Lecture 9

Multi-User MIMO Communications towards Massive MIMO

Dirk Slock

Communication Systems Department, EURECOM

Topics

- Broadcast Channel (BC): Single Cell Multi-User (MU) MIMO
  - beamforming (linear precoding)
  - Dirty Paper Coding (DPC)
- Broadcast Interference Channel (IBC): Multi-Cell Multi-User (MU) MIMO
  - Interference Alignment
- utility optimization, Uplink/Downlink (UL/DL) Duality
  - rate balancing, Perron Frobenius theory
  - max weighted sum rate (WSR), interference leakage aware beamforming and water filling
  - partial CSIT, Expected WSR (EWSR), Expected Weighted Sum MSE (EWSMSE), Expected Signal and Interference Power (ESIP) WSR

Outline

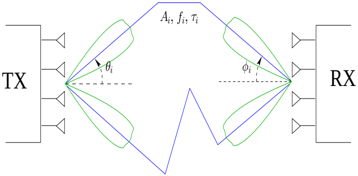
Multi-Antenna Communications

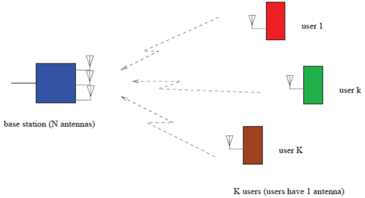
What can we do with multiple antennas:

- Single User (SU) MIMO
- Single Cell Multi-User (MU) MIMO (Broadcast Channel) - Zero Forcing
- Multi-Cell Multi-User MIMO (Interfering Broadcast Channel) - Interference Alignment

SDMA considerations

- Whereas single user (SU) MIMO communications represented a big breakthrough and are now integrated in a number of wireless communication standards, the next improvement is indeed multi-user MIMO (MU MIMO).
- This topic is nontrivial as e.g. illustrated by the fact that standardization bodies took years to get an agreement on the topic to get it included in the LTE-A standard.
- MU MIMO is a further evolution of SDMA, which was THE hot wireless topic throughout the nineties.





MU MIMO key elements

- SDMA is a suboptimal approach to MU MIMO, with transmitter precoding limited to linear beamforming on multipath components, whereas optimal MU MIMO requires Dirty Paper Coding (DPC) and channel impulse responses.
- Channel feedback has gained much more acceptance, leading to good Channel State Information at the Transmitter (CSIT), a crucial enabler for MU MIMO, whereas SDMA was either limited to TDD systems (channel CSIT through reciprocity) or Covariance (multipath) CSIT. In the early nineties, the only feedback that existed was for slow power control.
- Since SDMA, the concepts of multiuser diversity and user selection have emerged and their impact on the MU MIMO sum rate is now well understood. Furthermore, it is now known that user scheduling allows much simpler precoding schemes to be close to optimal.

MU MIMO key elements (2)

- Whereas SU MIMO allows to multiply transmission rate by the spatial multiplexing factor, when mobile terminals have multiple antennas, MU MIMO allows to reach this same gain with single antenna terminals.
- Whereas in SU MIMO, various degrees of CSIT only lead to a variation in coding gain (the constant term in the sum rate), in MU MIMO however CSIT affects the spatial multiplexing factor (= Degrees of Freedom (DoF)) (multiplying the log(SNR) term in the sum rate).
- Intermediate: Dumb Antennas: beamspace or fixed "'beams"'/antenna patterns, UE feedback of Rx gains on different beams (UE side beam preferences)

CSIT considerations

In the process attempting to integrate MU-MIMO into the LTE-A standard, a number of LTE-A contributors had recently become extremely sceptical about the usefulness of the available MU-MIMO proposals. The isue is that they currently do MU-MIMO in the same spirit as SU-MIMO, i.e. with feedback of CSI limited to just a few bits! However, MU-MIMO requires very good CSIT! Some possible solutions:

- Increase CSI feedback enormously (possibly using analog transmission).
- Exploit channel reciprocity in TDD (electronics calibration issue though).
- Limit MU-MIMO to LOS users and extract essential CSIT from DoA or location information.

Singular Value Decomposition (SVD)

- Any rectangular $n \times m$ complex matrix $\mathbf{H}$ can be uniquely decomposed as

$$\mathbf{H} = \mathbf{U}\,\boldsymbol{\Sigma}\,\mathbf{V}^H$$

  where $\mathbf{U}$ is $n \times n$ unitary, $\mathbf{U}^{-1} = \mathbf{U}^H$, $\mathbf{V}$ is $m \times m$ unitary, $\mathbf{V}^{-1} = \mathbf{V}^H$, $\boldsymbol{\Sigma}$ is a $n \times m$ "diagonal" matrix with diagonal elements $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$ which are called resp. the left and right singular vectors, and associated (non-zero) singular values, and r is the rank of $\mathbf{H}$.

- Relations SVD and eigen decompositions:

$$\mathbf{H}\mathbf{H}^H = \mathbf{U}\,\boldsymbol{\Sigma}\boldsymbol{\Sigma}^H\,\mathbf{U}^H \geq 0\ ,\quad \mathbf{H}^H\mathbf{H} = \mathbf{V}\,\boldsymbol{\Sigma}^H\boldsymbol{\Sigma}\,\mathbf{V}^H \geq 0$$

SU MIMO w Perfect CSIT: Optimal Tx Covariance

- For a given known time-invariant channel $\mathbf{H}$, in the presence of additive spatiotemporal white Gaussian noise and under a Tx power constraint, the mutual information maximizing input is a stationary temporally white Gaussian noise.
- It's spatial covariance $\mathbf{T}\mathbf{T}^H$ can be interpreted as the covariance of i.i.d. streams spatially filtered by $\mathbf{T}$.
- The capacity achieving optimal Tx filter $\mathbf{T}$ can be found as

$$
\begin{aligned}
\mathbf{C} &= \max_{\mathbf{T}:\mathrm{tr}\{\mathbf{T}\mathbf{T}^H\}=P} \log_2 \det(I + \tfrac{1}{\sigma_v^2}\,\mathbf{H}\mathbf{T}\mathbf{T}^H\mathbf{H}^H) \\
&= \max_{\mathbf{T}:\mathrm{tr}\{\mathbf{T}\mathbf{T}^H\}=P} \log_2 \det(I + \tfrac{1}{\sigma_v^2}\,\mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^H\mathbf{T}\mathbf{T}^H\mathbf{V}\,\mathbf{\Sigma}^H\,\mathbf{U}^H) \\
&= \max_{\mathbf{T}:\mathrm{tr}\{\mathbf{T}\mathbf{T}^H\}=P} \log_2 \det(I + \tfrac{1}{\sigma_v^2}\,\mathbf{\Sigma}\,\mathbf{V}^H\mathbf{T}\mathbf{T}^H\mathbf{V}\,\mathbf{\Sigma}^H\,\mathbf{U}^H\mathbf{U}) \\
&= \max_{\mathbf{T}:\mathrm{tr}\{\mathbf{T}\mathbf{T}^H\}=P} \log_2 \det(I + \tfrac{1}{\sigma_v^2}\,\mathbf{\Sigma}\,\mathbf{V}^H\mathbf{T}\mathbf{T}^H\mathbf{V}\,\mathbf{\Sigma}^H\,) \\
&= \max_{\mathbf{P}:\mathrm{tr}\{\mathbf{P}\}=P} \log_2 \det(I + \tfrac{1}{\sigma_v^2}\,\mathbf{P}\,\mathbf{\Sigma}^H\,\mathbf{\Sigma})
\end{aligned}
$$

where we used $\det(I+\mathbf{X}\mathbf{Y}) = \det(I+\mathbf{Y}\mathbf{X})$, $\mathbf{U}^H\mathbf{U} = I$, and we introduced the transformation $\mathbf{T}\mathbf{T}^H = \mathbf{V}\,\mathbf{P}\,\mathbf{V}^H$ in which $\mathbf{P} = \mathbf{P}^H \geq 0$.

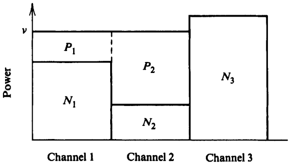
SU MIMO w Perfect CSIT: Optimal Tx Covariance (2)

- Note that for the diagonal part: $\text{diag}\{I + \frac{1}{\sigma_v^2}\,\mathbf{P}\,\mathbf{\Sigma}^H\,\mathbf{\Sigma}\} = I + \frac{1}{\sigma_v^2}\,\mathbf{\Sigma}^H\,\mathbf{\Sigma}\,\text{diag}\{\mathbf{P}\}$, whereas the power constraint only depends on $\text{diag}\{\mathbf{P}\}$. On the other hand, for given (fixed) $\text{diag}\{\mathbf{A}\}$ with $\mathbf{A} = \mathbf{A}^H \geq 0$, $\det(\mathbf{A}) \leq \det(\text{diag}\{\mathbf{A}\})$ (off-diagonal elements lower the determinant). Hence the optimal $\mathbf{P}$ is diagonal. Optimal Tx filter: $\mathbf{T} = \mathbf{V}\,\mathbf{P}^{\frac{1}{2}}$.

SU MIMO w Perfect CSIT: Water Filling

- Stream power optimization (with $r = \min(M, N)$)

$$\mathbf{C} = \max_{\mathbf{P}:\text{tr}\{\mathbf{P}\}=P} \log_2 \det(I + \frac{1}{\sigma_v^2}\mathbf{P}\,\mathbf{\Sigma}^H\,\mathbf{\Sigma})$$
$$= \max_{p_i \geq 0:\, \sum_{i=1}^r p_i = P} \sum_{i=1}^{r} \log_2(1 + \frac{p_i\,\sigma_i^2}{\sigma_v^2})$$



- Lagrangian : $\sum_{i=1}^r \log_2(1 + \frac{p_i\,\sigma_i^2}{\sigma_v^2}) + \lambda(P - \sum_{i=1}^r p_i)$ of which the derivative w.r.t. p_i gives $\frac{\sigma_i^2}{\sigma_v^2}/(1 + \frac{p_i\,\sigma_i^2}{\sigma_v^2}) = \lambda \ln 2$ (if $p_i > 0$). Together with the requirement $p_i \geq 0$ this leads to

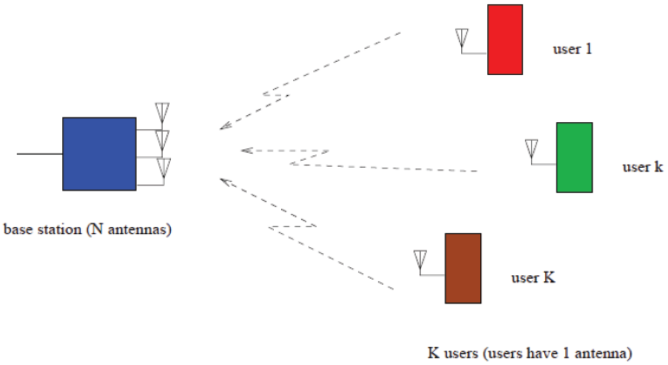
$$p_i = \left[\frac{1}{\lambda \ln 2} - \frac{\sigma_v^2}{\sigma_i^2}\right]_+ \quad (= \;[\nu - N_i]_+ \text{ in figure})$$

where $[.]_+$ denotes the non-negative part of the argument, and the Lagrange multiplier λ can be determined by the bisection method to satisfy $\sum_{i=1}^r p_i = P$.

SDMA system model

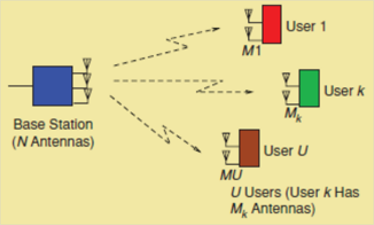
- Rx signal at user k:

$$y_k = \mathbf{h}_k^H \mathbf{x} + n_k, \qquad k = 1, 2, \ldots, K$$



MIMO BroadCast

- MIMO BC = Multi-User MIMO Downlink
- N_t transmission antennas.
- K users with N_k receiving antennas.
- Assume perfect CSI
- Possibly multiple streams/user d_k.
- Power constraint P
- Noise variance $\sigma^2 = 1$.
- $\mathbf{H}_k$ the MIMO channel for user k.



$$\begin{aligned}
\mathbf{F}_k\mathbf{y}_k &= \mathbf{F}_k\mathbf{H}_k \sum_{i=1}^{K} \mathbf{G}_i\mathbf{s}_i + \mathbf{F}_k\mathbf{z}_k \\
&= \underbrace{\mathbf{F}_k\mathbf{H}_k\mathbf{G}_k\mathbf{s}_k}_{\text{useful signal}} + \underbrace{\sum_{i=1, i\neq k}^{K} \mathbf{F}_k\mathbf{H}_k\mathbf{G}_i\mathbf{s}_i}_{\text{inter-user interference}} + \underbrace{\mathbf{F}_k\mathbf{z}_k}_{\text{noise}}
\end{aligned}$$

System Model (2)

- Gaussian signaling
- Rx signal: $\mathbf{y}_k = \mathbf{H}_k\mathbf{x} + \mathbf{z}_k = \mathbf{H}_k \sum_{i=1}^{K} \mathbf{G}_i\mathbf{s}_i + \mathbf{z}_k$
- $$\underbrace{\mathbf{F}_k}_{d_k\times N_k} \underbrace{\mathbf{y}_k}_{N_k\times 1} = \underbrace{\mathbf{F}_k}_{d_k\times N_k} \underbrace{\mathbf{H}_k}_{N_k\times N_t} \sum_{i=1}^{K} \underbrace{\mathbf{G}_i}_{N_t\times d_i} \underbrace{\mathbf{s}_i}_{d_i\times 1} + \underbrace{\mathbf{F}_k}_{d_k\times N_k} \underbrace{\mathbf{z}_k}_{N_k\times 1}$$
- [Christensen etal:T-WC08]: use of linear receivers in MIMO BC is not suboptimal (full CSIT, // SU MIMO): can prefilter $\mathbf{G}_k$ with a $d_k \times d_k$ unitary matrix to make interference plus noise prewhitened channel matrix - precoder cascade of user k orthogonal (columns)
Also: per stream $\equiv$ per user.

Zero-Forcing (ZF)

- ZF-BF $\mathbf{F}_{1:i}\mathbf{H}_{1:i}\mathbf{G}_{1:i} =$

$$\begin{bmatrix} \mathbf{F}_1 & 0 \cdots 0 \\ 0 & \mathbf{F}_2 \ddots \vdots \\ \vdots & \ddots 0 \\ 0 \cdots 0 & \mathbf{F}_i \end{bmatrix} \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_i \end{bmatrix} [\mathbf{G}_1\ \mathbf{G}_2 \cdots \mathbf{G}_i] = \begin{bmatrix} \mathbf{F}_1\mathbf{H}_1\mathbf{G}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{F}_2\mathbf{H}_2\mathbf{G}_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{F}_i\mathbf{H}_i\mathbf{G}_i \end{bmatrix}$$

- ZF-DPC (modulo reordering issues) $\mathbf{F}_{1:i}\mathbf{H}_{1:i}\mathbf{G}_{1:i} =$

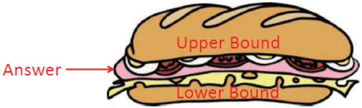
$$\begin{bmatrix} \mathbf{F}_1 & 0 \cdots 0 \\ 0 & \mathbf{F}_2 \ddots \vdots \\ \vdots & \ddots 0 \\ 0 \cdots 0 & \mathbf{F}_i \end{bmatrix} \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_i \end{bmatrix} [\mathbf{G}_1\ \mathbf{G}_2 \cdots \mathbf{G}_i] = \begin{bmatrix} \mathbf{F}_1\mathbf{H}_1\mathbf{G}_1 & 0 & \cdots & 0 \\ * & \mathbf{F}_2\mathbf{H}_2\mathbf{G}_2 & & \vdots \\ \vdots & & \ddots & 0 \\ * & \cdots & * & \mathbf{F}_i\mathbf{H}_i\mathbf{G}_i \end{bmatrix}$$

- BF-style selection, DPC-style selection: as if it's going to be used in BF/DPC

Tx determination and UL/DL Duality (BC/MAC)

- beautifully explained in [ViswanathTse:T-ITaug03]
- start from SU MIMO channel $\boldsymbol{H}$ w stream Tx & Rx filters $\boldsymbol{G}$, $\boldsymbol{F}$ and SINRs: $(\boldsymbol{FHG})^H = \boldsymbol{G}^H\boldsymbol{H}^H\boldsymbol{F}^H$, UL/DL duality for any filters and SINRs, same power feasibility and sum power constraint (and SU MIMO: Gaussian signaling)
- SIMO MAC (Multiple Access Channel) (MU UL) = special case of SU MIMO with $\boldsymbol{G} = \boldsymbol{I}_K$,
  MAC SR$=\ln\det(\boldsymbol{I} + \boldsymbol{HDH}^H)$, tr$\{\boldsymbol{D}\} = P$, $\boldsymbol{D}$ = diagonal
  Rx = stripping (successive interference cancellation and LMMSE)
- MISO BC (Broadcast Channel) (MU DL) = special case of SU MIMO with $\boldsymbol{F} = \boldsymbol{I}_K$,
  duality (same rates, SINRs) for BC/MAC with same Tx/Rx filters and same (sum) power constraint
- Costa: $y = x + s + v$, s known to Tx, has same capacity as $y = x + v$, Dirty Paper Coding (DPC)
- Costa rate region of MISO BC = rate region of SIMO MAC w stripping = MISO rate region lower bound

Tx determination and UL/DL Duality (BC/MAC) (2)



- Sato upper bound: rate region of BC is upper bounded by that of corresponding SU MIMO (Rx's cooperate)
- Observe: difference between "corresponding" MISO BC and SU MIMO: consider SU MIMO with spatially colored noise covariance matrix, only its diagonal elements count in MISO BC.
  Can show that there exists a noise covariance matrix for which coperation between Rx's does not help (via UL/DL relation).
  Hence: Costa lower bound reaches Sato upper bound and hence BC rate region = MAC rate region with sum power constraint.
- Can be immediately extended to MIMO BC and MIMO MAC.
- DPC in "practice": Tomlinson-Harashima (TH), Vector Precoding (VP = vector TH)

Stream Selection Criterion from Sum Rate

- At high SNR, both
  - optimized (MMSE style) filters vs. ZF filters
  - optimized vs. uniform power allocation

  only leads to $\frac{1}{\mathsf{SNR}}$ terms in rates.
- At high SNR, the sum rate is of the form

$$\underbrace{N_t}_{\text{DoF}} \log(\mathsf{SNR}/N_t) + \underbrace{\sum_i \log\det(\mathbf{F}_i \mathbf{H}_i \mathbf{G}_i)}_{\text{constant}} + \mathrm{O}(\frac{1}{\mathsf{SNR}}) + \underbrace{\mathrm{O}(\log\log(\mathsf{SNR}))}_{\text{noncoherent Tx}}$$

for properly normalized ZF Rx $\mathbf{F}_i$ and ZF Tx $\mathbf{G}_i$ (BF or DPC).

User Selection Motivation

- Optimal MIMO BC design requires DPC, which is significantly more complicated than BF.
- User selection allows to
  - improve the rates of DPC
  - bring the rate of BF close to those of DPC
- Optimal user/stream selection requires selection of optimal combination of N_t streams: too complex. Greedy user/stream selection (GUS): select one stream at a time $\Rightarrow$ complexity $\approx$ N_t times the complexity of selecting one stream $(K \gg N_t)$.
- Multiple receive antennas cannot improve the sum rate prelog.
  So what benefit can they bring?
  Of course: cancellation of interference from other transmitters (spatially colored noise): not considered here.

MISO DPC-style GUS

- GUS: Greedy User Selection
  Gram-Schmidt orthogonalize $\mathbf{h}_k$ w.r.t. those of already selected users and choose user with maximum residual norm (matched to DPC).
- MISO: $h_k = \mathbf{H}_k^H$, k_i = user selected at stage i, $H_i = h_{k_{1:i}}^H$.
- (single stream MIMO: $h_k = \mathbf{H}_k^H \boldsymbol{f}_k$, Rx - MIMO channel cascade = MISO channel, all ZF can be done by Tx, for any Rx)
- $$\det(H_i H_i^H) = \prod_{j=1}^{i} \|P^{\perp}_{h_{k_{1:j-1}}} h_{k_j}\|^2$$
- at stage i: $k_i = \arg\max_k \|P^{\perp}_{h_{k_{1:i-1}}} h_k\|^2$
- Introduce ϕ_i = angle between h_{k_i} and $h_{k_{1:i-1}}$ $\Rightarrow$ can write $\|P^{\perp}_{h_{k_{1:i-1}}} h_{k_i}\|^2 = \|h_{k_i}\|^2 \sin^2\phi_i$.

MISO BF-style GUS

$$k_i \text{ maximizes } (\det(\text{diag}\{(H_i H_i^H)^{-1}\}))^{-1} =$$
$$\|P^{\perp}_{h_{k_{1:i-1}}} h_{k_i}\|^2 \prod_{j=1}^{i-1} (\|P^{\perp}_{h_{k_{1:i-1} \setminus k_j}} h_{k_j}\|^2 - \frac{|h_{k_i}^H P^{\perp}_{h_{k_{1:i-1} \setminus k_j}} h_{k_j}|^2}{\|P^{\perp}_{h_{k_{1:i-1} \setminus k_j}} h_{k_i}\|^2})$$

For sufficiently large K, the BF-style user selection process will lead to the selection of channel vectors that are close to being mutually orthogonal. As a result we can write up to first order the contribution of stream i to the sum rate offset

$$\begin{aligned} &\|P^{\perp}_{h_{k_{1:i-1}}} h_{k_i}\|^2 \prod_{j=1}^{i-1} \sin^2 \phi_{ij} \approx \|P^{\perp}_{h_{k_{1:i-1}}} h_{k_i}\|^2 \sin^2 \phi_i \\ &= \|h_{k_i}\|^2 \sin^4 \phi_i = \|P^{\perp}_{h_{k_{1:i-1}}} h_{k_i}\|^4 / \|h_{k_i}\|^2 . \end{aligned} \quad (1)$$

DPC offset is $\|P^{\perp}_{h_{k_{1:i-1}}} h_{k_i}\|^2 = \|h_{k_i}\|^2 \sin^2 \phi_i =$ certain compromise between max $\|h_{k_i}\|^2$ and min $\cos^2 \phi_i$. In the case of BF, $\|h_{k_i}\|^2 \sin^4 \phi_i$ leads to a similar compromise, but with more emphasis on orthogonality.

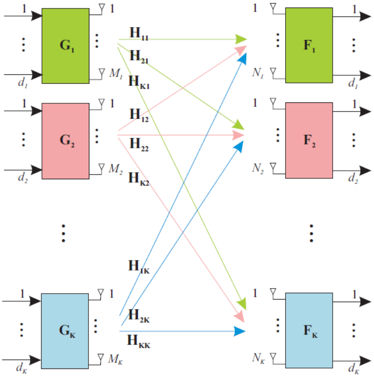
Multi-Cell Multi-User MIMO

Interfering Broadcast Channel - IBC

Interference Alignment

MIMO IC Introduction

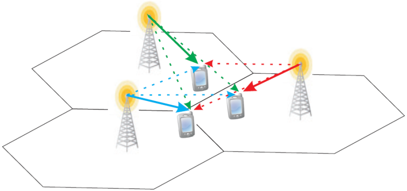
- Interference Alignment (IA) was introduced in [Cadambe,Jafar 2008]
- The objective of IA is to design the Tx beamforming matrices such that the interference at each non intended receiver lies in a common interference subspace
- If alignment is complete at the receiver simple Zero Forcing (ZF) can suppress interference and extract the desired signal
- In [SPAWC2010] we derive a set of interference alignment (IA) feasibility conditions for a K-link frequency-flat MIMO interference channel (IC)
- $d = \sum_{k=1}^{K} d_k$



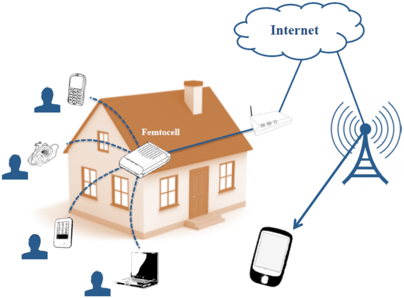
MIMO Interference Channel

Possible Application Scenarios

- Multi-cell cellular systems, modeling intercell interference. Difference from Network MIMO: no exchange of signals, "only" of channel impulse responses.



- HetNets: Coexistence of macrocells and small cells, especially when small cells are considered part of the cellular solution.



Why IA?

- The number of streams (degrees of freedom (DoF)) appearing in a feasible IA scenario correspond to prelogs of feasible multi-user rate tuples in the multi-user rate region.
  Max Weighted Sum Rate (WSR) becomes IA at high SNR.
- Noisy IC: interfering signals are not decoded but treated as (Gaussian) noise.
  Apparently enough for DoF.
- Lots of work more generally on rate prelog regions: involves time sharing, use of fractional power.

Various IA Flavors

- *linear* IA [GouJafar:IT1210], also called *signal space* IA, only uses the spatial dimensions introduced by multiple antennas.
- *asymptotic* IA [CadambeJafar:IT0808] uses symbol extension (in time and/or frequency), leading to (infinite) symbol extension involving diagonal channel matrices, requiring infinite channel diversity in those dimensions. This leads to infinite latency also. The (sum) DoF of asymptotic MIMO IA are determined by the *decomposition* bound [WangSunJafar:isit12].
- *ergodic* IA [NazerGastparJafarVishwanath:IT1012] explains the factor 2 loss in DoF of SISO IA w.r.t. an interference-free Tx scenario by transmitting the same signal twice at two paired channel uses in which all cross channel links cancel out each other: group channel realizations H_1, H_2 s.t. offdiag$(H_2) = -$offdiag(H_1). Ergodic IA also suffers from uncontrolled latency but provides the factor 2 rate loss at any SNR. The DoF of ergodic MIMO IA are also determined by the decomposition bound [LejosneSlockYuan:icassp14].
- *real* IA [MotahariGharanMaddah-AliKhandani:arxiv09], also called it signal scale IA, exploits discrete signal constellations and is based on the Diophantine equation. Although this approach appears still quite exploratory, some related work based on lattices appears promising.

IA as a Constrained Compressed SVD

- (compressed) SVD:

$$H = F\,D'\,G'^H = F\,[D\;0]\left[G\;G''\right]^H = F\,D\,G^H \;\Rightarrow\; F^H H G = D$$

- $F_k^H : d_k \times N_k,\;\; H_{ki} : N_k \times M_i,\;\; G_i : M_i \times d_i$ $\qquad F^H H G =$

$$\begin{bmatrix} F_1^H & 0 & \cdots & 0 \\ 0 & F_2^H & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & F_K^H \end{bmatrix} \begin{bmatrix} H_{11} & H_{12} & \cdots & H_{1K} \\ H_{21} & H_{22} & \cdots & H_{2K} \\ \vdots & & \ddots & \vdots \\ H_{K1} & H_{K2} & \cdots & H_{KK} \end{bmatrix} \begin{bmatrix} G_1 & 0 & \cdots & 0 \\ 0 & G_2 & \ddots & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & G_K \end{bmatrix} = \begin{bmatrix} F_1^H H_{11} G_1 & 0 & \cdots & 0 \\ 0 & F_2^H H_{22} G_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & F_K^H H_{KK} G_K \end{bmatrix}$$

F^H, G can be chosen to be unitary for IA

- per user vs per stream approaches:

  IA: can absorb the $d_k \times d_k$ $F_k^H H_{kk} G_k$ in either F_k^H (per stream LMMSE Rx) or G_k or both.

  WSR: can absorb unitary factors of SVD of $F_k^H H_{kk} G_k$ in F_k^H, G_k without loss in rate $\Rightarrow$ $F^H H G =$ diagonal.

Interference Alignment: Feasibility Conditions (1)

- To derive the existence conditions we consider the ZF conditions

$$\underbrace{\mathbf{F}_k^H}_{d_k \times N_k} \underbrace{\mathbf{H}_{kl}}_{N_k \times M_l} \underbrace{\mathbf{G}_l}_{M_l \times d_l} = \mathbf{0} \;, \quad \forall l \neq k$$

$$\mathrm{rank}(\mathbf{F}_k^H \mathbf{H}_{kk} \mathbf{G}_k) = d_k \;, \quad \forall k \in \{1, 2, \ldots, K\}$$

- rank requirement $\Rightarrow$ SU MIMO condition: $d_k \leq \min(M_k, N_k)$
- The total number of variables in $\mathbf{G}_k$ is $d_k M_k - d_k^2 = d_k(M_k - d_k)$
  Only the subspace of $\mathbf{G}_k$ counts, it is determined up to a $d_k \times d_k$ mixture matrix.
- The total number of variables in $\mathbf{F}_k^H$ is $d_k N_k - d_k^2 = d_k(N_k - d_k)$
  Only the subspace of $\mathbf{F}_k^H$ counts, it is determined up to a $d_k \times d_k$ mixture matrix.

Interference Alignment: Feasibility Conditions (2)

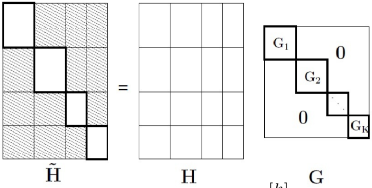
- A solution for the interference alignment problem can only exist if the total number of variables is greater than or equal to the total number of constraints i.e.,

$$\sum_{k=1}^{K} d_k(M_k - d_k) + \sum_{k=1}^{K} d_k(N_k - d_k) \geq \sum_{i\neq j=1}^{K} d_i\, d_j$$
$$\Rightarrow \;\; \sum_{k=1}^{K} d_k(M_k + N_k - 2d_k) \geq (\sum_{k=1}^{K} d_k)^2 - \sum_{k=1}^{K} d_k^2$$
$$\Rightarrow \;\; \sum_{k=1}^{K} d_k(M_k + N_k) \geq (\sum_{k=1}^{K} d_k)^2 + \sum_{k=1}^{K} d_k^2$$

- In the symmetric case: $d_k = d$, $M_k = M$, $N_k = N$: $d \leq \frac{M+N}{K+1}$
- For the $K = 3$ user case ($M = N$): $d = \frac{M}{2}$.
  With 3 parallel MIMO links, half of the (interference-free) resources are available!
  However $d \leq \frac{1}{(K+1)/2} M < \frac{1}{2} M$ for $K > 3$.

Interference Alignment: Feasibility Conditions (3)

The main idea of IA is to convert the alignment requirements at each RX into a rank condition of an associated interference matrix $\mathbf{H}_I^{[k]}$=[$\mathbf{H}_{k1}\mathbf{G}_1, \ldots \mathbf{H}_{k(k-1)}\mathbf{G}_{(k-1)}, \mathbf{H}_{k(k+1)}\mathbf{G}_{(k+1)}, \ldots \mathbf{H}_{kk}\mathbf{G}_K$], that spans the interference subspace at the k-th RX (the shaded blocks in each block row).Thus the dimension of the Interference subspace must satisfy $\mathrm{rank}(\mathbf{H}_I^{[k]}) = r_I^{[k]} \leq N_k - d_k$



The equation above prescribes an upperbound for $r_I^{[k]}$ but the nature of the channel matrix (full rank) and the rank requirement of the BF specifies the following lower bound $r_I^{[k]} \geq \max_{l \neq k}(d_l - [M_l - N_k]_+)$. Imposing a rank $r_I^{[k]}$ on $\mathbf{H}_I^{[k]}$ implies imposing $(N_k - r_I^{[k]})(\sum_{\substack{l=1 \\ l \neq k}}^{K} d_l - r_I^{[k]})$ constraints at RX k. Enforcing the minimum number of constraints on the system implies to have maximum rank: $r_I^{[k]} \leq \min(d_{tot}, N_k) - d_k$

Interference Alignment: Feasibility Conditions (3)

- [BreslerTse:arxiv11]: counting equations and variables not the whole story!
- appears in very "rectangular" ($\neq$ square) MIMO systems
- example: $(M, N, d)^K = (4, 8, 3)^3$ MIMO IFC system
  comparing variables and ZF equations: $d = \frac{M+N}{K+1} = \frac{4+8}{3+1} = \frac{12}{4} = 3$ should be possible
- supportable interference subspace dim. $= N - d =$ 8-3 $=$ 5
- however, the 2 interfering 8×4 cross channels generate 4-dimensional subspaces which in an 8-dimensional space do not intersect w.p. 1 !
- hence, the interfering 4×3 transmit filters cannot massage their 6-dimensional joint interference subspace into a 5-dimensional subspace!
- This issue is not captured by # variables vs # equations: $d = \frac{M+N}{K+1}$ only depends on $M + N$: $(5, 7, 3)^3$, $(6, 6, 3)^3$ work.

Feasibility Linear IA

- We shall focus here on linear IA, in which the spatial Tx filters align their various interference terms at a given user in a common subspace so that a Rx filter can zero force (ZF) it. Since linear IA only uses spatial filtering, it leads to low latency.
- The DoF of linear IA are upper bounded by the so-called **proper bound** [Negro:eusipco09], [Negro:spawc10], [YetisGouJafarKayran:SP10], which simply counts the number of filter variables vs. the number of ZF constraints.
- The proper bound is not always attained though because to make interference subspaces align, the channel subspaces in which they live have to sufficiently overlap to begin with, which is not always the case, as captured by the so-called **quantity bound** [Tingting:arxiv0913] and first elucidated in [BreslerCartwrightTse:allerton11], [BreslerCartwrightTse:itw11], [WangSunJafar:isit12].
- The transmitter coordination required for DL IA in a multi-cell setting corresponds to the Interfering Broadcast Channel (IBC). Depending on the number of interfering cells, the BS may run out of antennas to serve more than one user, which then leads to the Interference Channel (IC).

Necessary & Sufficient Conditions IA Feasibility

- From bilinear to linear system feasibility.
  Feasibility = generic (channel) property. Hence $d\boldsymbol{H}_{kl}$ causes $d\boldsymbol{F}_k$, $d\boldsymbol{G}_l$ s.t.

$$\boldsymbol{F}_k^H \boldsymbol{H}_{kl} \boldsymbol{G}_l = \boldsymbol{0} \;\Rightarrow\; (\boldsymbol{F}_k + d\boldsymbol{F}_k)^H (\boldsymbol{H}_{kl} + d\boldsymbol{H}_{kl})(\boldsymbol{G}_l + d\boldsymbol{G}_l) = \boldsymbol{0}$$

$$\Rightarrow \underbrace{\boldsymbol{F}_k^H \boldsymbol{H}_{kl} \boldsymbol{G}_l}_{=\boldsymbol{0}} + d\boldsymbol{F}_k^H \boldsymbol{H}_{kl} \boldsymbol{G}_l + \boldsymbol{F}_k^H \boldsymbol{H}_{kl}\, d\boldsymbol{G}_l = -\boldsymbol{F}_k^H\, d\boldsymbol{H}_{kl}\, \boldsymbol{G}_l$$

- variable part of Tx/Rx (column space): $\boldsymbol{F}_k = \begin{bmatrix} \boldsymbol{I}_{d_k} \\ \overline{\boldsymbol{F}}_k \end{bmatrix}$, $\boldsymbol{G}_l = \begin{bmatrix} \boldsymbol{I}_{d_l} \\ \overline{\boldsymbol{G}}_l \end{bmatrix}$
  Develop around $\overline{\boldsymbol{F}}_k = \boldsymbol{0}$, $\overline{\boldsymbol{G}}_l = \boldsymbol{0}$, resulting from selecting

$$\boldsymbol{H}_{kl} = \begin{bmatrix} \boldsymbol{H}_{kl}^{(1)} & \boldsymbol{H}_{kl}^{(2)} \\ \boldsymbol{H}_{kl}^{(3)} & \boldsymbol{H}_{kl}^{(4)} \end{bmatrix} = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{H}_{kl}^{(2)} \\ \boldsymbol{H}_{kl}^{(3)} & \boldsymbol{0} \end{bmatrix} \Rightarrow \begin{array}{l} \boldsymbol{F}_k^H \boldsymbol{H}_{kl} \boldsymbol{G}_l = [\boldsymbol{I}\ \boldsymbol{0}]\ \boldsymbol{H}_{kl} \begin{bmatrix} \boldsymbol{I} \\ \boldsymbol{0} \end{bmatrix} \\ = \boldsymbol{H}_{kl}^{(1)} = \boldsymbol{0} \end{array}$$

- Then we get for the feasibility of the perturbed system: stack $\forall\, k, l$

$$\underbrace{\begin{bmatrix} (\boldsymbol{H}_{kl}^{(3)T} \otimes \boldsymbol{I}) & (\boldsymbol{I} \otimes \boldsymbol{H}_{kl}^{(2)}) \end{bmatrix}}_{\boldsymbol{J}} \begin{bmatrix} \mathrm{vec}(d\overline{\boldsymbol{F}}_k^H) \\ \mathrm{vec}(d\overline{\boldsymbol{G}}_l) \end{bmatrix} = \begin{bmatrix} -\mathrm{vec}(d\boldsymbol{H}_{kl}^{(1)}) \end{bmatrix}$$

  IA feasible iff Jacobian $\boldsymbol{J}$ has full row rank.
  Requires # columns $\geq$ # rows, i.e. # variables $\geq$ # ZF constraints: proper condition = necessary. Sufficient conditions depend on scenario.

I and Q components: IA with Real Symbol Streams

- Using real signal constellations in place of complex constellations, transmission over a complex channel of any given dimension can be interpreted as transmission over a real channel of double the original dimensions (by treating the I and Q components as separate channels).
- This doubling of dimensions provides additional flexibility in achieving the total DoF available in the network.
- Split complex quantities in I and Q components:

$$\mathbf{H}_{ij} = \begin{bmatrix} \mathrm{Re}\{\mathbf{H}_{ij}\} & -\mathrm{Im}\{\mathbf{H}_{ij}\} \\ \mathrm{Im}\{\mathbf{H}_{ij}\} & \mathrm{Re}\{\mathbf{H}_{ij}\} \end{bmatrix} \quad \mathbf{x} = \begin{bmatrix} \mathrm{Re}\{\mathbf{x}\} \\ \mathrm{Im}\{\mathbf{x}\} \end{bmatrix}$$

- Example: GMSK in GSM: was considered as wasting half of the resources, but in fact unknowingly anticipated interference treatment: 3 interfering GSM links can each support one GMSK signal without interference by proper joint Tx/Rx design! (SAIC: handles 1 interferer, requires only Rx design).

Outline

1. Multi-Antenna Communication Scenarios

2. Utility Optimization Techniques

Multiple Antennas in Cognitve Radio (CR)

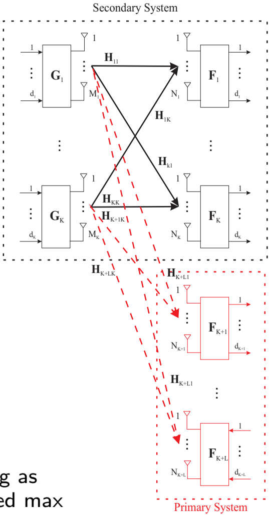
- spatial overlay: MIMO Interference Channel
  - overlay: primary and secondary collaborate
  - exploit multiple antennas and coordinate beamforming to achieve parallel interference-free channels
- spatial underlay
  - primary Rxs with multiple antennas
  - allows interference subspace of max dimension $=$ excess of N_{Rx}^{P} - # primary streams
  - consider interference to primary Rxs OK as long as interference subspace dimension does not exceed max
- spatial interweave
  - # secondary streams $\leq N_{Tx}^{S}$ - N_{Rx}^{P} possible if excess of Secondary Tx antennas over Primary Rx antennas
  - secondary BF nulls to primary Rx antennas w/o primary cooperation: location based (if LOS), reciprocity based in TDD

Spatial Overlay CR

- MIMO Interference Channel
  - also called Coordinated Beamforming
  - perhaps the right paradigm for multi-antenna Licensed Shared Access (LSA)
  - overlay: primary and secondary collaborate
  - A. Goldsmith, S. A. Jafar, I. Maric, and S. Srinivasa, "Breaking spectrum gridlock with cognitive radios: An information theoretic perspective," *Proceedings of the IEEE*, vol. 97, no. 5, pp. 894–914, May 2009.

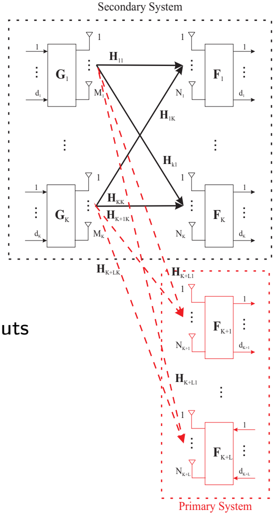
    Overlay: in principle: secondary to help primary transmission, by acting as relay for primary, besides secondary transmission. However, if account for transmission from primary Tx to secondary Tx,
    netDoF $\Rightarrow$ not worth it.
  - exploit multiple antennas and coordinate beamforming to achieve parallel interference-free channels

MIMO Underlay CR



- spatial underlay
  - primary Rxs with multiple antennas
  - allows interference subspace of max dimension $=$ excess of N_{Rx}^P - # primary streams
  - consider interference to primary Rxs OK as long as interference subspace dimension does not exceed max

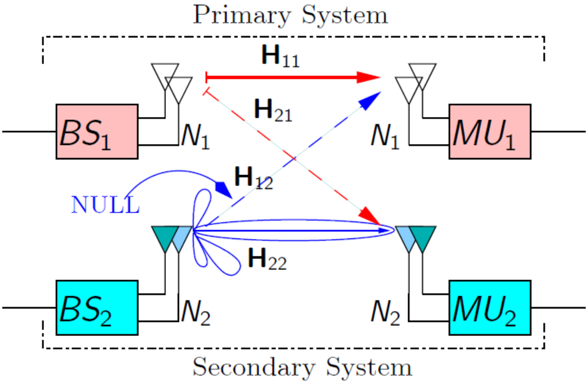
Spatial Interweave CR



- spatial interweave
  - 2 versions: ZF to Primary Rx antennas or Rx outputs
  - # secondary streams $\leq N_{Tx}^{S}$ - N_{Rx}^{P}
    possible if excess of Secondary Tx antennas
    over Primary Rx antennas
  - secondary BF nulls to primary Rx antennas
    w/o primary cooperation:
    location based (if LOS), reciprocity based in TDD

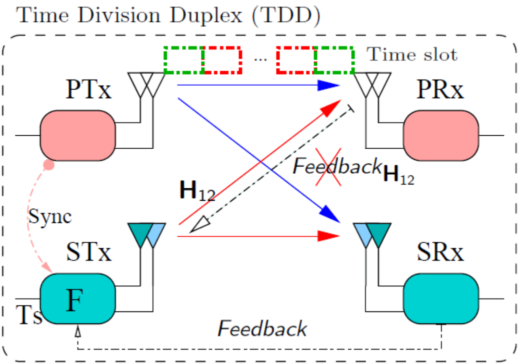
Spatial Interweave Scenario

- (Secondary) Multi-input multi-output (MIMO) system
- Interference avoidance: secondary link should cause no interference
- Spatial Interweave: transmit in the "spatial gaps" (low rank)



TDD Spatial Interweave Cognitive Radio

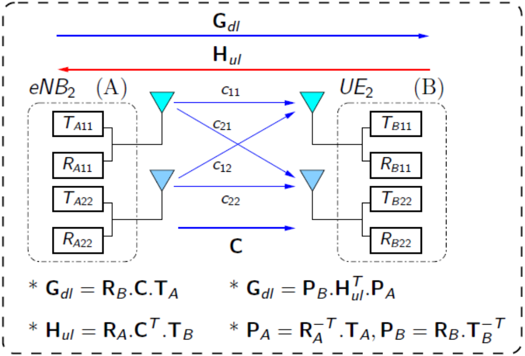
- Zero Forcing Beamforming (ZF-BF): $N2 > N1, \mathbf{R}_{s-p} = \mathbf{H}_{12}\mathbf{F}.\mathbf{T}_s = \mathbf{0}$



- $\mathbf{H}_{12}$ is required at secondary (Tx/BF), to cause null interference towards the primary. Impossibility to feedback crosslink CSI
- Exploit the channel reciprocity (due to TDD)

TDD Reciprocity calibration

- Reciprocity destroyed by *radio frequency (RF) front-ends*



- Calibration to restore the reciprocity: absolute or relative calibration

Reciprocity Calibration in Multiple Links

- in Secondary-to-Primary link:

$$G_{dl}^{SP} = P_B^P \, (H_{ul}^{PS})^T \, P_A^S$$

- in Secondary-to-Secondary link (same secondary):

$$G_{dl}^{SS} = P_B^S \, (H_{ul}^{SS})^T \, P_A^S$$

- spatial interweave beamformer F

$$G_{dl}^{SP} \, F = P_B^P \, (H_{ul}^{PS})^T \, P_A^S \, F = 0 \; \Leftrightarrow \; (H_{ul}^{PS})^T \, P_A^S \, F = 0$$

  insensitive to P_B^P (P_B^P,P_A^S square, G_{dl}^{SP},$(H_{ul}^{PS})^T$ fat, F tall).
- hence, the secondary side calibration matrix P_A^S for transmission to the primary can be derived from the secondary to secondary link.

TDD Massive MIMO

- Massive MIMO: many BS antennas simplifies MU-MIMO design
- requires (at least local) CSIT: use TDD and exploit channel reciprocity
- Thomas Marzetta, Giuseppe Caire, argos.rice.edu : reinvented that BS side calibration can be done by communication with any single UE
- TDD reciprocity calibration in 802.11n, but again removed in 802.11ac

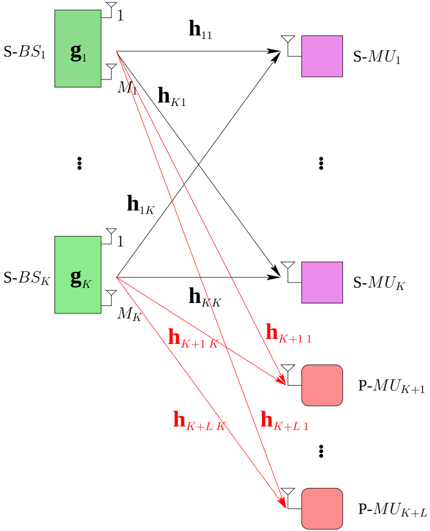
Underlay Cognitive Radio MISO IFC

- L Single antenna Primary MU (P-MU)
- Secondary network: K-user MISO IFC
- BS number k is equipped with M_k antennas
- $\mathbf{g}_k$ is the beamformer applied at the k-th BS
- Rx signal at S-MU_k:

$$y_k = \mathbf{h}_{kk}\mathbf{g}_k s_k + \sum_{j\neq k}^{K} \mathbf{h}_{kj}\mathbf{g}_j s_j + n_k$$

- Rx signal at P-MU_{K+l}:

$$y_{K+l} = \sum_{k=1}^{K} \mathbf{h}_{K+l,k}\mathbf{g}_k s_k$$



Weighted Sum Rate Maximization in the Cognitive MISO IFC

- Our objective is to find the set of BF vectors $\{\mathbf{g}_i\}$ that maximize the secondary network weighted sum rate (WSR)

$$\mathcal{R} = \sum_{k=1}^{K} u_k \ln \left(1 + \frac{\mathbf{h}_{kk}\mathbf{g}_k\mathbf{g}_k^H\mathbf{h}_{kk}^H}{\sum_{j\neq k}^{K} \mathbf{h}_{kj}\mathbf{g}_j\mathbf{g}_j^H\mathbf{h}_{kj}^H + \sigma_k^2} \right)$$

  where u_k is the weight factor for user number k

- Per BS power constraints

$$\text{(C.1)} \;\; \mathbf{g}_k^H\mathbf{g}_k \leq \mathcal{P}_{max,k}^{tx}; \quad \forall k$$

- Due to the Underlay transmission paradigm an Interference Temperature constraints should be imposed:
  - P-MU Interference temperature constraints

$$\text{(C.2)} \;\; \sum_{k=1}^{K} \mathbf{h}_{K+l,k}\mathbf{g}_k\mathbf{g}_k^H\mathbf{h}_{K+l,k}^H \leq \mathcal{P}_{max,l}^{int}; \quad \forall l$$

Weighted Sum Rate Maximization in the Cognitive MISO IFC (2)

- The optimization problem can be mathematically expressed as:

$$\begin{aligned} \max_{\{\mathbf{g}_k\}} \quad & \sum_{k=1}^{K} u_k \ln\left(1 + \frac{\mathbf{h}_{kk}\mathbf{g}_k\mathbf{g}_k^H\mathbf{h}_{kk}^H}{\sum_{j\neq k}^{K}\mathbf{h}_{kj}\mathbf{g}_j\mathbf{g}_j^H\mathbf{h}_{kj}^H + \sigma_k^2}\right) \\ \text{s.t:} \quad & \mathbf{g}_k^H\mathbf{g}_k \leq \mathcal{P}_{max,k}^{tx}; \quad \forall k \\ & \sum_{k=1}^{K}\mathbf{h}_{K+l,k}\mathbf{g}_k\mathbf{g}_k^H\mathbf{h}_{K+l,k}^H \leq \mathcal{P}_{max,l}^{int}; \quad \forall l \end{aligned}$$

- The given optimization Problem is a well known non-convex problem so a global optimum solution cannot be found using standard optimization tools
- We can study the KKT conditions for local optimal solutions

Weighted Sum Rate Maximization in the Cognitive MISO IFC (3)

- The Lagrangian of the proposed optimization problem reads as:

$$\begin{aligned}\mathcal{L}(\mathbf{g}_k, \lambda_k, \mu_k) = & \sum_{k=1}^{K} u_k \ln\left(1 + \frac{\mathbf{h}_{kk}\mathbf{g}_k\mathbf{g}_k^H\mathbf{h}_{kk}^H}{\sum_{j\neq k}^{K}\mathbf{h}_{kj}\mathbf{g}_j\mathbf{g}_j^H\mathbf{h}_{kj}^H + \sigma_k^2}\right) \\ - & \sum_{k=1}^{K}\lambda_k\left(\frac{\mathbf{g}_k^H\mathbf{g}_k}{\mathcal{P}_{max,k}^{tx}} - 1\right) \\ - & \sum_{l=1}^{L}\frac{\mu_l}{\mathcal{P}_{max,l}^{int}}\left(\sum_{k=1}^{K}\mathbf{h}_{K+l,k}\mathbf{g}_k\mathbf{g}_k^H\mathbf{h}_{K+l,k}^H - 1\right)\end{aligned}$$

  - λ_k s the Lagrange multiplier associated to the k-th BS transmission power constraint
  - μ_l is the Lagrange multiplier that takes into account the constraint on the l-th PU received interference.

5G Perspectives

- multi-user multi-cell interference management: theoretical possibilities, but (global) CSIT
  - FB delay $\Rightarrow$ channel prediction and channel Doppler models crucial
  - analog channel FB?
  - FDD: immediate channel FB
  - distributed : yes but watch for fast fading
- Massive MIMO simplifications: separating fast and slow fading channel components
- mmWave (beamforming, bandwidth), spectrum aggregation, full duplex radio
- beyond classical cellular:
  - HetNets (macro/small):
  - wireless/self backhauling

Outline

1. Multi-Antenna Communication Scenarios

2. Utility Optimization Techniques

Utility Functions

- single user (MIMO) in Gaussian noise: Gaussian signaling optimal (avg. power constr.)
- rate stream k : $R_k = \ln(1 + \text{SINR}_k)$
- SINR balancing: $\max_{BF} \min_k \text{SINR}_k / \gamma_k$ under Tx power P, fairness
- related: $\min$ Tx power under $\text{SINR}_k \geq \gamma_k$ GREEN
- max Weighted Sum Rate (WSR): $\max_{BF} \sum_k u_k R_k$, given P

  weights u_k may reflect state of queues (to minimize queue overflow)

  weights also allow to vary orientation of normal to Pareto boundary of rate region and hence to explore whole Pareto boundary if rate region convex
  Pareto boundary: cannot increase an R_k without decreasing some R_i.
- mixed utility functions: max WSR under minimum rate constraints

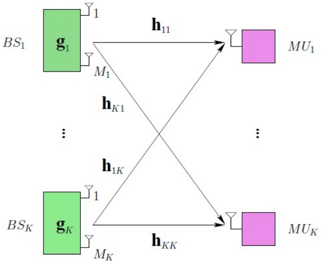
Rate Balancing

- Uplink/Downlink Duality
- Perron-Frobenius Theory for Power Optimization
- Lagrangian : from rate balancing to weighted sum rate

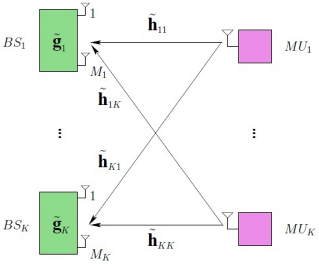
MISO Interference Channel

- K pairs of multiantenna Base Station (BS) and single antenna Mobile User (MU)
- BS number k is equipped with M_k antennas
- $\mathbf{g}_k$ ($\tilde{\mathbf{g}}_k$) is the beamformer (RX filter) applied at the k-th BS in DL (UL) transmission
- y_k is Rx signal at the k-th MU in the DL phase,
  $\tilde{r}_k$ is output of Rx filter at the k-th BS in the UL phase:

$$y_k = \mathbf{h}_{kk}\mathbf{g}_k s_k + \sum_{\substack{l=1\\l\neq k}}^{K} \mathbf{h}_{kl}\mathbf{g}_l s_l + n_k \qquad \tilde{r}_k = \tilde{\mathbf{g}}_k\tilde{\mathbf{h}}_{kk}\tilde{s}_k + \sum_{\substack{l=1\\l\neq k}}^{K} \tilde{\mathbf{g}}_k\tilde{\mathbf{h}}_{kl}\tilde{s}_l + \tilde{\mathbf{g}}_k\tilde{n}_k$$



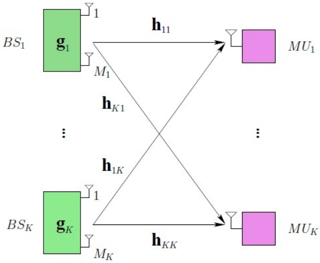
a) MISO DL



b) SIMO UL

UL-DL Duality in MISO/SIMO IFC Under Sum Power Constraint

- MISO DL IFC



- The SINR for the DL channel is:

$$SINR_k^{DL} = \frac{p_k \mathbf{g}_k^H \mathbf{h}_{kk}^H \mathbf{h}_{kk} \mathbf{g}_k}{\sum_{l \neq k} p_l \mathbf{g}_l^H \mathbf{h}_{kl}^H \mathbf{h}_{kl} \mathbf{g}_l + \sigma^2}$$

p_k is the TX power at the k-th BS.

UL-DL Duality in MISO/SIMO IFC Under Sum Power Constraint (2)

- Imposing a set of DL SINR constraints at each mobile station: $SINR_k^{DL} = \gamma_k$ we obtain in matrix notation:

$$\mathbf{\Phi}\mathbf{p} + \boldsymbol{\sigma} = \mathbf{D}\mathbf{p}$$

with:

$$[\mathbf{\Phi}]_{ij} = \begin{cases} \mathbf{g}_j^H \mathbf{h}_{ij}^H \mathbf{h}_{ij} \mathbf{g}_j = |\mathbf{h}_{ij}\mathbf{g}_j|^2, & j \neq i \\ 0, & j = i \end{cases}$$

$$\mathbf{D} = \text{diag}\{\frac{|\mathbf{h}_{11}\mathbf{g}_1|^2}{\gamma_1}, \ldots, \frac{|\mathbf{h}_{KK}\mathbf{g}_K|^2}{\gamma_K}\}, \quad \boldsymbol{\sigma} = \sigma^2 \mathbf{1}\,.$$

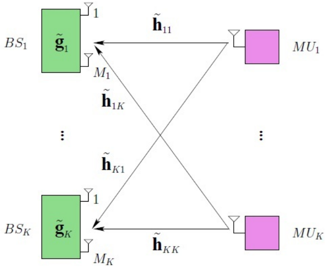
- We can determine the TX power solving w.r.t. **p** obtaining:

$$\mathbf{p} = (\mathbf{D} - \mathbf{\Phi})^{-1}\boldsymbol{\sigma} \tag{2}$$

Feasible ($\mathbf{p} > 0$) if $\max_i |\lambda_i(\mathbf{D}^{-1}\mathbf{\Phi})| < 1$.

UL-DL Duality in MISO/SIMO IFC Under Sum Power Constraint (3)

- SIMO UL IFC



- Assuming that $\tilde{\mathbf{h}}_{ij} = \mathbf{h}_{ji}^H$ and $\tilde{\mathbf{g}}_i = \mathbf{g}_i^H$ the SINR for the UL channel can be written as:

$$SINR_k^{UL} = \frac{q_k \mathbf{g}_k^H \mathbf{h}_{kk}^H \mathbf{h}_{kk} \mathbf{g}_k}{\mathbf{g}_k^H (\sum_{l \neq k} q_l \mathbf{h}_{lk}^H \mathbf{h}_{lk} + \sigma^2 \mathbf{I}) \mathbf{g}_k}$$ Rayleigh quotient

$$\arg\max_{\mathbf{g}_k} SINR_k^{UL} = V_{max}(q_k \mathbf{h}_{kk}^H \mathbf{h}_{kk}, \sum_{l \neq k} q_l \mathbf{h}_{lk}^H \mathbf{h}_{lk} + \sigma^2 \mathbf{I}) \propto (\sum_{l \neq k} q_l \mathbf{h}_{lk}^H \mathbf{h}_{lk} + \sigma^2 \mathbf{I})^{-1} \mathbf{h}_{kk}$$

Tx $\mathbf{g}_k$ = dual UL LMMSE Rx, with UL power q_k from the k-th MS.

Generalized Eigenvectors

- Consider $N \times N$ matrices $\mathbf{A} = \mathbf{A}^H \geq 0$, $\mathbf{B} = \mathbf{B}^H > 0$.
- Generalized eigen vectors V_i and eigen values λ_i

$$\mathbf{A}\, V_i = \lambda_i\, \mathbf{B}\, V_i \ , \ \mathbf{B}^{-1}\mathbf{A}\, V_i = \lambda_i\, V_i \ , \ \det(\mathbf{A} - \lambda\, \mathbf{B}) = 0$$

  $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N \geq 0$, $\mathbf{\Lambda} = \text{diag}\{\lambda_1, \ldots, \lambda_N\}$.
  Note that the V_i in $\boldsymbol{V} = [V_1 \cdots V_N]$ are not orthogonal.
- Let $\mathbf{B} = \mathbf{B}^{1/2}\mathbf{B}^{H/2}$ and ($\boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}$, $\mathbf{\Gamma}$ diagonal)
  eigen decomposition $\mathbf{B}^{-1/2}\mathbf{A}\,\mathbf{B}^{-H/2} = \boldsymbol{U}\mathbf{\Gamma}\boldsymbol{U}^H$. Then

$$\boldsymbol{V} = \mathbf{B}^{-H/2}\boldsymbol{U}\ ,\ \boldsymbol{V}^H\mathbf{A}\,\boldsymbol{V} = \mathbf{\Gamma}\ ,\ \boldsymbol{V}^H\mathbf{B}\,\boldsymbol{V} = \mathbf{\Xi} = \boldsymbol{I}\ ,\ \mathbf{A}\,\boldsymbol{V} = \mathbf{B}\,\boldsymbol{V}\,\mathbf{\Gamma}$$

  so $\mathbf{\Lambda} = \mathbf{\Gamma}$. If we normalize $diag(\boldsymbol{V}^H\boldsymbol{V}) = \boldsymbol{I}$, then $\mathbf{\Xi} \neq \boldsymbol{I}$ and $\mathbf{\Lambda} = \mathbf{\Xi}^{-1}\mathbf{\Gamma}$.
  The non-singular $\boldsymbol{V}$ simultaneously diagonalizes $\mathbf{A}$, $\mathbf{B}$.
- Rayleigh quotient

$$V_1 = V_{max}(\mathbf{A},\mathbf{B}) = \arg\max_V \frac{V^H\mathbf{A}\,V}{V^H\mathbf{B}\,V}\ ,\ \lambda_1 = \text{eig}_{max}(\mathbf{A},\mathbf{B}) = \max_V \frac{V^H\mathbf{A}\,V}{V^H\mathbf{B}\,V}$$

UL-DL Duality in MISO/SIMO IFC Under Sum Power Constraint (4)

- Imposing the same SINR constraints also in the UL: $SINR_k^{UL} = \gamma_k$ it is possible to rewrite that constraints as:

$$\tilde{\boldsymbol{\Phi}}\mathbf{q} + \boldsymbol{\sigma} = \mathbf{D}\mathbf{q}$$

with:

$$[\tilde{\boldsymbol{\Phi}}]_{ij} = \begin{cases} \mathbf{g}_i^H \mathbf{h}_{ji}^H \mathbf{h}_{ji} \mathbf{g}_i = |\mathbf{h}_{ji}\mathbf{g}_i|^2, & j \neq i \\ 0, & j = i \end{cases}$$

$$\mathbf{D} = \mathrm{diag}\{\frac{|\mathbf{h}_{11}\mathbf{g}_1|^2}{\gamma_1}, \ldots, \frac{|\mathbf{h}_{KK}\mathbf{g}_K|^2}{\gamma_K}\}.$$

- The power vector can be found as:

$$\mathbf{q} = (\mathbf{D} - \tilde{\boldsymbol{\Phi}})^{-1}\boldsymbol{\sigma} \tag{3}$$

EURECOM Sophia Antipolis

UL-DL Duality in MISO/SIMO IFC Under Sum Power Constraint (5)

- Comparing the definition we can see that $\tilde{\mathbf{\Phi}} = \mathbf{\Phi}^T$. This implies that there exists a duality relationship between the DL MISO and UL SIMO IFCs.
- We can extend the results for UL-DL duality for MAC/BC [Schubert & Boche'04] to the MISO/SIMO IFC:

Targets $\gamma_1, \ldots, \gamma_K$ are jointly feasible in UL and DL if and only if the spectral radius ρ of the weighted coupling matrix satisfies $\rho(\mathbf{D}^{-1}\mathbf{\Phi}) < 1$.

Both UL and DL have the same SINR feasible region under a sum-power constraint, i.e., target SINRs are feasible in the DL if and only if the same targets are feasible in the UL:

$$\sum_i q_i = \mathbf{1}^T\mathbf{q} = \sigma^2\mathbf{1}^T(\mathbf{D} - \mathbf{\Phi}^T)^{-1}\mathbf{1} = \sigma^2\mathbf{1}^T(\mathbf{D} - \mathbf{\Phi})^{-1}\mathbf{1} = \sum_i p_i \tag{4}$$

- Using this results it is possible to extend some BF design techniques used in the BC [Schubert & Boche'04] to the MISO IFC:
  - Max-Min SINR (SINR Balancing)
  - Power minimization under SINR constraints

Perron Frobenius Theory

- Weighted SINR balancing:

$$\max_{\mathbf{p}} \min_{k} \frac{SINR_k^{DL}}{\gamma_k} \Leftrightarrow \min_{\mathbf{p}} \max_{k} \frac{\gamma_k}{SINR_k^{DL}} = \min_{\mathbf{p}} \max_{k} \frac{1}{p_k}[\mathbf{D}^{-1}\mathbf{\Phi}\mathbf{p} + \mathbf{D}^{-1}\boldsymbol{\sigma}]_k$$

- Consider the power constraint $\mathbf{1}^T\mathbf{p} = P$, reparameterize $\mathbf{p} = \frac{P}{\mathbf{1}^T\mathbf{p}'}\,\mathbf{p}'$ which satisfies the power constraint for any $\mathbf{p}'$ and rename $\mathbf{p}'$ as $\mathbf{p}$. Then the SINR balancing becomes

$$\min_{\mathbf{p}} \max_{k} \frac{[\mathbf{\Lambda}\,\mathbf{p}]_k}{p_k} \quad \text{with} \quad \mathbf{\Lambda} = \mathbf{D}^{-1}[\mathbf{\Phi} + \frac{\sigma^2}{P}\,\mathbf{1}\mathbf{1}^T]$$

- Classical nonnegative vectors and matrices: Note that the optimal powers should satisfy the power constraint with equality. Furthermore, at the optimum we shall have th equality

$$\frac{SINR_k^{DL}}{\gamma_k} = \frac{1}{\Delta}, \forall k \quad \leftrightarrow \quad \mathbf{\Lambda}\,\mathbf{p} = \Delta\,\mathbf{p}$$

since otherwise the user with higher SINR can lower its power, reducing interference to the user with the lowest SINR, which then increases. So, $\mathbf{p}$ is an eigen vector of $\mathbf{\Lambda}$ with eigen value Δ.

Perron Frobenius Theory (2)

- Now, for a non-negative matrix $\mathbf{\Lambda}$, the eigenvalue of the largest magnitude is positive, and its corresponding eigenvector $\mathbf{p}$ can be chosen to be non-negative. For a non-negative matrix, the non-negative eigen vector corresponding to the eigenvalue of the largest norm is positive and so is the corresponding eigen value Δ. There is only one such positive eigen pair which is called Perron Frobenius.
- Actually, without physical motivation for the SINR equality, the Collatz-Wielandt formula for the Perron-Frobenius eigen pair is

$$\Delta = \min_{\mathbf{p}} \max_k \frac{[\mathbf{\Lambda}\,\mathbf{p}]_k}{p_k} = \text{eig}_{max}(\mathbf{\Lambda}),\ \mathbf{p}' = \arg\min_{\mathbf{p}} \max_k \frac{[\mathbf{\Lambda}\,\mathbf{p}]_k}{p_k} = V_{max}(\mathbf{\Lambda})$$

- The non-smooth optimization criterion $\min_{\mathbf{p}} \max_k \frac{\gamma_k}{SINR_k^{DL}}$ can be transformed into a smooth problem

$$\max_{\boldsymbol{\lambda}} \min_{\mathbf{p}} \{\, t + \sum_k \lambda_k (\frac{\gamma_k}{SINR_k^{DL}} - t)\,\}$$

where the $\lambda_k \geq 0$ are the Lagrange multipliers for the constraints $\frac{\gamma_i}{SINR_i^{DL}} \leq t = \max_k \frac{\gamma_k}{SINR_k^{DL}}$.

Perron Frobenius Theory (3)

- This leads to the Donsker-Varadhan-Friedland formula

$$\Delta = \text{eig}_{max}(\mathbf{\Lambda}) = \max_{\boldsymbol{\lambda}:\sum_k \lambda_k = 1} \min_{\mathbf{p}} \sum_k \lambda_k \frac{[\mathbf{\Lambda}\,\mathbf{p}]_k}{p_k}$$

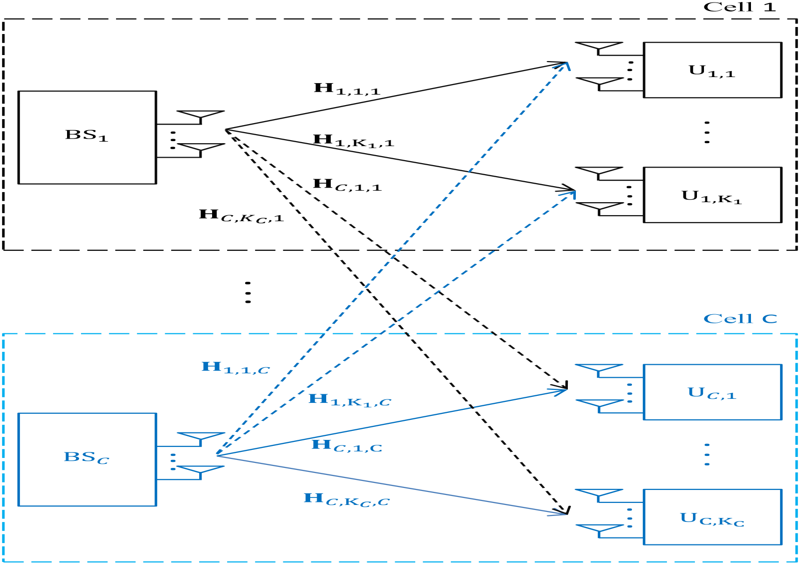
- A related formula is the Rayleigh quotient ($\lambda_k = \frac{p_k\,q_k}{\mathbf{q}^T\mathbf{p}}$)

$$\Delta = \max_{\mathbf{q}} \min_{\mathbf{p}} \frac{\mathbf{q}^T \mathbf{\Lambda}\,\mathbf{p}}{\mathbf{q}^T \mathbf{p}} \;\Rightarrow\; \mathbf{\Lambda}\,\mathbf{p} = \Delta\,\mathbf{p}\,,\; \mathbf{q}^T\mathbf{\Lambda} = \Delta\,\mathbf{q}^T$$

  which are the left and right Perron Frobenius eigen vectors.
- The beamformer optimization now only requires the dual UL powers q_k from either:
  - the left Perron Frobenius eigen vector $\mathbf{q}$, via Lagrangian duality, in particular $\mathbf{g}'_k = \arg\min_{\mathbf{g}_k} [\mathbf{q}^T\mathbf{\Lambda}]_k$,
  - by UL/DL duality from $\mathbf{p}$ and $SINR_k^{UL}(\mathbf{q}) = SINR_k^{DL}(\mathbf{p})$,
  - by iterating interference functions, which correspond to the power method for finding a largest eigen vector iteratively.

MIMO Interfering Broadcast Channel (IBC)

proper sumDoF bound: $C\,K\,\frac{M+N}{CK+1}$, C cells, K users/cell



-

Outline Coordinated BF w Perfect CSIT

- iterative local optimum finding
  - Weighted Sum MSE (WSMSE)
  - Minorization (Difference of Convex functions programming (DC))
  - Weighted Sum Unbiased MSE (WSUMSE)
- some options
  - with/without Receivers
  - power constraint enforcement (Lagrange multipliers)

MIMO IBC with Linear Tx/Rx, single stream

- IBC with C cells with a total of K users. System-wide user numbering: the $N_k \times 1$ Rx signal at user k in cell b_k is

$$\boldsymbol{y}_k = \underbrace{\boldsymbol{H}_{k,b_k}\,\boldsymbol{g}_k\,x_k}_{\text{signal}} + \underbrace{\sum_{\substack{i\neq k\\ b_i=b_k}} \boldsymbol{H}_{k,b_k}\,\boldsymbol{g}_i\,x_i}_{\text{intracell interf.}} + \underbrace{\sum_{j\neq b_k}\sum_{i:b_i=j} \boldsymbol{H}_{k,j}\,\boldsymbol{g}_i\,x_i}_{\text{intercell interf.}} + \boldsymbol{v}_k$$

  where x_k = intended (white, unit variance) scalar signal stream, $\boldsymbol{H}_{k,b_k} = N_k \times M_{b_k}$ channel from BS b_k to user k. BS b_k serves $K_{b_k} = \sum_{i:b_i=b_k} 1$ users. Noise whitened signal representation $\Rightarrow$ $\boldsymbol{v}_k \sim \mathcal{CN}(0, I_{N_k})$.

- The $M_{b_k} \times 1$ spatial Tx filter or beamformer (BF) is $\boldsymbol{g}_k$.
- Treating interference as noise, user k will apply a linear Rx filter $\boldsymbol{f}_k$ to maximize the signal power (diversity) while reducing any residual interference that would not have been (sufficiently) suppressed by the BS Tx. The Rx filter output is $\widehat{x}_k = \boldsymbol{f}_k^H \boldsymbol{y}_k$

$$\begin{aligned}\widehat{x}_k &= \boldsymbol{f}_k^H \boldsymbol{H}_{k,b_k}\,\boldsymbol{g}_k\,x_k + \sum_{i=1,\neq k}^{K} \boldsymbol{f}_k^H \boldsymbol{H}_{k,b_i}\,\boldsymbol{g}_i\,x_i + \boldsymbol{f}_k^H \boldsymbol{v}_k \\ &= \boldsymbol{f}_k^H \boldsymbol{h}_{k,k}\,x_k + \sum_{i\neq k} \boldsymbol{f}_k^H \boldsymbol{h}_{k,i}\,x_i + \boldsymbol{f}_k^H \boldsymbol{v}_k\end{aligned}$$

  where $\boldsymbol{h}_{k,i} = \boldsymbol{H}_{k,b_i}\,\boldsymbol{g}_i$ is the channel-Tx cascade vector.

Max Weighted Sum Rate (WSR)

- Weighted sum rate (WSR)

$$WSR = WSR(\boldsymbol{g}) = \sum_{k=1}^{K} u_k \, \ln \frac{1}{e_k}$$

where $\boldsymbol{g} = \{\boldsymbol{g}_k\}$, the u_k are rate weights

- MMSEs $e_k = e_k(\boldsymbol{g})$

$$\frac{1}{e_k} = 1 + \boldsymbol{g}_k^H \boldsymbol{H}_k^H \boldsymbol{R}_{\overline{k}}^{-1} \boldsymbol{H}_k \boldsymbol{g}_k = (1 - \boldsymbol{g}_k^H \boldsymbol{H}_k^H \boldsymbol{R}_k^{-1} \boldsymbol{H}_k \boldsymbol{g}_k)^{-1}$$

$$\boldsymbol{R}_k = \boldsymbol{R}_{\overline{k}} + \boldsymbol{H}_k \boldsymbol{g}_k \boldsymbol{g}_k^H \boldsymbol{H}_k^H \ , \ \boldsymbol{R}_{\overline{k}} = \sum_{i \neq k} \boldsymbol{H}_k \boldsymbol{g}_i \boldsymbol{g}_i^H \boldsymbol{H}_k^H + I_{N_k} \ ,$$

$\boldsymbol{R}_k$, $\boldsymbol{R}_{\overline{k}}$ = total, interference plus noise Rx cov. matrices resp.

- MMSE e_k obtained at the output $\widehat{x}_k = \boldsymbol{f}_k^H \boldsymbol{y}_k$ of the optimal (MMSE) linear Rx

$$\boldsymbol{f}_k = \boldsymbol{R}_k^{-1} \boldsymbol{H}_k \boldsymbol{g}_k \ .$$

From max WSR to min WSMSE

- For a general Rx filter $\boldsymbol{f}_k$ we have the MSE $e_k(\boldsymbol{f}_k, \boldsymbol{g})$

$$= (1 - \boldsymbol{f}_k^H \boldsymbol{H}_k \boldsymbol{g}_k)(1 - \boldsymbol{g}_k^H \boldsymbol{H}_k^H \boldsymbol{f}_k) + \sum_{i \neq k} \boldsymbol{f}_k^H \boldsymbol{H}_k \boldsymbol{g}_i \boldsymbol{g}_i^H \boldsymbol{H}_k^H \boldsymbol{f}_k + ||\boldsymbol{f}_k||^2$$
$$= 1 - \boldsymbol{f}_k^H \boldsymbol{H}_k \boldsymbol{g}_k - \boldsymbol{g}_k^H \boldsymbol{H}_k^H \boldsymbol{f}_k + \sum_i \boldsymbol{f}_k^H \boldsymbol{H}_k \boldsymbol{g}_i \boldsymbol{g}_i^H \boldsymbol{H}_k^H \boldsymbol{f}_k + ||\boldsymbol{f}_k||^2.$$

- The $WSR(\boldsymbol{g})$ is a non-convex and complicated function of $\boldsymbol{g}$. Inspired by [Christensen:TW1208], we introduced [Negro:ita10],[Negro:ita11] an augmented cost function, the Weighted Sum MSE,

$$WSMSE(\boldsymbol{g}, \boldsymbol{f}, w) = \sum_{k=1}^{K} u_k(w_k\, e_k(\boldsymbol{f}_k, \boldsymbol{g}) - 1 - \ln w_k) + \lambda(\sum_{k=1}^{K} ||\boldsymbol{g}_k||^2 - P)$$

  where λ = Lagrange multiplier and P = Tx power constraint.

- After optimizing over the aggregate auxiliary Rx filters $\boldsymbol{f}$ and weights w, we get the WSR back:

$$\min_{\boldsymbol{f}, w} WSMSE(\boldsymbol{g}, \boldsymbol{f}, w) = -WSR(\boldsymbol{g})$$

Alternating Minimization

- also called "cyclic minimization" or "block coordinate descent"
- Cost function to be minimized: $f(\boldsymbol{\theta})$
- Partition $\boldsymbol{\theta} = \boldsymbol{\theta}_{1:m} = \{\theta_1, \cdots, \theta_m\} = \{\boldsymbol{\theta}_{1:k-1}, \theta_k, \boldsymbol{\theta}_{k+1:m}\}$.
- Sweep through the partition. In sweep i:

$$\theta_k^{(i)} = \arg\min_{\theta_k} f(\boldsymbol{\theta}_{1:k-1}^{(i)}, \theta_k, \boldsymbol{\theta}_{k+1:m}^{(i-1)}) \;\; k = 1, \ldots, m$$

  Partitioning done in such a way that these reduced optimization problems are "easy".
- Guaranteed to converge to a local minimum:

$$f(\boldsymbol{\theta}_{1:k}^{(i)}, \boldsymbol{\theta}_{k+1:m}^{(i-1)}) \;\leq\; f(\boldsymbol{\theta}_{1:k-1}^{(i)}, \boldsymbol{\theta}_{k:m}^{(i-1)})$$

- Not necessary to do regular sweeps through the partition, can minimize in any order.

From max WSR to min WSMSE (2)

- Advantage augmented cost function: alternating optimization $\Rightarrow$ solving simple quadratic or convex functions

$$
\begin{aligned}
&\min_{w_k} WSMSE \;\Rightarrow\; w_k = 1/e_k \\
&\min_{\boldsymbol{f}_k} WSMSE \;\Rightarrow\; \boldsymbol{f}_k = (\sum_i \boldsymbol{H}_k \boldsymbol{g}_i \boldsymbol{g}_i^H \boldsymbol{H}_k^H + I_{N_k})^{-1} \boldsymbol{H}_k \boldsymbol{g}_k \\
&\min_{\boldsymbol{g}_k} WSMSE \;\Rightarrow\; \\
&\qquad \boldsymbol{g}_k = (\sum_i u_i w_i \boldsymbol{H}_i^H \boldsymbol{f}_i \boldsymbol{f}_i^H \boldsymbol{H}_i + \lambda I_M)^{-1} \boldsymbol{H}_k^H \boldsymbol{f}_k u_k w_k
\end{aligned}
$$

- UL/DL duality: optimal Tx filter g_k of the form of a MMSE linear Rx for the dual UL in which λ plays the role of Rx noise variance and $u_k w_k$ plays the role of stream variance.

Optimal Lagrange Multiplier(s) λ

- Solving dual problem for constrained optimization: optimize over $\boldsymbol{g}$ for each tested value of λ. Dual problem convex $\Rightarrow$ can do bisection for λ.
- (bisection) line search on $\sum_{k=1}^{K} ||\boldsymbol{g}_k||^2 - P = 0$ [Luo:SP0911].
- Or updated analytically as in [Negro:ita10],[Negro:ita11] by exploiting $\sum_k \boldsymbol{g}_k^H \frac{\partial WSMSE}{\partial \boldsymbol{g}_k^*} = 0$.
- This leads to the same result as in [Hassibi:TWC0906]: λ avoided by reparameterizing the BF to satisfy the power constraint: $\boldsymbol{g}_k = \sqrt{\frac{P}{\sum_{i=1}^{K} ||\boldsymbol{g}_i^{'}||^2}}\, \boldsymbol{g}_k^{'}$ with $\boldsymbol{g}_k^{'}$ now unconstrained

$$\text{SINR}_k = \frac{|\boldsymbol{f}_k \boldsymbol{H}_k \boldsymbol{g}_k^{'}|^2}{\sum_{i=1,\neq k}^{K} |\boldsymbol{f}_k \boldsymbol{H}_k \boldsymbol{g}_i^{'}|^2 + \frac{1}{P}||\boldsymbol{f}_k||^2 \sum_{i=1}^{K} ||\boldsymbol{g}_i^{'}||^2} .$$

- This leads to the same Lagrange multiplier expression obtained in [Christensen:TW1208] on the basis of a heuristic that was introduced in [Joham:isssta02] as was pointed out in [Negro:ita10].

Minorization (Difference of Convex functions (DC) programming, Successive Convex Approximation (SCA)) (single cell notation)

- Let $\boldsymbol{Q}_k = \boldsymbol{G}_k \boldsymbol{G}_k^H$ be the transmit covariance for user k $\Rightarrow$

$$WSR = \sum_{k=1}^{K} u_k [\ln\det(\boldsymbol{R}_k) - \ln\det(\boldsymbol{R}_{\overline{k}})]$$

with $\boldsymbol{R}_k = \boldsymbol{H}_k(\sum_i \boldsymbol{Q}_i)\boldsymbol{H}_k^H + I_{N_k}$, $\boldsymbol{R}_{\overline{k}} = \boldsymbol{H}_k(\sum_{i\neq k} \boldsymbol{Q}_i)\boldsymbol{H}_k^H + I_{N_k}$.

- Consider the dependence of WSR on $\boldsymbol{Q}_k$ alone:

$$WSR = u_k \ln\det(\boldsymbol{R}_{\overline{k}}^{-1}\boldsymbol{R}_k) + WSR_{\overline{k}}\,,\; WSR_{\overline{k}} = \sum_{i=1,\neq k}^{K} u_i \ln\det(\boldsymbol{R}_{\overline{i}}^{-1}\boldsymbol{R}_i)$$

where $\ln\det(\boldsymbol{R}_{\overline{k}}^{-1}\boldsymbol{R}_k)$ is concave in $\boldsymbol{Q}_k$ and $WSR_{\overline{k}}$ is convex in $\boldsymbol{Q}_k$. Since a linear function is simultaneously convex and concave, consider the first order Taylor series expansion in $\boldsymbol{Q}_k$ around $\widehat{\boldsymbol{Q}}$ (i.e. all $\widehat{\boldsymbol{Q}}_i$) with e.g. $\widehat{\boldsymbol{R}}_i = \boldsymbol{R}_i(\widehat{\boldsymbol{Q}})$, then

$$WSR_{\overline{k}}(\boldsymbol{Q}_k, \widehat{\boldsymbol{Q}}) \approx WSR_{\overline{k}}(\widehat{\boldsymbol{Q}}_k, \widehat{\boldsymbol{Q}}) - \text{tr}\{(\boldsymbol{Q}_k - \widehat{\boldsymbol{Q}}_k)\widehat{\mathbf{A}}_k\}$$

$$\widehat{\mathbf{A}}_k = -\left.\frac{\partial WSR_{\overline{k}}(\boldsymbol{Q}_k, \widehat{\boldsymbol{Q}})}{\partial \boldsymbol{Q}_k^T}\right|_{\widehat{\boldsymbol{Q}}_k, \widehat{\boldsymbol{Q}}} = \sum_{i=1,\neq k}^{K} u_i \boldsymbol{H}_i^H (\widehat{\boldsymbol{R}}_{\overline{i}}^{-1} - \widehat{\boldsymbol{R}}_i^{-1})\boldsymbol{H}_i$$

Minorization (DC) (2)

- Note that the linearized (tangent) expression for $WSR_{\overline{k}}$ constitutes a lower bound for it.
- Now, dropping constant terms, reparameterizing $Q_k = \boldsymbol{G}_k \boldsymbol{G}_k^H$ and performing this linearization for all users,

$$WSR(\boldsymbol{G}, \widehat{\boldsymbol{G}}) = \sum_{k=1}^{K} u_k \ln\det(I + \boldsymbol{G}_k^H \boldsymbol{H}_k^H \widehat{\boldsymbol{R}}_{\overline{k}}^{-1} \boldsymbol{H}_k \boldsymbol{G}_k) - \text{tr}\{\boldsymbol{G}_k^H (\widehat{\mathbf{A}}_k + \lambda I) \boldsymbol{G}_k\} + \lambda P$$

The gradient of this concave WSR lower bound is actually still the same as that of the original WSR or of the WSMSE criteria! Allows generalized eigenvector interpretation:

$$\boldsymbol{H}_k^H \widehat{\boldsymbol{R}}_{\overline{k}}^{-1} \boldsymbol{H}_k \boldsymbol{G}_k = (\widehat{\mathbf{A}}_k + \lambda I) \boldsymbol{G}_k \frac{1}{u_k} (I + \boldsymbol{G}_k^H \boldsymbol{H}_k^H \widehat{\boldsymbol{R}}_{\overline{k}}^{-1} \boldsymbol{H}_k \boldsymbol{G}_k)$$

or hence $\boldsymbol{G}_k' = V_{max}(\boldsymbol{H}_k^H \widehat{\boldsymbol{R}}_{\overline{k}}^{-1} \boldsymbol{H}_k, \widehat{\mathbf{A}}_k + \lambda I)$
which (MISO iCSIT case) is proportional to the "LMMSE" $\boldsymbol{g}_k$,
with max eigenvalue $\sigma_k = \sigma_{max}(\boldsymbol{H}_k^H \widehat{\boldsymbol{R}}_{\overline{k}}^{-1} \boldsymbol{H}_k, \widehat{\mathbf{A}}_k + \lambda I)$.

(DC) = Optimally Weighted SLNR (per stream approach)

- Again, minorization BF:

$$\boldsymbol{g}_k^{'} = V_{max}(\boldsymbol{H}_k^H \widehat{\boldsymbol{R}}_{\bar{k}}^{-1} \boldsymbol{H}_k, \sum_{i=1,\neq k}^{K} u_i \boldsymbol{H}_i^H (\widehat{\boldsymbol{R}}_{\bar{i}}^{-1} - \widehat{\boldsymbol{R}}_i^{-1}) \boldsymbol{H}_i + \lambda I)$$

- This can be viewed as an optimally weighted version of SLNR (Signal-to-Leakage-plus-Noise-Ratio) [Sayed:SP0507]

$$SLNR_k = \frac{||\boldsymbol{H}_k \boldsymbol{g}_k||^2}{\sum_{i\neq k} ||\boldsymbol{H}_i \boldsymbol{g}_k||^2 + \sum_i ||\boldsymbol{g}_i||^2/P} \text{ vs}$$

$$SINR_k = \frac{||\boldsymbol{H}_k \boldsymbol{g}_k||^2}{\sum_{i\neq k} ||\boldsymbol{H}_k \boldsymbol{g}_i||^2 + \sum_i ||\boldsymbol{g}_i||^2/P}$$

- SLNR takes as Tx filter

$$\boldsymbol{g}_k^{'} = V_{max}(\boldsymbol{H}_k^H \boldsymbol{H}_k, \sum_{i\neq k} \boldsymbol{H}_i^H \boldsymbol{H}_i + I)$$

[KimGiannakis:IT0511] Interference Aware WF

- Let $\sigma_k^{(1)} = \boldsymbol{g}_k^{'H} \boldsymbol{H}_k^H \widehat{\boldsymbol{R}}_{\overline{k}}^{-1} \boldsymbol{H}_k \boldsymbol{g}_k^{'}$ and $\sigma_k^{(2)} = \boldsymbol{g}_k^{'H} \widehat{\mathbf{A}}_k \boldsymbol{g}_k^{'}$.
- The advantage of this formulation is that it allows straightforward power adaptation: substituting $\boldsymbol{g}_k = \sqrt{p_k}\, \boldsymbol{g}_k^{'}$ yields

$$WSR = \lambda P + \sum_{k=1}^{K} \{ u_k \ln(1 + p_k \sigma_k^{(1)}) - p_k(\sigma_k^{(2)} + \lambda) \}$$

which leads to the following interference leakage aware water filling

$$p_k = \left(\frac{u_k}{\sigma_k^{(2)} + \lambda} - \frac{1}{\sigma_k^{(1)}} \right)^+ .$$

- For a given λ, $\boldsymbol{g}$ needs to be iterated till convergence.
- And λ can be found by duality (line search):

$$\min_{\lambda \geq 0} \max_{\boldsymbol{g}} \lambda P + \sum_k \{ u_k \ln\det(\boldsymbol{R}_{\overline{k}}^{-1} \boldsymbol{R}_k) - \lambda p_k \} = \min_{\lambda \geq 0} WSR(\lambda).$$

- In contrast to this optimization duality in [KimGiannakis:IT0511], we propose to jointly solve for $\boldsymbol{p}$,λ (for given $\boldsymbol{g}$) as in standard water filling.

Minorization: Interference Aware WF (Multi-Stream)

- Let $\Sigma_k^{(1)} = \boldsymbol{G}_k^{'H} \boldsymbol{H}_k^H \widehat{\boldsymbol{R}}_{\overline{k}}^{-1} \boldsymbol{H}_k \boldsymbol{G}_k^{'}$ and $\Sigma_k^{(2)} = \boldsymbol{G}_k^{'H} \widehat{\mathbf{A}}_k \boldsymbol{G}_k^{'}$.
- The advantage of this formulation is that it allows straightforward power adaptation: substituting $\boldsymbol{G}_k = \boldsymbol{G}_k^{'} P_k^{\frac{1}{2}}$ yields

$$WSR = \lambda P + \sum_{k=1}^{K} [u_k \ln\det(I + P_k \Sigma_k^{(1)}) - \text{tr}\{P_k(\Sigma_k^{(2)} + \lambda I)\}]$$

which leads to the following interference leakage aware water filling (jointly for the P_k and λ)

$$P_k = \left(u_k(\Sigma_k^{(2)} + \lambda I)^{-1} - \Sigma_k^{-(1)}\right)^{+}, \quad \sum_k \text{tr}\{P_k\} = P\,.$$

- In [KimGiannakis:IT0511]: for given λ, iterate $\boldsymbol{g}$ till convergence and λ found by duality (line search):

$$\min_{\lambda \geq 0} \max_{\boldsymbol{g}} \lambda P + \sum_k \{u_k \ln\det(\boldsymbol{R}_{\overline{k}}^{-1} \boldsymbol{R}_k) - \lambda p_k\} = \min_{\lambda \geq 0} WSR(\lambda).$$

WSMSE-DC BF relation

- min WSMSE iteration $(i+1)$

$$
\begin{aligned}
\mathbf{A}_k^{(i)} &= \sum_j u_j w_j^{(i)} \boldsymbol{H}_i^H \boldsymbol{f}_i^{(i)} \boldsymbol{f}_j^{(i)H} \boldsymbol{H}_j + \lambda^{(i)} I_M \\
\boldsymbol{g}_k^{(i+1)} &= (\mathbf{A}_k^{(i)})^{-1} \boldsymbol{H}_k^H \boldsymbol{f}_k^{(i)} u_k w_k^{(i)} \\
&= (\mathbf{A}_k^{(i)})^{-1} \mathbf{B}_k^{(i)} \ \boldsymbol{g}_k^{(i)} \ u_k w_k^{(i)} \\
\mathbf{B}_k^{(i)} &= \boldsymbol{H}_k^H \boldsymbol{R}_k^{-(i)} \boldsymbol{H}_k
\end{aligned}
$$

WSMSE does one power iteration of DC !!

$$
\boldsymbol{g}_k^{(i+1)} = V_{max}\{(\mathbf{A}_k^{(i)})^{-1} \mathbf{B}_k^{(i)}\}
$$

High/Low SNR Behavior

- At high SNR, max WSR BF converges to ZF solutions with uniform power

$$\boldsymbol{g}_k^H = \boldsymbol{f}_k \boldsymbol{H}_k P^{\perp}_{(\boldsymbol{f}\boldsymbol{H})^H_{\overline{k}}} / ||\boldsymbol{f}_k \boldsymbol{H}_k P^{\perp}_{(\boldsymbol{f}\boldsymbol{H})^H_{\overline{k}}}||$$

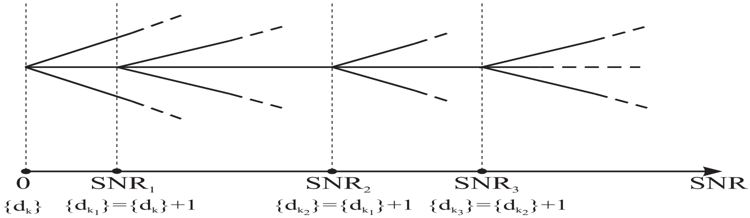
  where $P^{\perp}_{\boldsymbol{X}} = I - P_{\boldsymbol{X}}$ and $P_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}^H\boldsymbol{X})^{-1}\boldsymbol{X}^H$ projection matrices
  $(\boldsymbol{f}\boldsymbol{H})_{\overline{k}}$ denotes the (up-down) stacking of $\boldsymbol{f}_i \boldsymbol{H}_i$ for users
  $i = 1, \ldots, K,\ i \neq k$.
- At low SNR, matched filter for user with largest $||\boldsymbol{H}_k||_2$
  (max singular value)

Deterministic Annealing

- At high SNR: max WSR solutions are ZF. When ZF is possible (IA feasible), multiple ZF solutions typically exist.
  Homotopy on the MIMO channel SVD:
  $$H_{ji} = \sum_{k=1}^{d} \sigma_{jik}\mathbf{u}_{jik}\mathbf{v}_{jik}^H + t \sum_{k=d+1} \sigma_{jik}\mathbf{u}_{jik}\mathbf{v}_{jik}^H$$
  The IA (ZF) condition for rank 1 link $i-j$ can be written as
  $$\sigma_{ji}\mathbf{f}_j^H\mathbf{u}_{ji}\mathbf{v}_{ji}^H\mathbf{g}_i = 0$$
- Two configurations are possible: $\mathbf{f}_j^H\mathbf{u}_{ji} = 0$ or $\mathbf{v}_{ji}^H\mathbf{g}_i = 0$
  Either the Tx or the Rx suppresses one particular interfering stream
- These different ZF solutions are the possible local optima for max WSR at infinite SNR. By homotopy, this remains the number of max WSR local optima as the SNR decreases from infinity. As the SNR decreases further, a stream for some user may get turned off until only a single stream remains at low SNR. Hence, the number of local optima reduces as streams disappear at finite SNR.
- At intermediate SNR, the number of streams may also be larger than the DoF though.

Deterministic Annealing (2)

- Homotopy for finding global optimum: at low SNR, noise dominates interference $\Rightarrow$ optimal: one stream per power constraint, matched filter Tx/Rx. Gradually increasing SNR allows lower SNR solution to be in region of attraction of global optimum at next higher SNR.
  Phase transitions: add a stream.
- As a corollary, in the MISO case, the max WSR optimum is unique, since there is only one way to perform ZF BF.



Difference of Convex Functions vs Minorization

- Difference of Convex functions: linearize convex part in terms of Tx covariance matrices $\boldsymbol{Q}_k$ to make it concave
- afterwards work with BF in $\boldsymbol{Q}_k = \boldsymbol{g}_k \boldsymbol{g}_k^H$
- but the linearization in $\boldsymbol{Q}_k$ does not correspond to second-order Taylor series or any precise development in $\boldsymbol{g}_k$

- other interpretation: minorization: replace cost function to be maximized by one below it that touches the original one in one point [Stoica:SPmagJan04]
- specifically: matrix version of $-(x-1) \leq -\ln(x)\,,\ x > 0:$ (tangent above $\ln()$ curve) Itakura-Saito distance in AR modeling ($x =$ ratio of true spectrum and AR model spectrum)
  $-tr\{(R')^{-1}R - I\} \leq -\ln\det((R')^{-1}R)\ ,\ \ R = R' + \Delta R$
- minorized cost function can be optimized with any parameterization

MIMO Tx/Rx Design w Other Utility Functions

- in all cases (e.g. also SINR balancing),

  Rx filter = LMMSE
  Tx filter = LMMSE in dual uplink

- influence of precise utility function is in the design of the actual & dual stream powers and noise variances

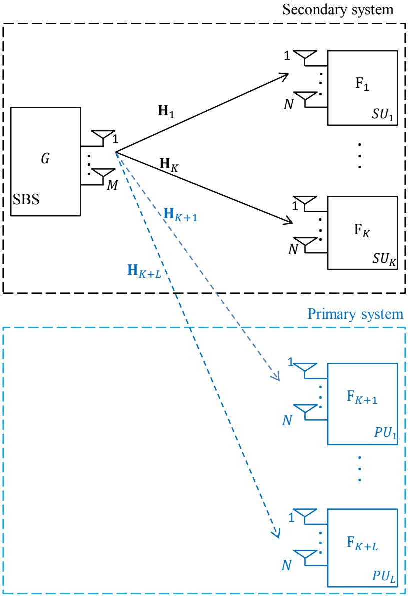
Optimization Discussion

- WSMSE vs Minorization:
  - WSMSE adapts both BF direction and stream powers
    Minorization adapts separately stream powers w Water Filling
  - in MIMO (vs MISO): Minorization computes generalized eigenvectors whereas WSMSE does Power Method to iteratively converge towards max generalized eigenvector

    MISO: max generalized eigenvector of rank 1 signal matrix $\Rightarrow$ MMSE solution
  - WSMSE introduces Rxs (hence have to know # streams) wheras Minorization can work with only Txs and find the feasible # streams automatically
- Minorization can also be formulated with Rxs, then still retain the power Water Filling

Application: WSR in Underlay Cognitive MU-MIMO (CBC)

Cognitive Broadcast Channel



CBC Constrained WSR Optimization

- Tx covariance matrix $\boldsymbol{Q} = \sum_{k=1}^{K} \boldsymbol{g}_k \boldsymbol{g}_k^H$

$$\max_{\boldsymbol{g}} \{WSR = WSR(\boldsymbol{g}) = \sum_{k=1}^{K} u_k \ln \frac{1}{e_k}\} \text{ s.t. } \mathrm{tr}\{\boldsymbol{Q}\} \leq P \,,\ \mathrm{tr}\{\boldsymbol{Q}\, \boldsymbol{H}_{K+l}^H \boldsymbol{H}_{K+l}\} \leq P_l,$$

- DC programming Lagrangian ($P_{L+1} = P$)

$$WSR(\boldsymbol{g}, \widehat{\boldsymbol{g}}, \boldsymbol{\lambda}) = \sum_{l=1}^{L+1} \lambda_l P_l + \sum_{k=1}^{K} u_k \ln(1 + \boldsymbol{g}_k^H \widehat{\mathbf{B}}_k \boldsymbol{g}_k) - \boldsymbol{g}_k^H (\widehat{\mathbf{A}}_k + \sum_{l=1}^{L+1} \lambda_l \boldsymbol{D}_l) \boldsymbol{g}_k$$

$$\widehat{\mathbf{A}}_k = \sum_{i=1, \neq k}^{K} u_i \boldsymbol{H}_i^H \boldsymbol{H}_i (\widehat{\boldsymbol{R}}_{\overline{i}}^{-1} - \widehat{\boldsymbol{R}}_i^{-1}), \widehat{\mathbf{B}}_k = \boldsymbol{H}_k^H \boldsymbol{H}_k \widehat{\boldsymbol{R}}_{\overline{k}}^{-1}, \boldsymbol{D}_l = \boldsymbol{H}_{K+l}^H \boldsymbol{H}_{K+l}$$

- max generalized eigen vector solution

$$\widehat{\mathbf{B}}_k\, \boldsymbol{g}_k = \frac{1 + \boldsymbol{g}_k^H \widehat{\mathbf{B}}_k \boldsymbol{g}_k}{u_k} (\widehat{\mathbf{A}}_k + \sum_{l=1}^{L+1} \lambda_l \boldsymbol{D}_l) \boldsymbol{g}_k \Rightarrow \boldsymbol{g}_k^{'} = V_{max}(\widehat{\mathbf{B}}_k, \widehat{\mathbf{A}}_k + \sum_{l} \lambda_l \boldsymbol{D}_l)$$

- ILA WF: let $\sigma_k^{(1)} = \boldsymbol{g}_k^{'H} \widehat{\mathbf{B}}_k \boldsymbol{g}_k^{'}$, $\sigma_k^{(2)} = \boldsymbol{g}_k^{'H} \widehat{\mathbf{A}}_k \boldsymbol{g}_k^{'}$, $\sigma_{k,l} = \boldsymbol{g}_k^{'H} \boldsymbol{D}_l \boldsymbol{g}_k^{'}$,
  $\boldsymbol{g}_k = \sqrt{p_k}\, \boldsymbol{g}_k^{'}$

$$\min_{\lambda \geq 0} \max_{\boldsymbol{p} \geq 0} \{WSR = \sum_{l} \lambda_l P_l + \sum_{k=1}^{K} [u_k \ln(1 + p_k \sigma_k^{(1)}) - p_k(\sigma_k^{(2)} + \sum_{l} \lambda_l \sigma_{k,l})]\}$$

CBC Constrained WSR Optimization (2)

- Interference Leakage Aware Water Filling (ILA WF) solution

$$p_k = \left(\frac{u_k}{\sigma_k^{(2)} + \sum_l \lambda_l \sigma_{k,l}} - \frac{1}{\sigma_k^{(1)}} \right)^+$$

where the λ need to be increased beyond zero to satisfy the power/interference constraints. This can be done by the ellipsoid algorithm or with a greedy approach (alternating bisection method) that focuses on the constraint violation terms in order of decreasing sensitivity.

- Contrast this limited $\min_{\lambda \geq 0} \max_{\boldsymbol{p} \geq 0}$ for given $\boldsymbol{g}'$ with the normal constrained optimization duality

$$\min_{\lambda \geq 0} \max_{\boldsymbol{g}} \sum_l \lambda_l P_l + \sum_k \{ u_k \ln\det(\boldsymbol{R}_{\overline{k}}^{-1} \boldsymbol{R}_k) - p_k \sum_l \lambda_l \sigma_{k,l} \}$$

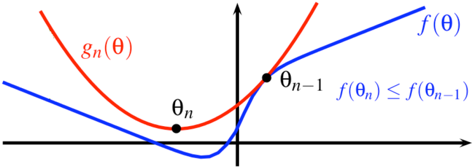
which requires to solve for $\boldsymbol{g}$ for every vector λ to be evaluated!

- The same water filling trick can be applied in other scenarios with multiple power constraints, such as in the case of per antenna power constraints.

WSR Minorizer Maximization Unification

- WSMSE: in terms of $\boldsymbol{E}_k$
- DC programming: in terms of $\boldsymbol{Q}_k$
- rate minorizer in terms of $\boldsymbol{R}_{\overline{k}}$

Majorization Minimization (or Minorizer Maximization)



- Cost function to be minimized: $f(\boldsymbol{\theta})$. At iteration $n{-}1$: have $\boldsymbol{\theta}^{(n-1)}$.
- A majorizer at iteration $n{-}1$ is a function $g_n(\boldsymbol{\theta})$ such that

$$g_n(\boldsymbol{\theta}^{(n-1)}) = f(\boldsymbol{\theta}^{(n-1)})$$
$$g_n(\boldsymbol{\theta}) \geq f(\boldsymbol{\theta})\,,\ \forall\,\boldsymbol{\theta}$$

where $g_n(\boldsymbol{\theta})$ is possibly convex or in any case "easy" to minimize or just decrease.

- Then

$$\boldsymbol{\theta}^{(n)} = \arg\min_{\boldsymbol{\theta}} g_n(\boldsymbol{\theta})$$
$$f(\boldsymbol{\theta}^{(n)}) \leq g_n(\boldsymbol{\theta}^{(n)}) \leq g_n(\boldsymbol{\theta}^{(n-1)}) = f(\boldsymbol{\theta}^{(n-1)})$$

Hence guaranteed to converge to a local minimum.

WSMSE as WSR Minorizer

- The rate of user k can also be represented as

$$r_k(\boldsymbol{G}) = \ln\det(\boldsymbol{E}_k^{-1}(\boldsymbol{G})) = \max_{\boldsymbol{W}_k,\boldsymbol{F}_k} r_k^l(\boldsymbol{W}_k,\boldsymbol{F}_k,\boldsymbol{G}),$$

$$r_k^l = \ln\det(\boldsymbol{W}_k) - \mathrm{tr}\{\boldsymbol{W}_k\,\boldsymbol{E}_k(\boldsymbol{F}_k,\boldsymbol{G}) - \boldsymbol{I}_{d_k}\}$$

where

$$\begin{aligned}\boldsymbol{E}_k(\boldsymbol{F}_k,\boldsymbol{G}) &= \mathbb{E}\left[(\hat{\mathbf{s}}_k - \mathbf{s}_k)(\hat{\mathbf{s}}_k - \mathbf{s}_k)^{\mathrm{H}}\right] \\ &= \boldsymbol{I} - \boldsymbol{F}_k^H\,\boldsymbol{H}_{k,b_k}\,\boldsymbol{G}_k - \boldsymbol{G}_k^H\,\boldsymbol{H}_{k,b_k}^H\,\boldsymbol{F}_K + \boldsymbol{F}_k^H\,\boldsymbol{R}_k\,\boldsymbol{F}_k\end{aligned}$$

and optimal $\boldsymbol{W}_k = \boldsymbol{E}_k^{-1}$, $\boldsymbol{F}_k = \boldsymbol{R}_k^{-1}\boldsymbol{H}_{k,b_k}\boldsymbol{G}_k$.

- classical viewpoint:

$$\max_{\boldsymbol{G}} \sum_k u_k\, r_k(\boldsymbol{G}) = \max_{\boldsymbol{W},\boldsymbol{F},\boldsymbol{G}} \sum_k u_k\, r_k^l(\boldsymbol{W}_k,\boldsymbol{F}_k,\boldsymbol{G})$$

augmented parameterization and alternating optimization

- minorizer viewpoint:

$r_k(\boldsymbol{G}) \geq r_k^l(\boldsymbol{W}_k,\boldsymbol{F}_k,\boldsymbol{G})$ equality for $\boldsymbol{G} = \boldsymbol{G}'$ if $\boldsymbol{W}_k = \boldsymbol{W}_k(\boldsymbol{G}')$, $\boldsymbol{F}_k = \boldsymbol{F}_k(\boldsymbol{G}')$

WSR Minorizer in terms of $\boldsymbol{R}_{\overline{k}}$

- As for the MSE based minorizer, this is a minorizer per rate r_k
- r_k is convex in $\boldsymbol{R}_{\overline{k}}$ (as it was in $\boldsymbol{E}_k$), hence

$$r_k(\boldsymbol{G}) \geq r_k^l(\boldsymbol{G}) = r_k(\boldsymbol{G}') + \text{tr}\{\frac{\partial r_k}{\partial \boldsymbol{R}_{\overline{k}}^T}(\boldsymbol{G}')\,(\boldsymbol{R}_{\overline{k}}(\boldsymbol{G}) - \boldsymbol{R}_{\overline{k}}(\boldsymbol{G}'))\}$$

which leads to, with $\boldsymbol{R}'_{\overline{k}} = \boldsymbol{R}_{\overline{k}}(\boldsymbol{G}')$ etc., $\mathbf{S}_k(\boldsymbol{G}_k) = \boldsymbol{H}_{k,b_k}\boldsymbol{G}_k\boldsymbol{G}_k^H\boldsymbol{H}_{k,b_k}^H$,

$$r_k(\boldsymbol{G}) \geq r_k^l(\boldsymbol{G}) = \ln\det(\boldsymbol{I} + \boldsymbol{R}'_{\overline{k}}\mathbf{S}_k(\boldsymbol{G}_k)) + \text{tr}\{\boldsymbol{W}_k^R(\boldsymbol{R}_{\overline{k}}(\boldsymbol{G}) - \boldsymbol{R}'_{\overline{k}})\}$$
$$\boldsymbol{W}_k^R = (\boldsymbol{R}'_{\overline{k}})^{-1} - (\boldsymbol{R}'_k)^{-1}, \boldsymbol{R}'_k = \boldsymbol{R}'_{\overline{k}} + \mathbf{S}'_k$$

- This leads to the same weighted sum rate result as the DC programming approach, but with a minorizer per individual rate instead of for the overall sum rate, with a minorizer in terms of $\boldsymbol{R}_{\overline{k}}$ instead of $\boldsymbol{Q}_k = \boldsymbol{G}_k\boldsymbol{G}_k^H$.
- Decoupled minorizer per user, can be optimized in parallel.
- In contrast to MSE $\boldsymbol{E}_k$ based minorization, no Rx needs to be introduced here.
- The $\boldsymbol{R}_{\overline{k}}$ based minorizer can be extended to imperfect CSIT and for rate balancing (in which case the power p_k needs to factored out).