A Unified Framework for Training-Based Channel Matrix and Norm Estimation in MIMO Systems

Emil Björnson and Björn Ottersten
ACCESS Linnaeus Center, Signal Processing
Royal Institute of Technology (KTH)
Stockholm, Sweden
Overview

- Narrowband Communication
  - Downlink communication from base station
  - One or several receiving users
  - Multiple antennas at both sides
  - Block fading environment
Overview (2)

- **Channel State Information (CSI)**
  - Channel between transmit and receive antennas
  - Complex coefficient (describes gain and phase shift)

- **Instantaneous CSI**
  - Current values of coefficients
  - Needs to be estimated and used with short delay

- **Statistical CSI**
  - How are the coefficient correlated?
  - Can be estimated slowly with a long time window
  - Assumed to be known perfectly at both sides
Overview (3)

- System Operation (Block Fading):

  - Perfect Channel Estimation at Receiver
    - Often assumed when focus is on transmission design

  - Instantaneous Channel Information useful:
    - Receive processing (Interference suppression, detection)
    - Feedback (User selection, precoding, rate adaptation)
Outline

• System Model
• Channel Matrix Estimation
  ▪ MMSE Estimator and Training Design
• Length of Training Sequence
• Channel Norm Estimation
  ▪ MMSE Estimator and Training Design
• Numerical Examples
• Summary and References
System Model

• MIMO Communication:
  ▪ $n_T$ transmit antennas, $n_R$ receive antennas

• Communication model to user $k$:
  \[
  \begin{align*}
  \mathbf{y}_k(t) &= \mathbf{H}_k \cdot \mathbf{x}(t) + \mathbf{n}_k(t) \\
  &\quad \left(n_R \times 1\right) \quad \left(n_R \times n_T\right) \quad \left(n_T \times 1\right) \quad \left(n_R \times 1\right)
  \end{align*}
  \]
  ▪ $\mathbf{x}(t)$ transmitted signal, $\mathbf{y}_k(t)$ received signal
  ▪ $\mathbf{n}_k(t)$ potentially correlated complex Gaussian noise
  ▪ Rician distributed channel matrix:
  \[
  \text{vec}(\mathbf{H}_k) \in \mathcal{CN}(\text{vec}(\tilde{\mathbf{H}}_k), \mathbf{R}_k)
  \]
System Model (2)

• Problem description:
  ▪ Estimate properties of the channel matrix $H_k$
  ▪ In general, we are interested in some function $f(H_k)$ (receiver structure, modulation, precoding)

• In this work we estimate two parameters
  ▪ $H_k$ channel matrix (many applications)
  ▪ $\|H_k\|^2_F$ channel gain (for user-selection, rate adaptation)

  ▪ It will be illustrated that calculation of $\|H_k\|^2_F$ from an estimation of $H_k$ gives poor performance
System Model (3)

• Training-based channel estimation
  - Training sequence of length $n_T$
    (maximal length if no per-symbol power constraint)
  - Represented by matrix $\mathbf{P}_k \in \mathbb{C}^{n_T \times n_T}$
  - Training power constraint: $\text{tr}(\mathbf{P}^H_k \mathbf{P}_k) = \mathcal{P}$
  - Transmission of $\mathbf{P}_k$ over $n_T$ symbol slots:

$$\mathbf{Y}_k = \mathbf{H}_k \mathbf{P}_k + \mathbf{N}_k$$

$$\mathbf{Y}_k = [y_k(1), \ldots, y_k(n_T)], \quad \mathbf{N}_k = [n_k(1), \ldots, n_k(n_T)]$$

- General disturbance statistics:
  $$\text{vec}(\mathbf{N}_k) \in \mathcal{CN}(\text{vec}(\tilde{\mathbf{N}}_k), \mathbf{S}_k)$$
Kronecker Product

- **Definition:**
  \[
  A \otimes B = \begin{bmatrix}
  a_{11}B & \ldots & a_{1n}B \\
  \vdots & \ddots & \vdots \\
  a_{m1}B & \ldots & a_{mn}B
  \end{bmatrix}
  \]

- **Useful to analyze matrix multiplication:**
  \[
  \text{vec}(CH_kD) = (D^T \otimes C) \text{vec}(H_k)
  \]
  - Training matrix \(P_k\) multiplied from the right:
    \[
    \text{vec}(H_kP_k) = (P_k^T \otimes I) \text{vec}(H_k)
    \]
  - To analyze impact of \(P_k\) we will later assume Kronecker-structured channel properties and use that
    \[
    (A \otimes B)(E \otimes F) = (AE \otimes BF)
    \]
Channel Matrix Estimation

• MMSE Estimation of $\mathbf{H}_k$ in a unified way

We consider Rician fading and Rician disturbance

Linear MMSE estimators have previously been derived:


Has also been done in the wrong way (suboptimally):


Channel Matrix Estimation (2)

- MMSE Estimator:

\[ \text{vec}(\hat{H}_{\text{MMSE}}) = \]
\[ \text{vec}(\bar{H}_k) + R_k \tilde{P}_k^H \left( \tilde{P}_k R_k \tilde{P}_k^H + S_k \right)^{-1} \]
\[ \times \left( \text{vec}(Y_k) - \tilde{P}_k \text{vec}(\bar{H}_k) - \text{vec}(\bar{N}_k) \right) \]

\[ \text{MSE} = \text{tr} \left\{ \left( R_k^{-1} + \tilde{P}_k^H S_k^{-1} \tilde{P}_k \right)^{-1} \right\} \]

where \( \tilde{P}_k = (P_k^T \otimes I) \)

- Linear/Affine (also LMMSE for non-Gaussian systems)
- Mean values do not affect the MSE
- Training matrix is clearly affecting the MSE
Channel Matrix Estimation (3)

- MSE minimizing training sequence design

\[
\min_{\mathbf{P}_k} \text{tr}\left\{ \left( \mathbf{R}_k^{-1} + (\mathbf{P}_k^T \otimes \mathbf{I}) \mathbf{H} \mathbf{S}_k^{-1} (\mathbf{P}_k^T \otimes \mathbf{I}) \right)^{-1} \right\}
\]

subject to \( \text{tr}(\mathbf{P}_k^H \mathbf{P}_k) = \mathcal{P} \).

- Training for multiple users: \( \mathbf{P}_k = \sqrt{\mathcal{P}/n_T} \mathbf{I} \)
- Training for single user:
  - Adapt training to channel and disturbance statistics
  - More training power in eigenmodes with strong SINRs

- Kronecker separability necessary for analysis: (dropped indices)

\[
\mathbf{R} = \mathbf{R}_T^T \otimes \mathbf{R}_R \quad \text{Transmit side} \quad \text{Receive side}
\]
\[
\mathbf{S} = \mathbf{S}_T^T \otimes \mathbf{S}_R \quad \text{Temporal} \quad \text{Receive side}
\]
Channel Matrix Estimation (4)

- **Eigenvalue decompositions:**
  \[ R_T = U_T \text{diag}(\lambda_1^{(T)}, \ldots, \lambda_{n_T}^{(T)}) U_T^H \]
  \[ S_T = V_T \text{diag}(\sigma_1^{(T)}, \ldots, \sigma_{n_T}^{(T)}) V_T^H \]

  - Opposite ordering of eigenvalues

- **Conclusions from training optimization:**
  - Matrix structure: \( P = U_T \text{diag}(\sqrt{p_1}, \ldots, \sqrt{p_{n_T}}) V_T^H \)

  Eigenvectors from transmit channel and temporal covariance (\( R_T \) and \( S_T \)), in opposite order of dominance.

  **Conclusion:** Separation in \( n_T \) virtual channels with SINRs \( p_j \lambda_j^{(T)} / \sigma_j^{(T)} \) where large \( \lambda_j^{(T)} \) assigned to small \( \sigma_j^{(T)} \).
Channel Matrix Estimation (5)

• Additional conclusions:
  - MSE with optimal training is Schur-concave in transmit channel covariance eigenvalues.

  *Conclusion*: It is good to have a spread of eigenvalues, since spatial correlation improves estimation.

  - Asymptotics:
    - Low SINR: All power in strongest eigenmode
    - High SINR: Proportional to noise standard deviation

\[
P = U_T \text{diag}(\sqrt{p_1}, \ldots, \sqrt{p_{nT}}) V_T^H
\]
Channel Matrix Estimation (6)

- Mathematical tools used in the proofs
  - MSE is a convex function (w.r.t. training powers)
  - Majorization theory
    - Training matrix based on channel/disturbance eigenvectors
    - Strong channel eigenvalues allocated to weak disturbance
    - Spatial correlation improves estimation performance
  - Convex optimization
    - Asymptotic optimal training, high/low SINR
    - Explicit optimal solutions in certain cases (next slide)
Channel Matrix Estimation (7)

- Training powers explicitly in certain cases
  - Identity as transmit channel and temporal covariance:
    \[ R_T = S_T = I \]
    Result: Equal power allocation: \[ p_j = P/n_T \forall j \]
  - Identical receive covariance for channel and disturbance:
    \[ R_R = S_R \]
    Result: Convex optimization problem
    \[ p_j = \max \left( \sqrt{\frac{\sigma_j^{(T)}}{\alpha}} - \frac{\sigma_j^{(T)}}{\lambda_j^{(T)}}, 0 \right) \]
    \[ \alpha \text{ Lagrange multiplier} \]
    Gives good numerical results also for general covariance
Length of training sequence

- Waterfilling of training powers
  - For low training power $P$ and/or large eigenvalue spread, some training powers will become zero.
  - In the previous case, $P$ is full rank if
    \[
    P > \sum_{j=1}^{n_T-1} \frac{\sqrt{\sigma_j(T)\sigma_{n_T}(T)}}{\lambda_{n_T}(T)} - \frac{\sigma_j(T)}{\lambda_j(T)}
    \]
    and otherwise have rank $m < n_T$ where
    \[
    \sum_{j=1}^{m-1} \frac{\sqrt{\sigma_j(T)\sigma_m(T)}}{\lambda_m(T)} - \frac{\sigma_j(T)}{\lambda_j(T)} < P \leq \sum_{j=1}^{m} \frac{\sqrt{\sigma_j(T)\sigma_{m+1}(T)}}{\lambda_{m+1}(T)} - \frac{\sigma_j(T)}{\lambda_j(T)}.
    \]
Length of training sequence (2)

- If \( \text{rank}(P) = m < n_T \)
  - \( m \) is the maximal necessary training sequence length
    (only approximately if disturbance contains information)

- Example:
Length of training sequence (3)

• Conclusion:
  ▪ The optimal number of training symbols can be smaller than the number of transmit antennas (in spatially correlated systems, limited power)

How is this related to:

“When the training and data powers are allowed to vary, we show that the optimal number of training symbols is equal to the number of transmit antennas”


▪ Their result is shown for uncorrelated systems, but the result have been cited for other applications!
Channel Norm Estimation

- MMSE Estimation of $\|H\|_F^2$ in similar way
  - Considerably more difficult to analyze
  - We limit ourself to zero-mean Kronecker channels
  - No previous results in the area, by our knowledge

- Conjecture: Structure of training matrix
  - Same structure as in the channel matrix estimation
  - Makes it possible to estimate $\|H\|_F^2$ as a sum of independent variables.
Channel Norm Estimation (2)

- MMSE estimator of $\rho = \|H\|_F^2$ and MSE:

$$\hat{\rho}_{\text{MMSE}} = 1^T B \Sigma 1 + \tilde{y}^H \tilde{D} B^2 \tilde{D} \tilde{y}$$

$$E\{|\rho - \hat{\rho}_{\text{MMSE}}|^2\} = 1^T B (\tilde{\Sigma}^2 + 2 \tilde{D} \tilde{\Sigma} \tilde{\Lambda} \tilde{D}) B 1$$

where

- $\tilde{y} = \text{vec}(U_R^H YV_T \Pi)$,
- $B = \Lambda (\tilde{D} \Lambda \tilde{D} + \Sigma)^{-1}$,
- $\tilde{D} = (\text{diag}(\sqrt{p_1}, \ldots, \sqrt{p_{n_T}}) \otimes I)$,
- $\Lambda = \Lambda_T \otimes \Lambda_R$,
- $\Sigma = (\Pi \Sigma_T \Pi^T) \otimes (\Pi \Sigma_R \Pi^T)$,
- $1 = [1, \ldots, 1]^T$. 
Channel Norm Estimation (3)

• Training sequence design
  - More difficult to solve since the MSE is not convex

• Two approaches:
  - Small set of potential explicit solutions can be derived
    (in the case of $R_R = S_R$, otherwise approximately)
  - An additional constraint can make the problem convex

• Asymptotic results:
  - Low SINR: All power allocated to strongest eigenmode
  - High SINR: Proportional distribution to standard deviation
    of channel and disturbance eigenmodes
    *(different from the channel matrix case)*
Numerical Examples

• Numerical illustrations of performance
  ▪ MMSE estimators compared with other estimators
  ▪ Uniform training compared with optimal training

• System parameters
  ▪ Kronecker-structure of covariance matrices
  ▪ Uncolored disturbance (noise-limited system)
  ▪ Transmit and receive channel covariance modeled with exponential model (model of a Uniform Linear Array)
Numerical Examples (2)

• Channel Matrix Estimation
  - Comparison of four different estimators, optimal training
  - Normalized MSE: $E\{\|H - \hat{H}_{\text{MMSE}}\|_F^2\}/\text{tr}(R)$

---

8 Transmit Antennas

4 Receive Antennas

Correlation Parameter: 0.8

2009-02-11
Numerical Examples (3)

- Channel Matrix Estimation
  - Comparison of different training sequences
  - Normalized MSE: \[ E\left\{ \|H - \widehat{H}_{\text{MMSE}}\|_F^2 \right\}/\text{tr}(R) \]

8 Transmit Antennas
4 Receive Antennas
Correlation Parameter: 0.8
Numerical Examples (4)

- Channel Squared Norm Estimation
  - Comparison of MMSE and indirect estimation
  - Normalized MSE: \( E\{\|H\|_F^2 - \|\hat{H}_{\text{MMSE}}\|_F^2\}/\text{tr}(RR^H) \)

8 Transmit Antennas (corr: 0.8)

4 Receive Antennas (corr: 0)
Summary

• Training-based channel estimation
  ▪ Narrowband multi-antenna system
  ▪ Channel Matrix, Channel Squared Norm

• MMSE Estimation
  ▪ The general MMSE estimators becomes linear

• MSE minimizing training design
  ▪ Training matrix is a weighting of eigenmodes of channel and disturbance covariance matrices
  ▪ Waterfilling structure on power allocation
  ▪ Explicit power allocation results: For certain covariance structures and asymptotically.
Summary (2)

• Main contributions:
  ▪ **Channel Matrix Estimation**:  
    • Generalization to Rician channels/disturbance  
    • Unification of previous results:  
      Which results depend on which assumptions?  
    • Analysis of the optimal training length  
    • Identification of mistakes in other papers  
  ▪ **Channel Norm Estimation**:  
    • Novel estimation and training optimization results  
    • Clear gain in MSE compared to indirect estimation
References


