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

Emil Björnson and Björn Ottersten

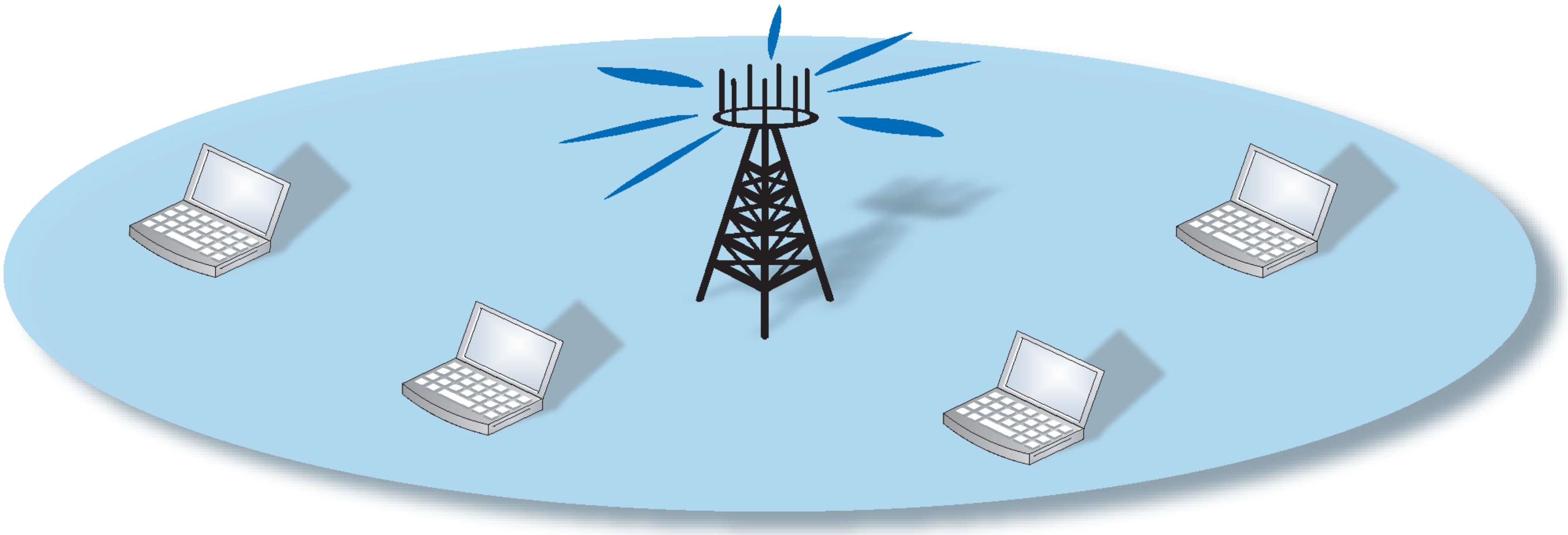
ACCESS Linnaeus Center, Signal Processing

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



ROYAL INSTITUTE OF TECHNOLOGY

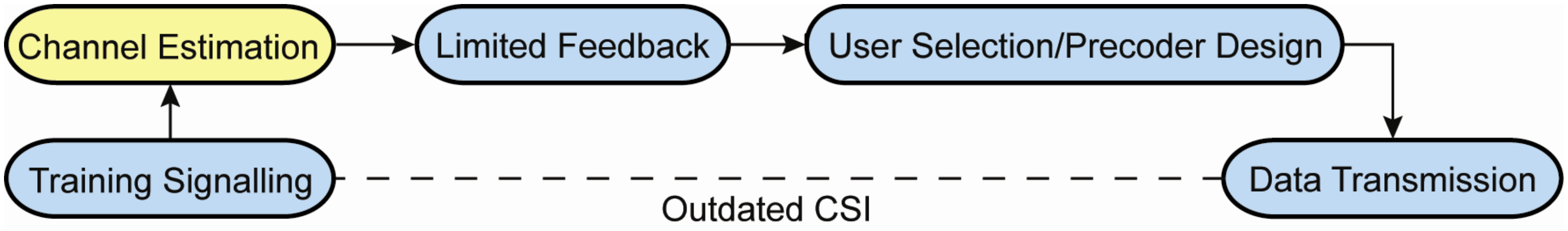
Overview (2)

- 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

ROYAL INSTITUTE OF TECHNOLOGY

Overview (3)

- System Operation (Block Fading):



ROYAL INSTITUTE OF TECHNOLOGY

- 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

ROYAL INSTITUTE OF TECHNOLOGY

System Model

- MIMO Communication:
  - n_T transmit antennas, n_R receive antennas

ROYAL INSTITUTE OF TECHNOLOGY

- Communication model to user k:

$$\underbrace{\mathbf{y}_k(t)}_{n_R \times 1} = \underbrace{\mathbf{H}_k}_{n_R \times n_T} \cdot \underbrace{\mathbf{x}(t)}_{n_T \times 1} + \underbrace{\mathbf{n}_k(t)}_{n_R \times 1}$$

  - $\mathbf{x}(t)$ transmitted signal, $\mathbf{y}_k(t)$ received signal
  - $\mathbf{n}_k(t)$ potentially correlated complex Gaussian noise
  - Rician distributed channel matrix:

$$\mathrm{vec}(\mathbf{H}_k) \in \mathcal{CN}(\mathrm{vec}(\bar{\mathbf{H}}_k), \mathbf{R}_k)$$

System Model (2)

- Problem description:
  - Estimate properties of the channel matrix $\mathbf{H}_k$
  - In general, we are interested in some function $f(\mathbf{H}_k)$ (receiver structure, modulation, precoding)

- In this work we estimate two parameters
  - $\mathbf{H}_k$ channel matrix (many applications)
  - $\|\mathbf{H}_k\|_F^2$ channel gain (for user-selection, rate adaptation)

  - It will be illustrated that calculation of $\|\mathbf{H}_k\|_F^2$ from an estimation of $\mathbf{H}_k$ gives poor performance

ROYAL INSTITUTE OF TECHNOLOGY

System Model (3)

ROYAL INSTITUTE OF TECHNOLOGY

- 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: $\mathrm{tr}(\mathbf{P}_k^H \mathbf{P}_k) = \mathcal{P}$
  - Transmission of $\mathbf{P}_k$ over n_T symbol slots:

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

$$\mathbf{Y}_k = [\mathbf{y}_k(1), \ldots, \mathbf{y}_k(n_T)], \, \mathbf{N}_k = [\mathbf{n}_k(1), \ldots, \mathbf{n}_k(n_T)]$$

  - General disturbance statistics:

$$\mathrm{vec}(\mathbf{N}_k) \in \mathcal{CN}(\mathrm{vec}(\bar{\mathbf{N}}_k), \mathbf{S}_k)$$

Kronecker Product

- Definition:

$$\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{11}\mathbf{B} & \dots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{bmatrix}$$

ROYAL INSTITUTE OF TECHNOLOGY

- Useful to analyze matrix multiplication:

$$\text{vec}(\mathbf{C}\mathbf{H}_k\mathbf{D}) = (\mathbf{D}^T \otimes \mathbf{C})\,\text{vec}(\mathbf{H}_k)$$

  - Training matrix $\mathbf{P}_k$ multiplied from the right:

$$\text{vec}(\mathbf{H}_k\mathbf{P}_k) = (\mathbf{P}_k^T \otimes \mathbf{I})\,\text{vec}(\mathbf{H}_k)$$

  - To analyze impact of $\mathbf{P}_k$ we will later assume Kronecker-structured channel properties and use that

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{E} \otimes \mathbf{F}) = (\mathbf{AE} \otimes \mathbf{BF})$$

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:

- **Rayleigh fading, uncolored noise:** J. Kotecha and A. Sayeed, "*Transmit signal design for optimal estimation of correlated MIMO channels*," 2004.
- **Rayleigh fading, colored noise:** Y. Liu, T. Wong, and W. Hager, "*Training signal design for estimation of correlated MIMO channels with colored interference*," 2007.

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

- **Rayleigh fading, uncolored noise:** M. Biguesh and A. Gershman, "*Training-based MIMO channel estimation: a study of estimator tradeoffs and optimal training signals*," 2006.
- **Rayleigh fading, colored noise:** D. Katselis, E. Kofidis, and S. Theodoridis, "*Training-based estimation of correlated MIMO fading channels in the presence of colored interference*," 2007.

ROYAL INSTITUTE OF TECHNOLOGY

General linear estimator

$$\mathbf{B}\text{vec}(\mathbf{Y}_k)$$

Suboptimal estimator

$$\mathbf{Y}_k\mathbf{A} \Leftrightarrow (\mathbf{A}^T \otimes \mathbf{I})\text{vec}(\mathbf{Y}_k)$$

Channel Matrix Estimation (2)

- MMSE Estimator:

$$\mathrm{vec}(\widehat{\mathbf{H}}_{\mathrm{MMSE}}) =$$
$$\mathrm{vec}(\bar{\mathbf{H}}_k) + \mathbf{R}_k \widetilde{\mathbf{P}}_k^H \left(\widetilde{\mathbf{P}}_k \mathbf{R}_k \widetilde{\mathbf{P}}_k^H + \mathbf{S}_k \right)^{-1}$$
$$\times \left(\mathrm{vec}(\mathbf{Y}_k) - \widetilde{\mathbf{P}}_k \mathrm{vec}(\bar{\mathbf{H}}_k) - \mathrm{vec}(\bar{\mathbf{N}}_k) \right)$$

$$\mathrm{MSE} = \mathrm{tr} \left\{ \left(\mathbf{R}_k^{-1} + \widetilde{\mathbf{P}}_k^H \mathbf{S}_k^{-1} \widetilde{\mathbf{P}}_k \right)^{-1} \right\}$$

where $\widetilde{\mathbf{P}}_k = (\mathbf{P}_k^T \otimes \mathbf{I})$

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

KTH VETENSKAP OCH KONST
ROYAL INSTITUTE OF TECHNOLOGY

Channel Matrix Estimation (3)

- MSE minimizing training sequence design

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

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

  - 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)

$$\mathbf{R} = \underbrace{\boxed{\mathbf{R}_T^T}}_{\text{Transmit side}} \otimes \underbrace{\boxed{\mathbf{R}_R}}_{\text{Receive side}} \qquad \mathbf{S} = \underbrace{\boxed{\mathbf{S}_T^T}}_{\text{Temporal}} \otimes \underbrace{\boxed{\mathbf{S}_R}}_{\text{Receive side}}$$

ROYAL INSTITUTE OF TECHNOLOGY

Channel Matrix Estimation (4)

- Eigenvalue decompositions:

$$\mathbf{R}_T = \mathbf{U}_T \,\text{diag}(\lambda_1^{(T)}, \ldots, \lambda_{n_T}^{(T)}) \mathbf{U}_T^H$$
$$\mathbf{S}_T = \mathbf{V}_T \,\text{diag}(\sigma_1^{(T)}, \ldots, \sigma_{n_T}^{(T)}) \mathbf{V}_T^H$$

  - Opposite ordering of eigenvalues

- Conclusions from training optimization:
  - Matrix structure: $\mathbf{P} = \mathbf{U}_T \text{diag}(\sqrt{p_1}, \ldots, \sqrt{p_{n_T}}) \mathbf{V}_T^H$
    Eigenvectors from transmit channel and temporal covariance ($\mathbf{R}_T$ and $\mathbf{S}_T$), 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 $\sigma_j^{(T)}$

ROYAL INSTITUTE OF TECHNOLOGY

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

$$\mathbf{P} = \mathbf{U}_T \text{diag}(\sqrt{p_1}, \ldots, \sqrt{p_{n_T}}) \mathbf{V}_T^H$$

ROYAL INSTITUTE OF TECHNOLOGY

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)

ROYAL INSTITUTE OF TECHNOLOGY

Channel Matrix Estimation (7)

- Training powers explicitly in certain cases
  - Identity as transmit channel and temporal covariance:

$$\mathbf{R}_T = \mathbf{S}_T = \mathbf{I}$$

Result: Equal power allocation: $p_j = \mathcal{P}/n_T \ \forall j$

  - Identical receive covariance for channel and disturbance:

$$\mathbf{R}_R = \mathbf{S}_R$$

Result: Convex optimization problem

$$p_j = \max\left(\sqrt{\frac{\sigma_j^{(T)}}{\alpha}} - \frac{\sigma_j^{(T)}}{\lambda_j^{(T)}}, 0\right) \qquad \alpha \text{ Lagrange multiplier}$$

Gives good numerical results also for general covariance

Length of training sequence

- Waterfilling of training powers
  - For low training power $\mathcal{P}$ and/or large eigenvalue spread, some training powers will become zero.
  - In the previous case, $\mathbf{P}$ is full rank if

$$\mathcal{P} > \sum_{j=1}^{n_T-1} \frac{\sqrt{\sigma_j^{(T)}\sigma_{n_T}^{(T)}}}{\lambda_{n_T}^{(T)}} - \frac{\sigma_j^{(T)}}{\lambda_j^{(T)}}$$

and otherwise have rank $m < n_T$ where

$$\sum_{j=1}^{m-1} \frac{\sqrt{\sigma_j^{(T)}\sigma_m^{(T)}}}{\lambda_m^{(T)}} - \frac{\sigma_j^{(T)}}{\lambda_j^{(T)}} < \mathcal{P} \leq \sum_{j=1}^{m} \frac{\sqrt{\sigma_j^{(T)}\sigma_{m+1}^{(T)}}}{\lambda_{m+1}^{(T)}} - \frac{\sigma_j^{(T)}}{\lambda_j^{(T)}}.$$

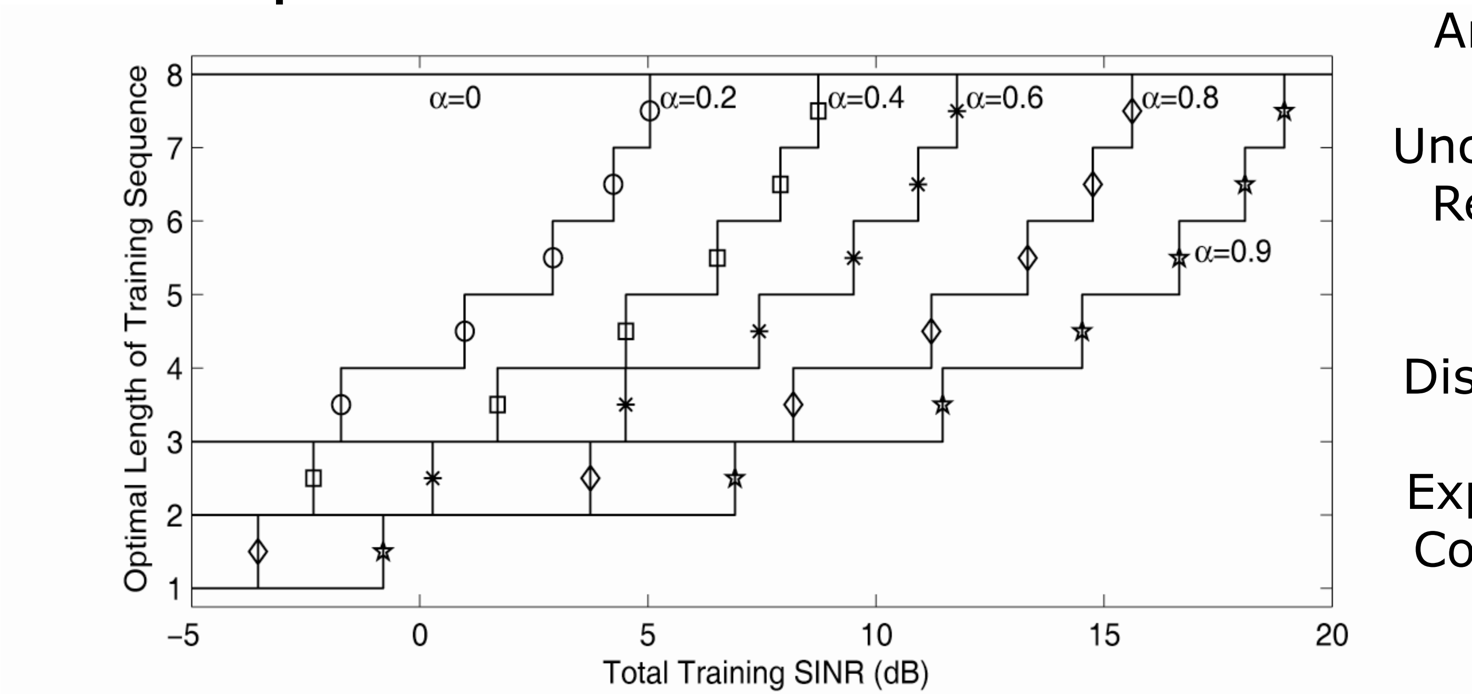
ROYAL INSTITUTE OF TECHNOLOGY

Length of training sequence (2)

- If rank$(\mathbf{P}) = m < n_T$
  - m is the maximal necessary training sequence length

    (only approximately if disturbance contains information)

- Example:



8 Transmit Antennas

Uncorrelated Receivers

White Disturbance

Exponential Correlation Model

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"

B. Hassibi and B. Hochwald, "*How much training is needed in multiple-antenna wireless links?*," 2003.

- Their result is shown for uncorrelated systems, but the result have been cited for other applications!

ROYAL INSTITUTE OF TECHNOLOGY

Channel Norm Estimation

- MMSE Estimation of $\|\mathbf{H}\|_F^2$ 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

ROYAL INSTITUTE OF TECHNOLOGY

- Conjecture: Structure of training matrix

$$\mathbf{P} = \mathbf{U}_T \text{diag}(\sqrt{p_1}, \ldots, \sqrt{p_{n_T}}) \mathbf{V}_T^H$$

  - Same structure as in the channel matrix estimation
  - Makes it possible to estimate $\|\mathbf{H}\|_F^2$ as a sum of independent variables.

Channel Norm Estimation (2)

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

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

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

where

$\tilde{\mathbf{y}} = \text{vec}(\mathbf{U}_R^H \mathbf{Y} \mathbf{V}_T \boldsymbol{\Pi})$,

$\mathbf{B} = \boldsymbol{\Lambda}(\widetilde{\mathbf{D}} \boldsymbol{\Lambda} \widetilde{\mathbf{D}} + \boldsymbol{\Sigma})^{-1}$,

$\widetilde{\mathbf{D}} = (\text{diag}(\sqrt{p_1}, \ldots, \sqrt{p_{n_T}}) \otimes \mathbf{I})$,

$\boldsymbol{\Lambda} = \boldsymbol{\Lambda}_T \otimes \boldsymbol{\Lambda}_R$,

$\boldsymbol{\Sigma} = (\boldsymbol{\Pi} \boldsymbol{\Sigma}_T \boldsymbol{\Pi}^T) \otimes (\widetilde{\boldsymbol{\Pi}} \boldsymbol{\Sigma}_R \widetilde{\boldsymbol{\Pi}}^T)$,

$\mathbf{1} = [1, \ldots, 1]^T$.

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 $\mathbf{R}_R = \mathbf{S}_R$, 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)*

ROYAL INSTITUTE OF TECHNOLOGY

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)

ROYAL INSTITUTE OF TECHNOLOGY

Numerical Examples (2)

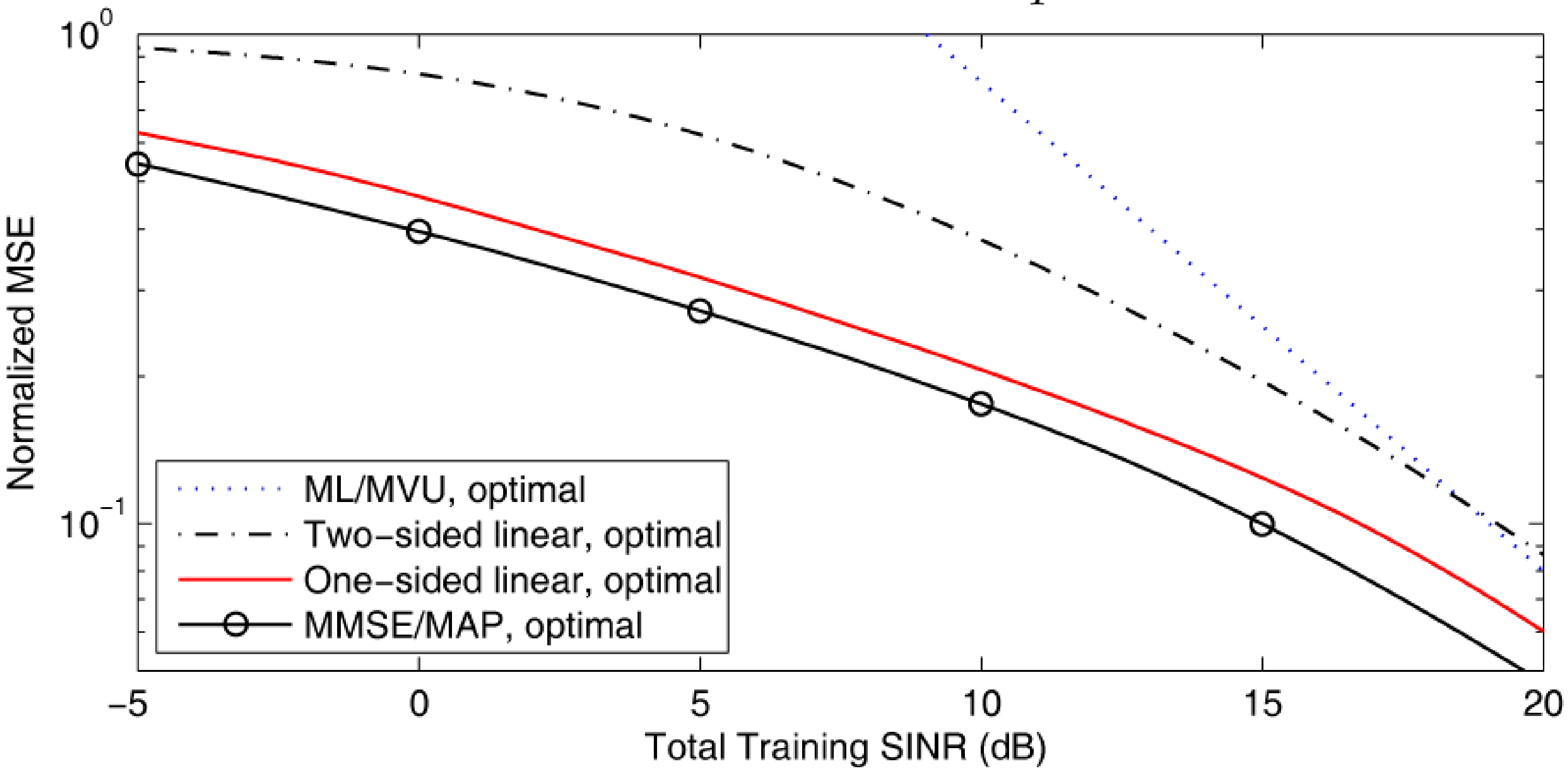
- Channel Matrix Estimation
  - Comparison of four different estimators, optimal training
  - Normalized MSE: $E\{\|\mathbf{H}-\widehat{\mathbf{H}}_{\text{MMSE}}\|_F^2\}/\text{tr}(\mathbf{R})$

ROYAL INSTITUTE OF TECHNOLOGY

8 Transmit Antennas

4 Receive Antennas

Correlation Parameter: 0.8



Numerical Examples (3)

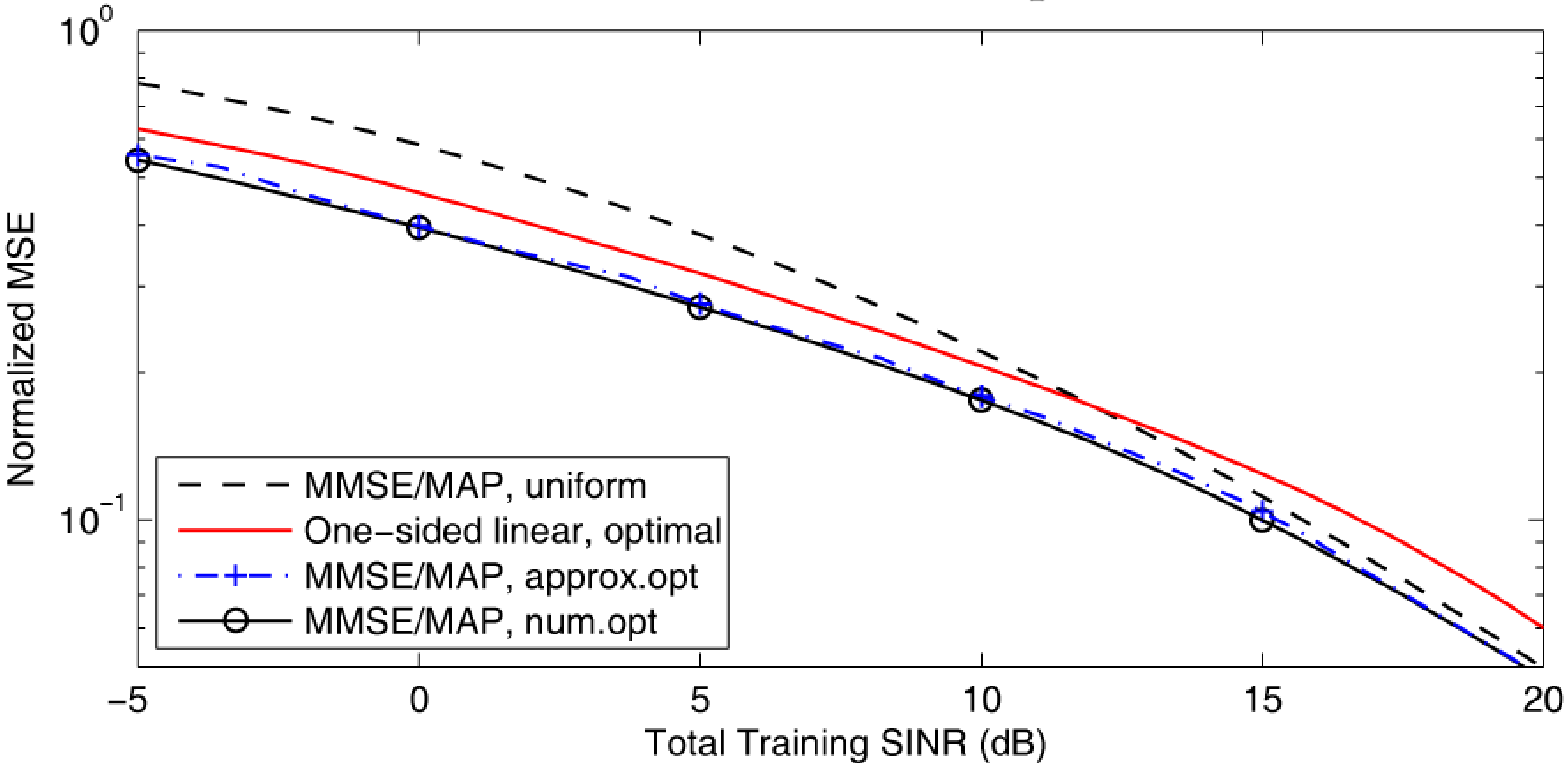
- Channel Matrix Estimation
  - Comparison of different training sequences
  - Normalized MSE: $E\{\|\mathbf{H}-\widehat{\mathbf{H}}_{\text{MMSE}}\|_F^2\}/\text{tr}(\mathbf{R})$

ROYAL INSTITUTE OF TECHNOLOGY

8 Transmit Antennas

4 Receive Antennas

Correlation Parameter: 0.8

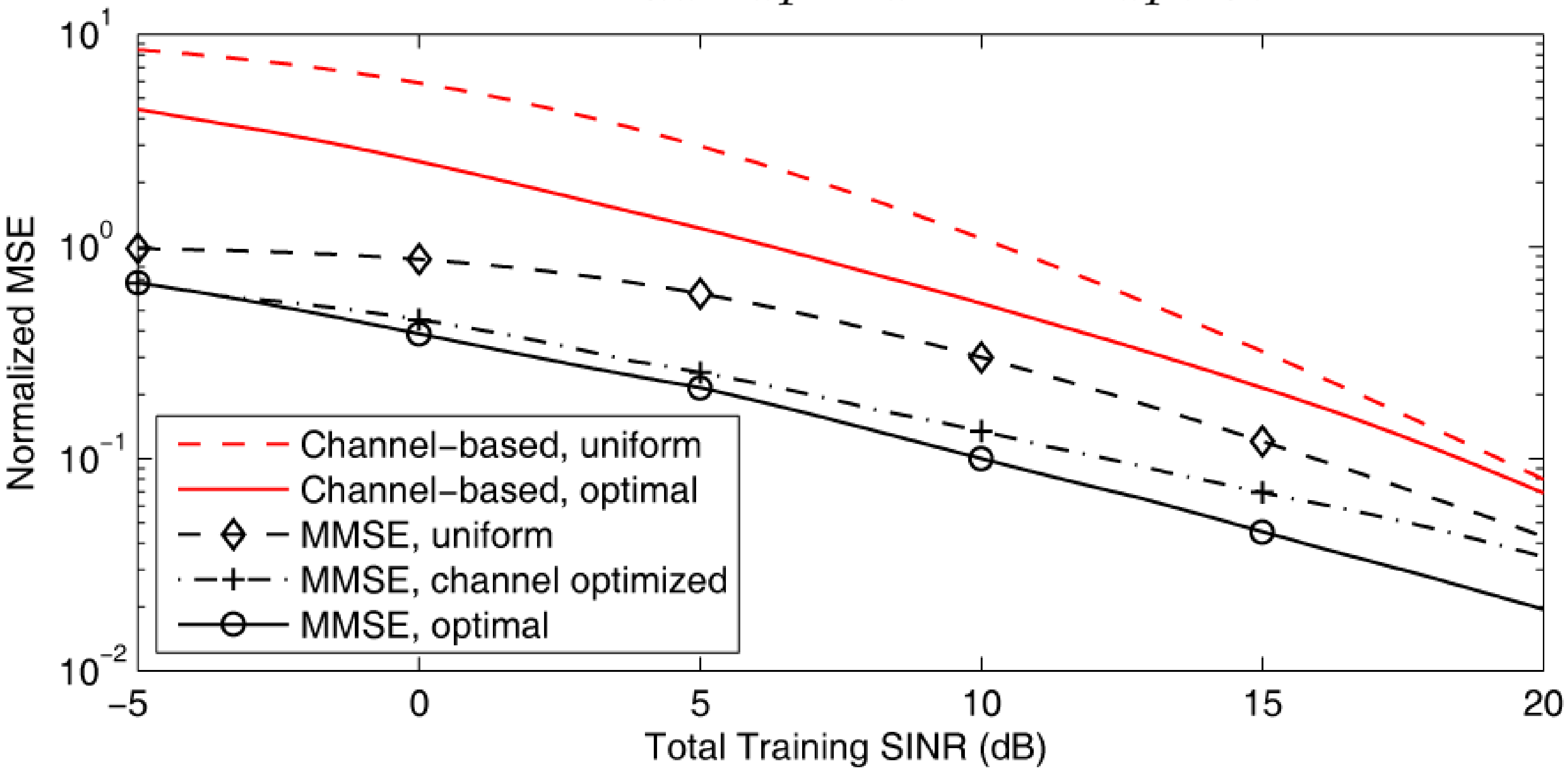


Numerical Examples (4)

- Channel Squared Norm Estimation
  - Comparison of MMSE and indirect estimation
  - Normalized MSE: $E\{|\|\mathbf{H}\|_F^2 - \|\widehat{\mathbf{H}}_{\mathrm{MMSE}}\|_F^2|^2\}/\mathrm{tr}(\mathbf{R}\mathbf{R}^H)$

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.

ROYAL INSTITUTE OF TECHNOLOGY

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

ROYAL INSTITUTE OF TECHNOLOGY

References

[1] E. Björnson and B. Ottersten, “*A Unified Framework for Training-Based Estimation in Arbitrarily Correlated Rician MIMO Channels with Rician Disturbance*,” Submitted to IEEE Trans. Signal Process., Dec 2008.

[2] E. Björnson and B. Ottersten, “*Training-Based Bayesian MIMO Channel and Channel Norm Estimation*”, IEEE ICASSP’09.

[3] J. Kotecha and A. Sayeed, “Transmit signal design for optimal estimation of correlated MIMO channels,” IEEE Trans. Signal Process., vol. 52, pp. 546–557, 2004.

[4] Y. Liu, T. Wong, and W. Hager, “Training signal design for estimation of correlated MIMO channels with colored interference,” IEEE Trans. Signal Process., vol. 55, pp. 1486–1497, 2007.

ROYAL INSTITUTE OF TECHNOLOGY