The power of (non-convex) relaxation

H. Ghauch, T. Kim, M. Bengtsson, M. Skoglund

ACCESS Linnaeus Center, Royal Insititue of Technology
Department of Electronic Engineering, City University of Hong Kong

ghauch@kth.se, taejokim@cityu.edu.hk, mats.bengtsson@ee.kth.se,
skoglund@ee.kth.se

December 5, 2013
Table of contents

1. Motivation

2. Proposed Scheme
   - Problem Formulation
   - Turbo Optimization
   - Backward Iteration
   - Algorithm Description

3. Numerical Results
   - Simulations
   - Discussion

4. Conclusion
We look at the problem of distributed transmit / receive filter optimization, in the context of MIMO inference channels. A plethora of algorithms that address this problem. They employ the so-called *forward-backward* iterations (F-B):

- Exploit channel reciprocity in TDD
- Receive filters are optimized in the forward iteration, transmit filters are optimized in backward iteration
Motivation

Limitations of current algorithms

- Require a significantly large number of F-B iterations to converge (typically 100 to 1000).
- An extremely high cost, since each F-B iteration corresponds to actual uses of the channel.

Proposed Solution

- We propose 2 algorithms that are constrained to require only a few F-B iterations
- We also introduce a turbo-iteration (inspired from the decoding of turbo-codes), that greatly speeds up the convergence.
Formulation of Leakage Minimization

Forward training: At the $l^{th}$ iteration, each receiver updates its interference covariance matrix, $Q_l \leftarrow Q_{l+1}$, and updates its receive filter $U_l \leftarrow U_{l+1}$, where $U_{l+1}$ is the solution to ($P1$).

\[
(P1) \quad \min_{U_{l+1}} \quad tr(U_{l+1}^T Q_{l+1} U_{l+1})
\]
\[
\text{s.t.} \quad \|U_{l+1}\|_F^2 = P
\]

\[
(LM) \quad \min_{U_{l+1}} \quad tr(U_{l+1}^T Q_{l+1} U_{l+1})
\]
\[
\text{s.t.} \quad U_{l+1}^T U_{l+1} = P/d \cdot I_d
\]

**Proposition 1**

($P1$) is a relaxation of the well-known leakage minimization (LM) problem (Gomadam etal.)

- Not convex, no closed form solution
- Direct solution can only yield rank-one filters
The Need for Structured Updates

One way of getting around the lack of convexity is to impose an update rule. Conceptually, imposing a structure of the variables has the effect of adding an implicit constraint on \((P1)\), i.e.,

\[
(P1) \min_{U_{l+1}} tr(U_{l+1}^\dagger Q_{l+1} U_{l+1})
\]

\[
s.t. \|U_{l+1}\|_F^2 = P
\]

\[
U_{l+1} \in S
\]

Additional constraint might change feasible set of \((P1)\). Thus \((P1)\) might no longer be a relaxation of the original \((LM)\) problem. If \(S\) is pick to be as 'wide' as possible, then it does not alter the feasible set of \((P1)\). This motivates the use of a generic update rule.
The Need for Generic Updates

Thus, we opt to use the following update rule,

\[ U_{l+1} = \Delta A_l + \Phi B_l, \quad l = 1, 2, \ldots \]  
\[ U_{l+1} \in \mathbb{C}^{N\times d}, \quad \Delta \in \mathcal{U}(N, d), \quad \Phi \in \mathcal{U}(N, N - d), \quad \Delta^\dagger \Phi = 0 \]  

Why? it is the most generic update rule

**Proposition 2**

Given \( \Delta, \Phi, \) and full rank \( \Theta \in \mathbb{C}^{N\times d} \), then there exists \( A \in \mathbb{C}^{d\times d}, \) \( B \in \mathbb{C}^{(N-d)\times d} \) such that \( \Theta = \Delta A + \Phi B \), where \( \Delta^\dagger \Phi = 0 \)

**Thus imposing the structure in (1) does not restrict the feasible set of (P1), and thus can be omitted.**
Equivalent Problem

We use (1) as a mapping that allows up to formulate an equivalent problem,

\[
    (P2) \min_{A_l, B_l} f(A_l, B_l) = \text{tr}[(\Delta A_l + \Phi B_l)\dagger Q_{l+1}(\Delta A_l + \Phi B_l)]
\]

\[
    \text{s.t. } \|A_l\|^2_F + \|B_l\|^2_F = P
\]

Since $\Delta, \Phi$ are fixed, the problem now reduced to optimizing the combining weights, $A_l, B_l$.

- The problem is still not jointly convex in the variables
- We then use a Block Coordinate Descent (BCD)
Motivation for Turbo iteration

The use of BCD results in 2 sub-problems - an ideal candidate to use a so-called turbo iteration. Inside the main F-B iteration, we run an additional inner / turbo iteration, where the weights are alternately and sequentially optimized. Given $B_{l,m}$ ($m = \text{turbo iteration number}$), the we define the updates $A_{l,m+1}, B_{l,m+1}$ as

\begin{align}
(K1) \quad A_{l,m+1} &= \arg\min_{A_{l,m}} f(A_{l,m}, B_{l,m}), \quad \text{s.t.} \quad \|A_{l,m+1}\|^2_F = P - \|B_{l,m}\|^2_F \\
(K2) \quad B_{l,m+1} &= \arg\min_{B_{l,m}} f(A_{l,m+1}, B_{l,m}), \quad \text{s.t.} \quad \|B_{l,m}\|^2_F = P - \|A_{l,m+1}\|^2_F
\end{align}

Even then, the non-affine constraint makes the sub-problems non-convex.
**Theorem 1**

There exists primal and dual variables, $A_{l,m+1,}$ $\mu_1$ (or $B_{l,m+1,}$ $\mu_2$) that are local minima to $K1$ (or $K2$), and are given as,

\[ A_{l,m+1} = - (\Delta_Q^{\dagger} \Delta + \mu_1 I)^{-1} \Delta_Q^{\dagger} \Delta B_{l,m} \]  \hspace{1cm} (2)

\[ B_{l,m+1} = - (\Phi_Q^{\dagger} \Phi + \mu_2 I)^{-1} \Phi_Q^{\dagger} \Phi A_{l,m+1} \]  \hspace{1cm} (3)

**Proof:** based on weak duality

**Convergence**

Since each of the updates in (2) and (3), can only decrease the cost function, the sequence $\{f(A_{l,m}, B_{l,m})\}_m$ is non-increasing and converges to a stationary point of $f$. 

H. Ghauch, T. Kim, M. Bengtsson, M. Skoglund  

The power of (non-convex) relaxation
Thus, despite lack of convexity, we see that the use of a turbo iteration allowed us to iteratively approach local minima of the leakage function.

All the details thus far dealt with the forward training phase (optimization of receive filters). The exact same derivations and results can be derived for the reverse training phase, to optimize the transmit filters.

This revealed an elegant 'duality' in the equations, i.e., the transmit filter updates can obtained from the receive filter updates, by replacing each variable by its 'dual' (e.g., $U_{l+1}$ by $V_{l+1}$, $Q_{l+1}$ by $\overline{Q}_{l+1}$, etc.).
Iterative Weight Update

**Forward Training**
- TX 1 → RX 1
  - Transmit Pilots
  - Optimize Receiver Weights
  - Until convergence

- TX 2 → RX 2
  - Transmit Pilots
  - Optimize Receiver Weights
  - Until convergence

- TX 3 → RX 3
  - Transmit Pilots
  - Optimize Receiver Weights
  - Until convergence

**Turbo Iteration**
- RX 1 → TX 1
  - Transmit Pilots
  - Optimize Transmit Weights
  - Until convergence

- RX 2 → TX 2
  - Transmit Pilots
  - Optimize Transmit Weights
  - Until convergence

- RX 3 → TX 3
  - Transmit Pilots
  - Optimize Transmit Weights
  - Until convergence

**Reverse Training**
- TX 1 → RX 1
  - Transmit Pilots

- TX 2 → RX 2
  - Transmit Pilots

- TX 3 → RX 3
  - Transmit Pilots

**Turbo Iteration**
- RX 1 → TX 1
  - Transmit Pilots
  - Optimize Transmit Weights
  - Until convergence

- RX 2 → TX 2
  - Transmit Pilots
  - Optimize Transmit Weights
  - Until convergence

- RX 3 → TX 3
  - Transmit Pilots
  - Optimize Transmit Weights
  - Until convergence

H. Ghauch, T. Kim, M. Bengtsson, M. Skoglund

The power of (non-convex) relaxation
Simulations

Ergodic Sum-rate for $K = 3$, $M = 4$, $N = 4$, $d = 2$, $it = 2$

- Distributed IA
- IWU ($T = 1$)
- IWU ($T = 2$)
- IWU ($T = 5$)

SNR(dB) vs. Ergodic Sum-rate (bps/Hz)
The power of (non-convex) relaxation

Ergodic Sum-rate for $K = 4$, $M = 10$, $N = 10$, $d = 4$, $i_t = 2$

- Distributed IA
- IWU - RR ($T = 1$)
- IWU - RR ($T = 2$)
- IWU - RR ($T = 10$)
We see that the gains proposed scheme kick-in in the medium-to-high SNR regime: this is due to the fact that reducing interference is necessary to increasing the sum-rate, in the high SNR regime.

The gains of the proposed scheme (w.r.t. original LM problem) are firstly derived from solving a relaxed problem, and secondly from introducing of the so-called turbo iteration.
We proposed a distributed scheme in the context of the leakage minimization problem, that delivers high spectral efficiency while requiring an amazingly low number of F-B iterations.

This was made possible by the introduction of the so called turbo iteration, that allowed us to transform the communication overhead to a computational overhead at each receiver.
THANK YOU!

Questions?