3}$
- Diminishing step-size for OGD, constant for OVRG and OM-App
- For OVRG
  - Length of each stage $T = 1000$
  - Mini-batch sizes: $$k_s = \begin{cases} 80, & s = 1 \\ 2k_{s-1}, & k_s < 640 \\ 640, & \text{otherwise} \end{cases}$$
- Fix maximum gradient budget for all methods
- Massive MIMO scenario (large no. of BS antennas)
  - Power budget: -6dbW per antenna
- Generate channels using GMM with 4 kernels
  - Equal mixture probabilities
  - Mean of each kernel modeled using different LOS component

Illustrative Example

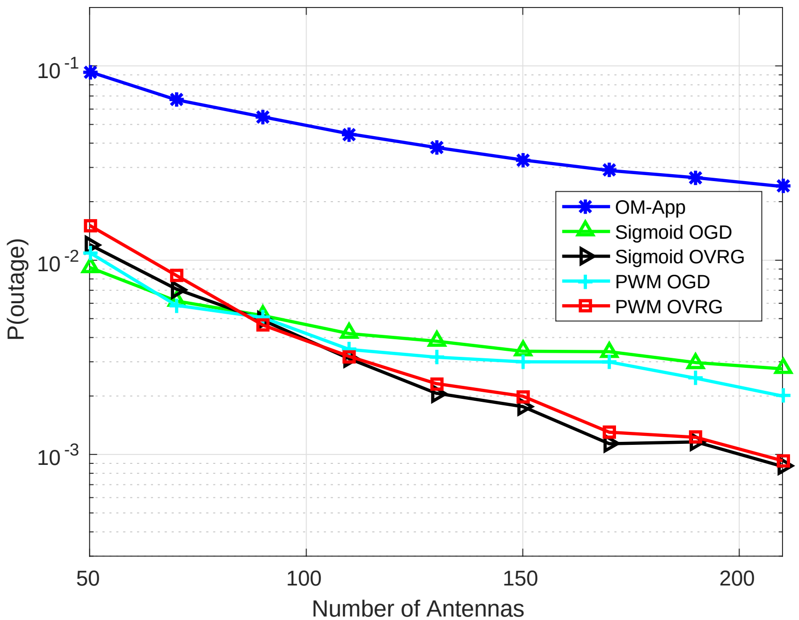
N = 100 BS antennas, Ks = 200, $\gamma = 4$



OVRG based methods attain very satisfactory performance

Detailed Experiments

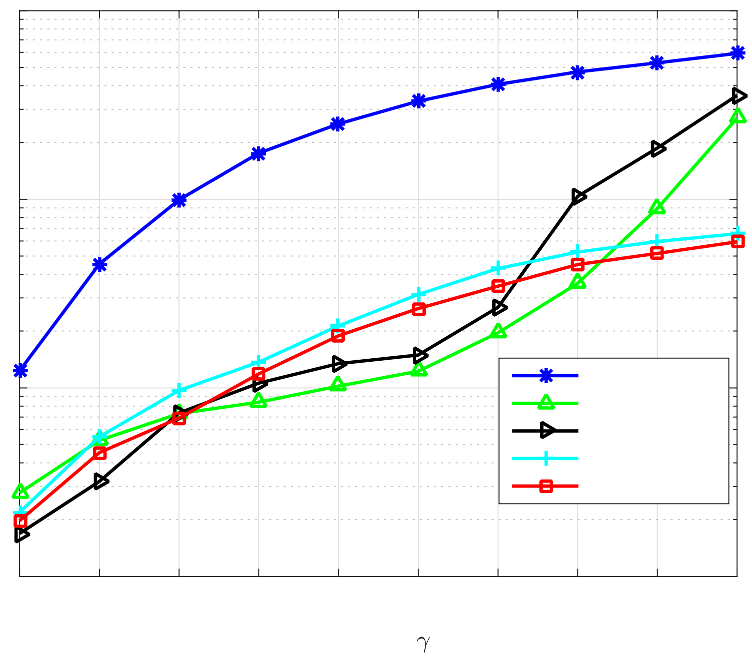
Outage probability vs. no. of BS antennas



$$\gamma = 4$$

OVRG based methods best overall

Outage probability vs. Threshold



$$N = 100$$

Performance of PWM based approximation schemes degrades more gracefully

Conclusions

- Transmit Beamforming for Minimum Outage
  - Minimize outage probability subject to power constraints
  - No prior knowledge of distribution required at BS
  - Reformulate as stochastic optimization problem
  - Construct smooth surrogate of indicator function
  - Use simple stochastic approximation based algorithms for computing solutions based on user feedback
    - Channel estimates can be intermittent/delayed/stale
  - Works well in practice
    - PWM OVRG performs best overall
  - Well suited for massive MIMO systems

Thank you!