Exploiting Large/Massive MIMO
Deterministic Limits for
Cooperative Beamforming with Partial CSIT

Dirk Slock

Communication Systems Department, EURECOM

Collaboration with Wassim Tabikh, Yi Yuan-Wu, Kalyana Gopala

TIW 2016
Overall Outline

- infinite SNR: MIMO Interference Alignment
- perfect CSIT coordinated beamforming designs
- partial CSIT deterministic approximations
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 (IFC)
\[ d = \sum_{k=1}^{K} d_k \]
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.
(compressed) SVD:

\[ H = F D' G' H = F [D \ 0] \begin{bmatrix} G & G'' \end{bmatrix}^H = F D G^H \Rightarrow F^H H G = D \]

- \( F_k^H : d_k \times N_k, \ H_{ki} : N_k \times M_i, \ G_i : M_i \times d_i \)

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

\( F^H, \ G \) 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 \( \Rightarrow F^H H G = \text{diagonal} \).
MIMO Interfering Broadcast Channel (IBC)

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


- HetNets: reverse time division duplex (R-TDD): some cells in uplink mode and the rest of them in downlink simultaneously

- Necessary feasibility condition for interference alignment in the multi-cell R-TDD scenario, which is then specialized to the particular case of symmetric demands and antenna distribution.

- For those symmetric networks for which the properness condition holds with equality, R-TDD does not improve the DoF performance of conventional synchronous TDD systems.

- Nevertheless, simulation results indicate that, in more asymmetric scenarios, significant DoF benefits can be achieved by applying the R-TDD approach.
Overall Outline

- infinite SNR: MIMO Interference Alignment
- perfect CSIT coordinated beamforming designs
- partial CSIT deterministic approximations
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)
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

\[
y_k = \sum_{i \neq k} \sum_{b_i = b_k} H_{k,b_k} g_i x_i + \sum_{j \neq b_k} \sum_{i : b_i = j} H_{k,j} g_i x_i + v_k
\]

where \( x_k = \) intended (white, unit variance) scalar signal stream, \( H_{k,b_k} = N_k \times M_{b_k} \) 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 \( \Rightarrow \ v_k \sim \mathcal{CN}(0, I_{N_k}) \).

The \( M_{b_k} \times 1 \) spatial Tx filter or beamformer (BF) is \( g_k \).

Treating interference as noise, user \( k \) will apply a linear Rx filter \( f_k \) 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

\[
\hat{x}_k = f_k^H y_k
\]

\[
\hat{x}_k = f_k^H H_{k,b_k} g_k x_k + \sum_{i = 1, \neq k} f_k^H H_{k,b_i} g_i x_i + f_k^H v_k
\]

\[
= f_k^H h_{k,k} x_k + \sum_{i \neq k} f_k^H h_{k,i} x_i + f_k^H v_k
\]

where \( h_{k,i} = H_{k,b_i} g_i \) is the channel-Tx cascade vector.
Max Weighted Sum Rate (WSR)

- Weighted sum rate (WSR)

\[
WSR = WSR(g) = \sum_{k=1}^{K} u_k \ln \frac{1}{e_k}
\]

where \( g = \{g_k\} \), the \( u_k \) are rate weights

- MMSEs \( e_k = e_k(g) \)

\[
\frac{1}{e_k} = 1 + g_k^H H_k^H R_k^{-1} H_k g_k = (1 - g_k^H H_k^H R_k^{-1} H_k g_k)^{-1}
\]

\[
R_k = R_k^- + H_k g_k g_k^H H_k^H, \quad R_k^- = \sum_{i \neq k} H_k g_i g_i^H H_k^H + I_{N_k},
\]

\( R_k, R_k^- = \) total, interference plus noise Rx cov. matrices resp.

- MMSE \( e_k \) obtained at the output \( \hat{x}_k = f_k^H y_k \) of the optimal (MMSE) linear Rx

\[
f_k = R_k^{-1} H_k g_k.
\]
From max WSR to min WSMSE

For a general Rx filter $f_k$ we have the MSE

$$e_k(f_k, g) = (1 - f_k^H H_k g_k)(1 - g_k^H H_k^H f_k) + \sum_{i \neq k} f_k^H H_k g_i g_i^H H_k^H f_k + \|f_k\|^2$$

$$= 1 - f_k^H H_k g_k - g_k^H H_k^H f_k + \sum_{i} f_k^H H_k g_i g_i^H H_k^H f_k + \|f_k\|^2.$$

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

$$= \sum_{k=1}^{K} u_k (w_k e_k(f_k, g) - \ln w_k) + \lambda \left( \sum_{k=1}^{K} \|g_k\|^2 - P \right)$$

where $\lambda$ = Lagrange multiplier and $P$ = Tx power constraint.

After optimizing over the aggregate auxiliary Rx filters $f$ and weights $w$, we get the WSR back:

$$\min_{f, w} WSMSE(g, f, w) = -WSR(g) + \sum_{k=1}^{K} u_k$$
Advantage augmented cost function: alternating optimization
⇒ solving simple quadratic or convex functions

\[
\begin{align*}
\min_{w_k} WSMSE & \Rightarrow w_k = \frac{1}{e_k} \\
\min_{f_k} WSMSE & \Rightarrow f_k = \left( \sum_i H_k g_i g_i^H H_k^H + I_{N_k} \right)^{-1} H_k g_k \\
\min_{g_k} WSMSE & \Rightarrow \\
& g_k = \left( \sum_i u_i w_i H_i^H f_i f_i^H H_i + \lambda I_M \right)^{-1} H_k^H f_k u_k w_k
\end{align*}
\]

UL/DL duality: optimal Tx filter \( g_k \) of the form of a MMSE linear Rx for the dual UL in which \( \lambda \) plays the role of Rx noise variance and \( u_k w_k \) plays the role of stream variance.
Let $Q_k = g_k g_k^H$ be the transmit covariance for stream $k$ ⇒

$$WSR = \sum_{k=1}^{K} u_k [\ln \det(R_k) - \ln \det(R_k^-)]$$

with $R_k = H_k (\sum_i Q_i) H_k^H + I_{N_k}$, $R_k^- = H_k (\sum_{i\neq k} Q_i) H_k^H + I_{N_k}$.

Consider the dependence of WSR on $Q_k$ alone:

$$WSR = u_k \ln \det(R_k^{-1} R_k) + WSR_k^- , \quad WSR_k^- = \sum_{i=1,\neq k}^{K} u_i \ln \det(R_i^{-1} R_i)$$

where $\ln \det(R_k^{-1} R_k)$ is concave in $Q_k$ and $WSR_k^-$ is convex in $Q_k$. Since a linear function is simultaneously convex and concave, consider the first order Taylor series expansion in $Q_k$ around $\hat{Q}$ (i.e. all $\hat{Q}_i$) with e.g. $\hat{R}_i = R_i(\hat{Q})$, then

$$WSR_k(Q_k, \hat{Q}) \approx WSR_k^- (\hat{Q}_k, \hat{Q}) - \text{tr}\{(Q_k - \hat{Q}_k) \hat{A}_k\}$$

$$\hat{A}_k = - \left. \frac{\partial WSR_k(Q_k, \hat{Q})}{\partial Q_k} \right|_{\hat{Q}_k, \hat{Q}} = \sum_{i=1,\neq k}^{K} u_i H_i^H (\hat{R}_i^{-1} - \hat{R}_i^{-1}) H_i$$
Note that the linearized (tangent) expression for \( WSR_k \) constitutes a lower bound for it.

Now, dropping constant terms, reparameterizing \( Q_k = g_k g_k^H \) and performing this linearization for all users,

\[
WSR(g, \hat{g}) = \sum_{k=1}^{K} u_k \ln(1 + g_k^H \hat{H}_k^H \hat{R}_k^{-1} H_k g_k) - g_k^H (\hat{A}_k + \lambda I) g_k + \lambda P.
\]

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:

\[
H_k^H \hat{R}_k^{-1} H_k g_k = \frac{1 + g_k^H \hat{H}_k^H \hat{R}_k^{-1} H_k g_k}{u_k} (\hat{A}_k + \lambda I) g_k
\]

or hence \( g_k' = V_{\text{max}}(H_k^H \hat{R}_k^{-1} H_k, \hat{A}_k + \lambda I) \)

which is proportional to the "LMMSE" \( g_k \),

with max eigenvalue \( \sigma_k = \sigma_{\text{max}}(H_k^H \hat{R}_k^{-1} H_k, \hat{A}_k + \lambda I) \).
Again, [KimGiannakis:IT0511] BF:

\[ g'_k = V_{\text{max}} (H_k^H \hat{R}_k^{-1} H_k, \sum_{i=1, i \neq k}^{K} u_i H_i^H (\hat{R}_i^{-1} - \hat{R}_i^{-1}) H_i + \lambda I) \]

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

\[
\text{SLNR}_k = \frac{||H_k g_k||^2}{\sum_{i \neq k} ||H_i g_k||^2 + \sum_i ||g_i||^2 / P} \quad \text{vs}
\]

\[
\text{SINR}_k = \frac{||H_k g_k||^2}{\sum_{i \neq k} ||H_k g_i||^2 + \sum_i ||g_i||^2 / P}
\]

SLNR takes as Tx filter

\[ g'_k = V_{\text{max}} (H_k^H H_k, \sum_{i \neq k} H_i^H H_i + I) \]
Let $\sigma^{(1)}_k = g_k^\prime H_k^H \hat{R}_k^{-1} H_k g_k^\prime$ and $\sigma^{(2)}_k = g_k^\prime \hat{A}_k g_k^\prime$.

The advantage of this formulation is that it allows straightforward power adaptation: substituting $g_k = \sqrt{p_k} g_k^\prime$ yields

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

which leads to the following interference leakage aware water filling

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

For a given $\lambda$, $g$ needs to be iterated till convergence.

And $\lambda$ can be found by duality (line search):

$$\min_{\lambda \geq 0} \max_g \{ \lambda P + \sum_k \{ u_k \ln \det(\hat{R}_k^{-1} R_k) - \lambda p_k \} = \min_{\lambda \geq 0} WSR(\lambda) \}.$$
From Max WSR to Min Weighted Sum Unbiased MSE (WSUMSE)

For the Rx output $\hat{x}_k$ to be an unbiased estimator for the Tx signal $x_k$, we require

$$E|x_k \hat{x}_k = x_k \Rightarrow f_k^H H_k, b_k g_k = 1.$$ 

If the Tx/Rx filters satisfy the unbiasedness constraint, then we get the Unbiased MSE (UMSE)

$$e_k^u(f_k, g) = \sum_{i \neq k} f_k^H H_k, b_i g_i^H H_k, b_i f_k + \| f_k \|^2.$$ 

In the complex case, it is more convenient to work with the alternative unbiasedness constraint

$$|f_k^H H_k, b_k g_k|^2 = 1$$

in which the phase uncertainty does not affect SINR or Gaussian capacity.
Weighted Sum UMSE (WSUMSE)

- augmented cost function:

\[
\text{WSUMSE}(g, f, w_1, w_2) = \sum_{k=1}^{K} u_k (w_{1,k} e_k^u(f_k, g) - \ln w_{1,k} - w_{2,k} (e_k^u(f_k, g) + 1) + \ln w_{2,k}) + \sum_{k=1}^{K} \mu_k (1 - |f_k^H H_k b_k g_k|^2) + \sum_{i=1}^{C} \lambda_i \left( \sum_{k : b_k = i} ||g_k||^2 - P_i \right)
\]

where the \( \lambda_i, \mu_k \) are Lagrange multipliers.

- After optimizing over the aggregate auxiliary Rx filters \( f \) and weights \( w_1, w_2 \), we get the WSR back:

\[
\min_{f} \min_{w_1} \max_{w_2} \text{WSUMSE}(g, f, w_1, w_2) = -WSR(g)
\]

where \( WSR = \sum_k u_k \ln(1 + \frac{1}{e_k^u}) \).
alternating optimization:

\[
\begin{align*}
\min_{w_{1,k}} \max_{w_{2,k}} WSUMSE & \Rightarrow w_{1,k} = 1/e_k^u, \quad w_{2,k} = 1/(e_k^u + 1) \\
\min_{f_k} WSUMSE & \Rightarrow f_k = R_k^{-1} H_{k,b_k} g_k \mu_k^g H_{k,b_k}^H f_k \\
\min_{g_k} WSUMSE & \Rightarrow g_k = (T_k^{-1} + \lambda_{b_k} I_M)^{-1} H_{k,b_k}^H f_k \mu_k' f_k^H H_{k,b_k} g_k
\end{align*}
\]

where \( w_k = w_{1,k} - w_{2,k} > 0 \). Note that \(|f_k^H H_{k,b_k} g_k| = 1\).

We can choose the phase such that \( f_k^H H_{k,b_k} g_k = 1 \). If we also reparameterize \( \mu_k = u_k w_k \mu_k' \), where the \( \mu_k \) are chosen to satisfy \( f_k^H H_{k,b_k} g_k = 1 \), then we can rewrite

\[
\begin{align*}
\beta_k &= R_k^{-1} H_{k,b_k} g_k \mu_k' \\
g_k &= (T_k^{-1} + \lambda_{b_k} I_M)^{-1} H_{k,b_k}^H f_k u_k w_k \mu_k'
\end{align*}
\]

which are proportional to the Tx/Rx found by the WSMSE approach.
Optimal Lagrange Multiplier $\lambda$

- (bisection) line search on $\sum_{k=1}^{K} \|g_k\|^2 - P = 0$ [Luo:SP0911].

- Or updated analytically as in [Negro:ita10],[Negro:ita11] by exploiting $\sum_k g_k^H \partial WSMSE/\partial g_k^* = 0$.

- This leads to the same result as in [Hassibi:TWC0906]: $\lambda$ avoided by reparameterizing the BF to satisfy the power constraint: $g_k = \sqrt{\frac{P}{\sum_{i=1}^{K} \|g'_i\|^2}} g'_k$ with $g'_k$ now unconstrained.

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

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

\[ g_k^H = f_k H_k P_{\perp}^H / \| f_k H_k P_{\perp}^H \| \]

where \( P_{\perp}^X = I - P_X \) and \( P_X = X (X^H X)^{-1} X^H \) projection matrices

\( (fH)_k^H \) denotes the (up-down) stacking of \( f_i H_i \) for users \( i = 1, \ldots, K, i \neq k \).

At low SNR, matched filter for user with largest \( \| H_k \|_2 \) (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:

$$H_{ji} = \sum_{k=1}^{d} \sigma_{jik} u_{jik} v_{jik}^H + t \sum_{k=d+1}^{d} \sigma_{jik} u_{jik} v_{jik}^H$$

The IA (ZF) condition for rank 1 link $i - j$ can be written as

$$\sigma_{ji} f_j^H u_{ji} v_{ji}^H g_i = 0$$

- Two configurations are possible: $f_j^H u_{ji} = 0$ or $v_{ji}^H g_i = 0$

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.
- Homotopy for finding global optimum: at low SNR, noise dominates interference $\Rightarrow$ optimal: one stream per power constraint, matched filter $T_x/R_x$. 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: linearize convex part in terms of Tx covariance matrices $Q_k$ to make it concave.

Afterwards work with BF in $Q_k = g_k g_k^H$.

But the linearization in $Q_k$ does not correspond to second-order Taylor series or any precise development in $g_k$.

Other interpretation: majorization: replace cost function to be maximized by one below it that touches the original one in one point [Stoica:SPmagJan04].

Specifically: matrix version of $x - 1 - \ln(x) \geq 0$, $x > 0$: Itakura-Saito distance in AR modeling ($x =$ ratio of true spectrum and AR model spectrum).

Majorized cost function can be optimized with any parameterization.
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
Overall Outline

- infinite SNR: MIMO Interference Alignment
- perfect CSIT coordinated beamforming designs
- partial CSIT deterministic approximations
Partial CSIT Outline

- Partial CSIT channel models
  - Gaussian CSIT: integrating channel estimates and covariance
- Expected/Ergodic Weighted Sum Rate (EWSR)
- lower bound: Expected Weighted Sum MSE (EWSMSE)
- Massive MIMO limit: ESEI-WSR $M, K \to \infty$
- upper bound: Massive $N$ MIMO $N \to \infty$
- large MIMO limit $M, N \to \infty$
Mean information about the channel can come from channel feedback or reciprocity, and prediction, or it may correspond to the non fading (e.g. LoS) part of the channel (note that an unknown phase factor $e^{j\phi}$ in the overall channel mean does not affect the BF design).

Covariance information may correspond to channel estimation (feedback, prediction) errors and/or to information about spatial correlations. The separable (or Kronecker) correlation model (for the channel itself, as opposed to its estimation error or knowledge) below is acceptable when the number of propagation paths $N_p$ becomes large ($N_p \gg MN$) as possibly in indoor propagation.

Given only mean and covariance information, the fitting maximum entropy distribution is Gaussian.
Massive MIMO: from spatial to spatiotemporal and back

- spatial: to null a user, need to null all paths of that user
- spatiotemporal: \# antennas > \# users
- spatial: \# antennas > \# paths \gg \# users
- but: paths are **slowly** fading, user channels are **fast** fading

- Keysight ieee comsoc M-MIMO tutorial, mmWave
Figure 1: MIMO transmission with $M$ transmit and $N$ receive antennas.

The antenna array responses are just functions of angles AoD, AoA in the case of standard antenna arrays with scatterers in the far field. In the case of distributed antenna systems, the array responses become a function of all position parameters of the path scatterers. The fast variation of the phases $\psi_j$ (due to Doppler) and possibly the variation of the $A_j$ correspond to the fast fading. All the other parameters vary on a slower time scale and correspond to slow fading.

MIMO channel transfer matrix at any particular subcarrier of a given OFDM symbol

$$H = \sum_{i=1}^{N_p} A_i e^{j\psi_i} h_r(\phi_i) h_t^T(\theta_i) = BA^H$$  \hspace{1cm} (1)$$

where there are $N_p$ (specular) pathwise contributions with

- $A_i > 0$: path amplitude
- $\theta_i$: direction of departure (AoD)
- $\phi_i$: direction of arrival (AoA)
- $h_t(\cdot)$, $h_r(\cdot)$: $M/N \times 1$ Tx/Rx antenna array response

and

$$B = \begin{bmatrix} e^{j\psi_1} \\ e^{j\psi_2} \\ \vdots \end{bmatrix}$$

$$A^H = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ h_t^T(\theta_1) \\ h_t^T(\theta_2) \end{bmatrix}$$  \hspace{1cm} (2)$$
Given only mean and (separable) covariance information, the fitting maximum entropy distribution is Gaussian. Hence consider $\text{vec}(H) \sim \mathcal{CN}(\text{vec}(\bar{H}), C_r^T \otimes C_r)$ which can be rewritten as

$$H = \bar{H} + C_r^{1/2} \tilde{H} C_t^{1/2}$$

(3)

where $C_r^{1/2}, C_t^{1/2}$ are Hermitian square-roots of the Rx and Tx side covariance matrices

$$E(H - \bar{H})(H - \bar{H})^H = \text{tr}\{C_t\} C_r$$
$$E(H - \bar{H})^H(H - \bar{H}) = \text{tr}\{C_r\} C_t$$

(4)

and the elements of $\tilde{H}$ are i.i.d. $\sim \mathcal{CN}(0, 1)$. In what follows, it will also be of interest to consider the total Tx side correlation matrix

$$R_t = EH^H = \bar{H}^H \bar{H} + \text{tr}\{C_r\} C_t.$$  

(5)

Note that the Gaussian CSIT model could be considered an instance of Ricean fading in which the ratio $\text{tr}\{(H^H \bar{H})/(\text{tr}\{C_r\} \text{tr}\{C_t\})$ could be considered the Ricean factor.

This Taylor series modeling of clusters is in contrast to the uniform DoA profile used in [Caire:mmWave], [Gesbert:arxiv1013].

Assuming the Tx disposes of not much more than the information about $r$ dominant path AoDs, we shall consider the following MIMO (Ricean) channel model

$$H = B A^H(\theta) + \sqrt{\beta} \tilde{H}'$$

(6)

which follows from (1), (2) except restricted to the $r$ strongest paths, with the rest modeled by $\sqrt{\beta} \tilde{H}'$ (elements i.i.d. $\sim \mathcal{CN}(0, \beta)$, independent of the $\psi_i$). Averaging over path phases $\psi_i \Rightarrow$ Tx side covariance matrix

$$C_t = AA^H + N/\beta I_M$$

(7)

since due to the normalization of the antenna array responses, $E B^H B = I$. Note that $\mu = \text{tr}\{AA^H\}/\beta NM$ could be considered a Ricean factor. When needed, we may also consider the $h_r$, the columns of $B$, to be isotropically distributed. Note that the rank of $AA^H$ can be substantially less than the number of paths. Consider e.g. a cluster of paths with narrow AoD spread, then we have $\theta_i = \theta + \Delta \theta_i$ where $\theta$ is the nominal AoD and $\Delta \theta_i$ is small $\Rightarrow$ $h_t(\theta_i) \approx h_t(\theta) + \Delta \theta_i h_t(\theta)$: rank 2 contribution to $AA^H$. 
Averaging over the (uniform) path phases $\psi_i$ leads to

$$C_{hh} = \sum_{i=1}^{N_p} A_i^2 \mathbf{h}_i \mathbf{h}_i^H = \sum_{i=1}^{N_p} A_i^2 (\mathbf{h}_r(\phi_i)\mathbf{h}_r^H(\phi_i)) \otimes (\mathbf{h}_t(\theta_i)\mathbf{h}_t^H(\theta_i))$$

where $C_{hh} = \mathbb{E} \mathbf{h}\mathbf{h}^H$, $\mathbf{h} = \text{vec}(\mathbf{H})$ and $\mathbf{h}_i = \mathbf{h}_t(\theta_i) \otimes \mathbf{h}_r(\phi_i)$. Note that the rank of $C_{hh}$ can be substantially less than the number of paths. Consider e.g. a cluster of paths with narrow AoD spread, then we have

$$\theta_i = \theta + \Delta \theta_i$$

where $\theta$ is the nominal AoD and $\Delta \theta_i$ is small. Hence

$$\mathbf{h}_t(\theta_i) \approx \mathbf{h}_t(\theta) + \Delta \theta_i \dot{\mathbf{h}}_t(\theta) .$$

Such a cluster of paths only adds a rank 2 contribution to $C_{hh}$. Not of Kronecker form.
Tx side Covariance CSIT

Tx side covariance matrix $C^t$, which only explores the channel correlations as they can be seen from the BS side

$$C^t = E H^H H$$

We can factor the channel response as

$$H = B A^H, \quad B = [h_r(\phi_1) \ h_r(\phi_1) \cdots]$$

$$A^H = \begin{bmatrix} A_1 \\ A_2 \\ \vdots \end{bmatrix} \begin{bmatrix} h_t^T(\theta_1) \\ h_t^T(\theta_2) \\ \vdots \end{bmatrix}$$

Averaging of the path phases $\psi_i$, we get for the Tx side covariance matrix

$$C^t = AA^H$$

since due to the normalization of the antenna array responses,

$$E B^H B = \text{diag}\{[h_r(\phi_1) \ h_r(\phi_2) \cdots]^H [h_r(\phi_1) \ h_r(\phi_1) \cdots]\} = I.$$
The ZF from BS $j$ to MT $(i, k)$ requires

$$F_{i,k}^H H_{i,k,j} G_{j,n} = F_{i,k}^H B_{i,k,j} A_{i,k,j}^H G_{j,n} = 0$$

which involves $\min(d_{i,k} d_{j,n}, d_{i,k} r_{i,k,j}, r_{i,k} d_{j,n})$ constraints to be satisfied by the $(N_i,k - d_{i,k}) d_{i,k}/(M_j - d_{j,n}) d_{j,n}$ variables parameterizing the column subspaces of $F_{i,k}/G_{j,n}$.

**IA feasibility singular MIMO IC with Tx/Rx decoupling**

$$F_{i,k}^H B_{i,k,j} = 0 \text{ or } A_{i,k,j}^H G_{j,n} = 0.$$  

This leads to a possibly increased number of ZF constraints $r_{i,k,j} \min(d_{i,k}, d_{j,n})$ and hence to possibly reduced IA feasibility. ZF of every cross link now needs to be partitioned between all Txs and Rxs, taking into account the limited number of variables each Tx or Rx has. The main goal of this approach however is that it leads to Tx/Rx decoupling and local CSI.
In massive MIMO, the Tx side channel covariance matrix is very likely to be (very) singular even though the channel response $H$ may not be singular:

$$\text{rank}(C_{i,k,j}^t = A_{i,k,j}A_{i,k,j}^H) = r_{i,k,j}, \ A_{i,k,j} : M_j \times r_{i,k,j}$$

Let $P_X = X(X^H X)^\# X^H$ and $P_X^\perp$ be the projection matrices on the column space of $X$ and its orthogonal complement resp. Consider now a massive MIMO IBC with $C$ cells containing $K_i$ users each to be served by a single stream. The following result states when this will be possible.

**Theorem**

**Sufficiency of Covariance CSIT for Massive MIMO IBC**  
*In the MIMO IBC with (local) covariance CSIT, all BS will be able to perform ZF BF if the following holds*

$$\|P_{A_{i,k,j}}^\perp A_{i,k,j}\| > 0, \ \forall i, k, j$$

*where $A_{i,k,j} = \{A_{n,m,j}, (n, m) \neq (i, k)\}$.***
These conditions will be satisfied w.p. 1 if
\[ \sum_{i=1}^{C} \sum_{k=1}^{K_i} r_{i,k,j} \leq M_j, \quad j = 1, \ldots, C. \]
In that case all the column spaces of the \( A_{i,k,j} \) will tend to be non-overlapping. However, the conditions could very well be satisfied even if these column spaces are overlapping, in contrast to what [Gesbert:arxiv1013],[Caire:arxiv0912] appear to require.

In Theorem 1, we assume that all ZF work is done by the BS. However, if the MT have multiple antennas, they can help to a certain extent.

**Theorem**

**Role of Receive Antennas in Massive MIMO IBC**  If MT \((i, k)\) disposes of \(N_{i,k}\) antennas to receive a stream, it can perform rank reduction of a total amount of \(N_{i,k} - 1\) to be distributed over \(\{r_{i,k,j}, \quad j = 1, \ldots, C\}\).

Such rank reduction (by ZF of certain path contributions) facilitates the satisfaction of the conditions in Theorem 1.
Pathwise Multi-User Multi-Cell

scrambler → path gains, DoAs; DL → dual UL LMMSE
prior MIMO (Ricean) channel model

\[ H = B A^H(\theta) + \sqrt{\beta} \tilde{H} \]

Averaging of the path phases \( \psi_i \), we get for the Tx side covariance matrix

\[ \mathbb{E} H^H H = N C_t = N (AA^H + \beta I_M) \]

The Tx dispose also of a (deterministic) channel estimate

\[ \hat{H} = H + \frac{1}{\sqrt{N}} \tilde{H}_d C_d^{1/2} \]

where \( \tilde{H}_d \) elements i.i.d. \( \sim \mathcal{CN}(0, 1) \), typically \( C_d = \sigma_h^2 \). Combo channel estimate + prior \( \rightarrow \) (posterior) LMMSE estimate.
To simplify, force a separable prior model by considering all of \( B \) to be unknown with elements that are i.i.d. \( \sim \mathcal{CN}(0, 1) \). As a result we can interpret \( \mathbf{H} \) to be of the form \( \mathbf{H} = \tilde{\mathbf{H}} \mathbf{C}_t^{1/2} \) and we get for the LMMSE estimate

\[
\hat{\mathbf{H}} = \hat{\mathbf{H}}_d (C_t + C_d)^{-1} C_t = \mathbf{H} + \tilde{\mathbf{H}}_p \mathbf{C}_p^{1/2}
\]

\[
C_p = C_d (C_t + C_d)^{-1} C_t
\]

where \( \hat{\mathbf{H}} \) and \( \tilde{\mathbf{H}}_p \) are independent.

Note that we get for the MMSE estimate of a quadratic quantity of the form

\[
E_{\mathbf{H}|\hat{\mathbf{H}}_d} \mathbf{H}^H \mathbf{H} = \hat{\mathbf{H}}^H \hat{\mathbf{H}} + N \mathbf{C}_p = \mathbf{W}.
\]

This MMSE estimate implies

\[
\mathbf{W} = \arg \min_{\mathbf{T}} E_{\mathbf{H}|\hat{\mathbf{H}}_d} \| \mathbf{H}^H \mathbf{H} - \mathbf{T} \|^2.
\]

It averages out to

\[
E_{\hat{\mathbf{H}}_d} \mathbf{W} = E_{\mathbf{H}, \hat{\mathbf{H}}_d} \mathbf{H}^H \mathbf{H} = E_{\mathbf{H}} \mathbf{H}^H \mathbf{H} = N \mathbf{C}_t.
\]
Max Expected WSR (EWSR)

- scenario of interest: perfect CSIR, partial CSIT
- Imperfect CSIT $\Rightarrow$ various possible optimization criteria: outage capacity,.... Here: expected weighted sum rate

$$E_H WSR(g, H) =$$

$$EWSR(g) = E_H \sum_k u_k \ln(1 + g_k^H H_k^H R_k^{-1} H_k g_k)$$

perfect CSIR: optimal Rx filters $f_k$ (fn of aggregate $H$) have been substituted: $WSR(g, H) = \max_f \sum_k u_k (-\ln(e_k(f_k, g)))$.

- 2 step averaging:

$$\max EWSR(g) = E_{\hat{H}} \max_g E_{H|\hat{H}} \sum_k u_k \ln(1 + g_k^H H_k^H R_k^{-1} H_k g_k)$$

- Note that

$$E_{H|\hat{H}} H Q H^H = \hat{H} Q \hat{H}^H + \text{tr}\{Q C_p\} I_N$$

$$E_{H|\hat{H}} H^H Q H = \hat{H}^H Q \hat{H} + \text{tr}\{Q\} C_p$$
EWSR Lower Bound: EWSMSE

- *EWSR*(g) : difficult to compute and to maximize directly.
  [Negro:iswcs12] [BjornsonLarsson:gc15] much more attractive to consider \( E_H e_k(f_k, g, H) \) since \( e_k(f_k, g, H) \) is quadratic in \( H \). Hence optimizing \( E_H WSMSE(g, f, w, H) \).

\[
\min_{f, w} E_H WSMSE(g, f, w, H) \\
\geq E_H \min_{f, w} WSMSE(g, f, w, H) = -EWSR(g)
\]

or hence \( EWSR(g) \geq -\min_{f, w} E_H WSMSE(g, f, w, H) \).

- So now only a lower bound to the EWSR gets maximized, which corresponds in fact to the CSIR being equally partial as the CSIT.

\[
E_H e_k = 1 - 2\Re\{f_k^H \hat{H}_k g_k\} + \sum_{i=1}^{K} f_k^H \hat{H}_k g_i g_i^H \hat{H}_k f_k
\]
\[
+ f_k^H R_{r, k} f_k \sum_{i=1}^{K} g_i^H R_{t, k} g_i + ||f_k||^2.
\]

\( \Rightarrow \) signal term disappears if \( \hat{H}_k = 0 \)! Hence the EWSMSE lower bound is (very) loose unless the Rice factor is high, and is useless in the absence of mean CSIT.
Alternating optimization as before leads to

\[
\begin{align*}
\min_{w_k} EWSMSE & \Rightarrow w_k = 1 / \hat{e}_k \\
\min_{f_k} EWSMSE & \Rightarrow f_k = \hat{R}_k^{-1} \hat{H}_{k,b_k} g_k \\
\min_{g_k} EWSMSE & \Rightarrow g_k = \left( \hat{T}_k + \lambda b_k I_M \right)^{-1} \hat{H}_{k,b_k} f_k u_k w_k
\end{align*}
\]

where

\[
\begin{align*}
\hat{R}_k &= \sum_i \hat{H}_{k,b_i} g_i g_i^H \hat{H}_{k,b_i} + (1 + \sum_i g_i^H C_{p,k,b_i} g_i) I_{N_k} \\
\hat{T}_k &= \sum_{i=1}^K u_i w_i (\hat{H}_{i,b_k} f_i f_i^H \hat{H}_{i,b_k} + \|f_i\|^2 C_{p,k,b_i}) .
\end{align*}
\]
Expected Weighted Sum Unbiased MSE (EWSUMSE)

- Take $E_{H|\hat{H}}$ of $WSUMSE(g, f, w_1, w_2)$

$$= \sum_{k=1}^{K} u_k(w_{1,k}e_k^u(f_k, g) - \ln w_{1,k} - w_{2,k}(e_k^u(f_k, g) + 1) + \ln w_{2,k})$$

$$+ \sum_{k=1}^{K} \mu_k(1 - |f_k^H H_k, b_k g_k|^2) + \sum_{i=1}^{C} \lambda_i(\sum_{k: b_k = i} \|g_k\|^2 - P_i)$$

where $e_k^u(f_k, g) = \sum_{i \neq k} f_k^H H_k, b_i g_i g_i^H H_k^H, b_i f_k + \|f_k\|^2$.

- Alternating optimization leads to

$$w_k = 1/(\hat{e}_k^u(\hat{e}_k^u + 1))$$

$$f_k = V_{\max}(\hat{H}_{k, b_k} g_k g_k^H \hat{H}_{k, b_k} + g_k^H C_{p, k, b_k} g_k I_{N_k}, \hat{R}_{k})$$

$$g_k = V_{\max}(\hat{H}_{k, b_k} f_k f_k^H \hat{H}_{k, b_k} + \|f_k\|^2 C_{p, k, b_k}, \hat{T}_{k} + \lambda_{b_k} I_{M})$$

where $\hat{R}_{k}$ and $\hat{T}_{k}$ are like $\hat{R}_{k}$ and $\hat{T}_{k}$ in EWSUMSE but without the term for user $k$. 
Note that in case of no channel estimate, $\hat{H}_{k,b_k} = 0$, then

$$f_k = V_{\min}(\hat{R}_k)$$

$$g_k = V_{\max}(C_{p,k,b_k}, \hat{T}_k + \lambda_{b_k} I_M).$$

The Rx and Tx filters optimizing the EWSUMSE criterion also optimize the ESEINR (Expected Signal to Expected Interference plus Noise Ratio) and (optimally weighted) ESELNR (Expected Signal to Expected Leakage plus Noise Ratio):

$$ESEINR_k = \frac{f_k^H (\hat{H}_{k,b_k} g_k g_k^H \hat{H}_{k,b_k} + g_k^H C_{p,k,b_k} g_k I_{N_k}) f_k}{f_k^H \hat{R}_k f_k}$$

$$ESELNR_k = \frac{g_k^H (\hat{H}_{k,b_k} f_k f_k^H \hat{H}_{k,b_k} + \|f_k\|^2 C_{p,k,b_k}) g_k}{g_k^H (\hat{T}_k + \lambda_{b_k} I_M) g_k}.$$
DC programming waterfilling: \( \|f_k\| = 1, \|g_k\| = 1 \). The stream powers and Lagrange multipliers get adapted with \( \hat{T}_k \) replacing \( T_k \) and \( S_k \) being replaced by

\[
\hat{S}_k = \hat{H}_{k,b_k} \hat{R}_k^{-1} \hat{H}_{k,b_k} + \text{tr}\{\hat{R}_k^{-1}\} C_{p,k,b_k}.
\]
Using the concavity of $\ln(.)$, we get

$$EWSR(g) \leq \sum_{k=1}^{K} u_k \ln(1 + E_{H_k} \text{SINR}_k(g, H_k)) .$$
Consider the approximation

$$E_H \ln(1 + \text{SINR}_k) \approx \ln\left(1 + \frac{E_S}{E_I + N}\right).$$

This can be solved as easily as $\min (E)W_{\text{SMSE}}$!

However, here the $\tilde{H}_k$ part in the signal gets also counted in the signal power, unlike in the EWSMSE criterion where it gets ignored.

Approximation exact in (reversed) Massive MIMO, $N \to \infty$.

Rewrite WSR (level of Rx signal iso Rx output)

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

w/ $\tilde{R}_k = (\sum_i Q_i) H_k^H H_k + I_M$, $\tilde{R}_k^{-1} = (\sum_{i \neq k} Q_i) H_k^H H_k + I_M$.

Can apply [KimGiannakis:IT0511], replacing $\tilde{R}_k$, $\tilde{R}_k^{-1}$ by $E \tilde{R}_k$, $E \tilde{R}_k^{-1}$, and hence $H_k^H H_k$ by $R_{t,k}$ and expressions of the form $H_k^H R^{-1} H_k$ by $R_{t,k} R^{-1}$.
We get a convergence for any term of the form

\[ HQH^H \xrightarrow{M \to \infty} E_H HQH^H = \bar{H}Q\bar{H}^H + \text{tr}\{QC_t\} C_r. \]

Go one step further in separable channel correlation model: 
\[ C_{r,k,b_i} = C_{r,k}, \ \forall b_i. \] This leads us to introduce

\[ H_k = [H_{k,1} \cdots H_{k,C}] = \bar{H}_k + C_{r,k}^{1/2} \sim H_k C_{t,k}^{1/2} \]

\[ Q = \begin{bmatrix} \sum_{i:b_i=1} Q_i \\ \vdots \\ \sum_{i:b_i=C} Q_i \end{bmatrix} = \sum_{j=1}^C \sum_{i:b_i=j} I_j Q_i I_j^H \]

\[ Q_{\sim}^k = Q - I_{b_i} Q_i I_{b_i}^H \]

where \( C_{t,k} = \text{blockdiag}\{C_{t,k,1}, \ldots, C_{t,k,C}\} \), and \( I_j \) is an all zero block vector except for an identity matrix in block \( j \). Then we get for the WSR \((= EWSR)\),

\[ \text{WSR} = \sum_{k=1}^K u_k \ln \det(\bar{R}_k^{-1} \tilde{R}_k) \]

where

\[ \tilde{R}_k = I_{N_k} + \bar{H}_k Q \bar{H}_k^H + \text{tr}\{QC_{t,k}\} C_{r,k} \]

\[ \tilde{R}_k = I_{N_k} + \bar{H}_k Q_k \bar{H}_k^H + \text{tr}\{Q_{\sim}^k C_{t,k}\} C_{r,k} \]
This leads to
\[
WSR = u_k \ln \det(I + \tilde{R}^{-1}_k \left( \overline{H}_{k,b_k} g_k g_k^H \overline{H}_{k,b_k}^H + \text{tr}\{g_k g_k^H C_{t,k,b_k} \} C_{r,k} \right)) + WSR_k^-
\]

Consider simplified case: "Ricean factor" $\mu \sim \text{SNR}$, for the direct links $H_{k,b_k}$ (only) (properly organized (intracell) channel estimation and feedback) $\Rightarrow$ approximation

\[
WSR = u_k \ln \det(I + g_k^H \tilde{B}_k g_k) + WSR_k^- \quad \text{with}
\]
\[
\tilde{B}_k = \overline{H}_{k,b_k}^H \tilde{R}^{-1}_k \overline{H}_{k,b_k} + \text{tr}\{C_{r,k} \tilde{R}^{-1}_k \} C_{t,k,b_k}
\]

The linearization of $WSR_k^-$ w.r.t. $Q_k$ now involves

\[
\tilde{A}_k = \sum_{i \neq k}^K u_i \left[ \overline{H}_{i,b_k}^H \left( \tilde{R}^{-1}_i - \tilde{R}^{-1}_k \right) \overline{H}_{i,b_k} + \text{tr}\left( \left( \tilde{R}^{-1}_i - \tilde{R}^{-1}_k \right) C_{r,i} \right) C_{t,i,b_k} \right].
\]

The rest of the development is now completely analogous to the case of perfect CSIT.
Figure: Expected sum rate comparison for $C = 2$, $K = 8$, $M = 8$, $N = 1$, $\text{rank}(C_{t,k,b_i}) = 2$, $\forall i$, $\forall k$ and $\tau^2 = \frac{1}{10}$. 
Large MIMO Asymptotics Refinement

- SU MIMO asymptotics from [Loubaton:IT0310], [Taricco:IT0808] (in which both $M, N \to \infty$, which tends to give more precise approximations when $M$ is not so large) for a term of the form $\ln \det(\mathbf{Q}\mathbf{H}^H\mathbf{H} + I)$ correspond to replacing $\mathbf{H}_k^H\mathbf{H}_k$ in the $\tilde{\mathbf{R}}_k$ and $\tilde{\mathbf{R}}_k^\perp$ with a kind of $\mathbf{R}_{t,k}$ with a different weighting of the $\mathbf{H}_k^H\mathbf{H}_k$ and $\mathbf{C}_{t,k}$ portions, of the form
  
  $$\tilde{\mathbf{R}}_{t,k}' = a_k \mathbf{C}_{t,k} + \mathbf{H}_k^H\mathbf{B}_k\mathbf{H}_k$$

  for some scalar $a_k$ and matrix $\mathbf{B}_k$ that depends on $\mathbf{C}_{r,k}$.

- For the general case of Gaussian CSIT with separable (Kronecker) covariance, get

  $$E_{\mathbf{H}} \ln \det(I + \mathbf{H} \mathbf{Q}\mathbf{H}^H)$$

  $$= \max_{z,w} \left\{ \ln \det \begin{bmatrix} I + w \mathbf{C}_{r} & \mathbf{H} \\ -\mathbf{Q}\mathbf{H}^H & I + z \mathbf{Q}\mathbf{C}_{t} \end{bmatrix} - zw \right\}.$$ 

  $\max_{z,w}$ interpretation is new.
The EWSR with large MIMO asymptotics now becomes

\[ EWSR = \sum_{k=1}^{K} u_k \left( \max_{z_k, w_k} \{ \ln \det S_k(Q, z_k, w_k) - z_k w_k \} \right) \]

\[ - \max_{z_k, w_k} \left\{ \ln \det S_k \left( Q_{-k}, z_k, w_k \right) - z_k w_k \right\} \]

where

\[ S_k(Q, z, w) = \begin{bmatrix} I + w C_{r,k} & \overline{H}_k \\ -Q \overline{H}_k^H & I + z QC_{t,k} \end{bmatrix}. \]

Note that

\[ \ln \det S_k(Q, z, w) = \ln \det(I + w C_{r,k}) + \ln \det(I + Q T_k(z, w)) \]

with

\[ T_k(z, w) = z C_{t,k} + \overline{H}_k^H (I + w C_{r,k})^{-1} \overline{H}_k \]

where \( T_k \) plays the role of some kind of total Tx side channel correlation matrix. Note that the weighting coefficients \( z, w \) depend on the BFs also though.
The EWSR expression can be maximized alternatingly over the \( \{g_k\} \), the \( \{z_k, w_k\} \) and the \( \{z_{\overline{k}}, w_{\overline{k}}\} \). For the optimization of the BFs \( g_k \), for given \( z, w \), introduce

\[
\hat{R}_{k,\overline{k}} = I + \hat{Q}_k \hat{T}_k(z_k, w_k)
\]
\[
\hat{R}_{\overline{k}} = I + \hat{Q}_{\overline{k}} \hat{T}_{\overline{k}}(z_{\overline{k}}, w_{\overline{k}}), \quad \hat{R}_k = I + \hat{Q}_k \hat{T}_k(z_k, w_k).
\]

Generalized eigenvector approach and interference aware WF of the perfect CSI case can be applied with

\[
\hat{B}_k = \hat{T}_k(z_k, w_k) \hat{R}_{k,\overline{k}}^{-1}
\]
\[
\hat{A}_k = \sum_{i \neq k} u_i \left[ \hat{T}_i(z_i, w_i) \hat{R}_i^{-1} - \hat{T}_i(z_{\overline{i}}, w_{\overline{i}}) \hat{R}_{\overline{i}}^{-1} \right].
\]

For the optimization \( \max_{z \geq 0, w \geq 0} \{ \ln \det S(Q, z, w) - zw \} \): KKT

\[
w = f(Q, z, w) = \text{tr}\{ QC_t [I + QT(z, w)]^{-1} \}
\]
\[
z = g(Q, z, w) = \text{tr}\{ C_r [I + wC_r\overline{H}(I + zQC_t)^{-1}Q\overline{H}^H]^{-1} \}.
\]
WSMSE-DC BF relation

- min WSMSE iteration \((i + 1)\)

\[
A_k^{(i)} = \sum_j u_j w_j^{(i)} H_i^H f_j^{(i)} f_j^{(i)H} H_j + \lambda^{(i)} I_M
\]

\[
g_k^{(i+1)} = (A_k^{(i)})^{-1} H_k^H f_k^{(i)} u_k w_k^{(i)}
\]

\[
= (A_k^{(i)})^{-1} B_k^{(i)} g_k^{(i)} u_k w_k^{(i)}
\]

\[
B_k^{(i)} = H_k^H R_k^{-(i)} H_k
\]

WSMSE does one power iteration of DC !!

\[
g_k^{(i+1)} = \text{V}_{max}\{(A_k^{(i)})^{-1} B_k^{(i)}\}
\]

- partial CSIT (or MaMIMO) case: modified WSMSE:

\[
g_k^{(i+1)} = (E H A_k^{(i)})^{-1} (E H B_k^{(i)}) g_k^{(i)} u_k w_k^{(i)}
\]
Various Asymptotic MIMO Regimes

- Large MIMO regime considered above: \(M, N \rightarrow \infty, M/N \rightarrow \text{constant}, K \text{ finite}\)
- Massive MIMO regime: \(M, K \rightarrow \infty, M/K \rightarrow \text{constant}, N \text{ finite}\)
- Considered in [1], for (R-)ZF BF, Gaussian channel vectors with arbitrary covariance matrices and CSIT errors
- Optimal BF (perfect CSIT) are considered in [2]


In the case of partial CSIT we get for the symbol estimate

\[ \hat{x}_k = f_k^H \hat{H}_{k,b_k} g_k x_k + f_k^H \tilde{H}_{k,b_k} g_k x_k \]

\[ \text{sig. ch. error} \]

\[ + \sum_{i=1, \neq k}^K (f_k^H \hat{H}_{k,b_i} g_i x_i + f_k^H \tilde{H}_{k,b_i} g_i x_i) + f_k^H v_k \]

\[ \text{interf. ch. error} \]

Naive EWSR (NEWSR) : just replace \( H \) by \( \hat{H} \) in a perfect CSIT approach. Ignore \( \tilde{H} \) everywhere.

EWSMSE: accounts for covariance CSIT in the interference. This can have significant impact, even on the DoF if the instantaneous channel CSIT quality does not scale with SNR. However, EWSMSE also moves the signal \( \tilde{H} \) term to the interference plus noise.
Discussion EWSR Approximations (2)

- **EWSUMSE**: signal $\tilde{H}$ term is accounted for in the signal power.
- In a MaMIMO setting, the way mean and covariance CSIT are combined in the WSUMSE approach for the interference terms becomes equally optimal as in the EWSR for a large number of users (CLT).
- EWSUMSE furthermore represents an improvement over EWSMSE for capturing the signal power (matched filtering and diversity aspects) but only leads to a finite (dB) gain in ESEINR, though its remaining approximation error over EWSR may be limited.
- In the MaMIMO setting, EWSUMSE represents a EWSR upper bound due to the concavity of $\ln(.)$. 
Relaxed generalized eigen decompositions

\[ g_k = V_{\text{max}}(\mathbf{h}_{k,b_k} \mathbf{h}_{k,b_k}^H + C_{p,k,b_k}, \tilde{\mathbf{A}}_k + \lambda_{b_k} \mathbf{I}) \] with
\[ \frac{1}{a_k} = \sigma_{\text{max}}(\mathbf{h}_{k,b_k} \mathbf{h}_{k,b_k}^H + C_{p,k,b_k}, \tilde{\mathbf{A}}_k + \lambda_{b_k} \mathbf{I}) \] then can compute \( g_k \) iteratively from
\[ a_k \mathbf{h}_{k,b_k} \mathbf{h}_{k,b_k}^H \mathbf{g}_k = (\tilde{\mathbf{A}}_k + \lambda_{b_k} \mathbf{I} - a_k C_{p,k,b_k}) \mathbf{g}_k \]

which leads to the following explicit solution
\[ \mathbf{g}'_k = (\tilde{\mathbf{A}}_k + \lambda_{b_k} \mathbf{I} - a_k C_{p,k,b_k})^{-1} \mathbf{h}_{k,b_k} \]

more generally, if \( \mathbf{g}_{\text{max}} = V_{\text{max}}(\mathbf{S}, \mathbf{T}) \) where \( \text{rank}(\mathbf{S}) = r \) then a \( V_{\text{max}} \) of a \( r \)-dimensional matrix suffices.
Concluding Remarks

- In Massive MIMO, the number of random sources may not scale with the number of antennas, hence no central limit theorem.
- In large system MIMO analysis, the true correlation structure is not separable (non-Kronecker).
- Covariance CSIT acquisition? Work with parameterized pathwise models to significantly reduce # unknowns and do joint communications & geopositioning?
- Beyond classical beamforming: network coding + multicast + space-time coding on random (unknown) channel part? [Kobayashi:tiw16]