Optimal algorithms and CRB for reciprocity calibration in Massive MIMO

K. GOPALA D.T.M. SLOCK

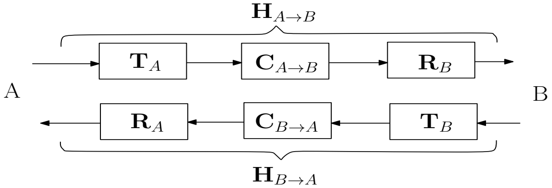
EURECOM Sophia Antipolis Communication Systems Department

Massive and mmWave MIMO, ICASSP, 2018

Outline

TDD Massive MIMO and Channel Reciprocity

- A Massive number of Transmit (Tx) Antennas at the base station (BS) talking to a small number of user equipment (UE) in Time Division Duplexing (TDD) mode.
- Channel observed in the digital domain from BS to UE (in general, from antennas A to antennas B) not reciprocal.



- $\mathbf{C}_{A\to B}$ and $\mathbf{C}_{B\to A}$ model the reciprocal propagation channels.
- Diagonal Matrices $\mathbf{T}_A$, $\mathbf{R}_A$, $\mathbf{T}_B$, $\mathbf{R}_B$ model the response of the transmit and receive RF front-ends

TDD Massive MIMO and Channel Reciprocity (Cont'd)

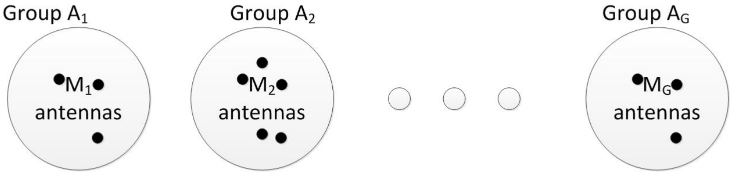
- $\mathbf{H}_{A\to B} = \mathbf{R}_B \mathbf{C}_{A\to B} \mathbf{T}_A, \qquad \mathbf{H}_{B\to A} = \mathbf{R}_A \mathbf{C}_{B\to A} \mathbf{T}_B,$
- DL (Downlink) channel is derived from the UL (Uplink) channel for Tx beamforming.
- As $\mathbf{C}_{A\to B} = \mathbf{C}_{B\to A}^T$,

$$\mathbf{H}_{A\to B} = \underbrace{\mathbf{R}_B \mathbf{T}_B^{-T}}_{\mathbf{F}_B^{-T}} \mathbf{H}_{B\to A}^T \underbrace{\mathbf{R}_A^{-T} \mathbf{T}_A}_{\mathbf{F}_A} = \mathbf{F}_B^{-T} \mathbf{H}_{B\to A}^T \mathbf{F}_A. \tag{1}$$

Relative calibration factor.

- It has been shown that calibration at the BS is more crucial for Multi-user MIMO performance.
- Hence, in what follows, calibration will be done internally at the BS without involving the UE.

General Framework and antenna grouping at BS



- Partition the M antennas into G groups with M_i antennas each.
- Proposed in Jiang et al., "A Framework for Over-the-air Reciprocity Calibration for TDD Massive MIMO Systems".
- Each group A_i transmits pilots $\mathbf{P}_i$ for L_i channel uses.

$$\text{Bidirectional Tx} \begin{cases} \mathbf{Y}_{i\to j} = \mathbf{R}_j \mathbf{C}_{i\to j} \mathbf{T}_i \mathbf{P}_i + \mathbf{N}_{i\to j}, \\ \mathbf{Y}_{j\to i} = \mathbf{R}_i \mathbf{C}_{j\to i} \mathbf{T}_j \mathbf{P}_j + \mathbf{N}_{j\to i}, \end{cases} \quad (2)$$

General Framework

- Eliminating the propagation channel $\mathbf{C}$ and after some matrix manipulations, we get

$$\mathcal{Y}(\mathbf{P})\mathbf{f} = \widetilde{\mathbf{n}}, \qquad \mathbf{F} = \text{diag}\{\mathbf{f}\} = \mathbf{R}^{-T}\mathbf{T}.$$

$$\mathcal{Y}(\mathbf{P}) = \underbrace{\begin{bmatrix} (\mathbf{Y}_{2\to1}^T * \mathbf{P}_1^T) & -(\mathbf{P}_2^T * \mathbf{Y}_{1\to2}^T) & 0 & \cdots \\ (\mathbf{Y}_{3\to1}^T * \mathbf{P}_1^T) & 0 & -(\mathbf{P}_3^T * \mathbf{Y}_{1\to3}^T) & \cdots \\ 0 & (\mathbf{Y}_{3\to2}^T * \mathbf{P}_2^T) & -(\mathbf{P}_3^T * \mathbf{Y}_{2\to3}^T) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}}_{(\sum_{j=2}^{G}\sum_{i=1}^{j-1} L_i L_j)\times M}.$$

where $*$ denotes the Khatri–Rao product (or column-wise Kronecker product[1]). $\widetilde{\mathbf{n}}$ is colored noise.

[1] $\mathbf{A} = \begin{bmatrix}\mathbf{a}_1 & \mathbf{a}_2 & \dots & \mathbf{a}_M\end{bmatrix}$ and $\mathbf{B} = \begin{bmatrix}\mathbf{b}_1 & \mathbf{b}_2 & \dots & \mathbf{b}_M\end{bmatrix}$ where $\mathbf{a}_i$ and $\mathbf{b}_i$ are column vectors for $i \in 1\dots M$, then, $\mathbf{A} * \mathbf{B} = \begin{bmatrix}\mathbf{a}_1\otimes\mathbf{b}_1 & \mathbf{a}_2\otimes\mathbf{b}_2 & \dots & \mathbf{a}_M\otimes\mathbf{b}_M\end{bmatrix}$.

General Framework (Cont'd)

$$\mathcal{Y}(\mathbf{P})\mathbf{f} = \widetilde{\mathbf{n}}$$

- Least squares (LS) estimate of $\mathbf{f}$

$$\begin{aligned}\hat{\mathbf{f}} &= \arg\min_{\mathbf{f}} \|\mathcal{Y}(\mathbf{P})\,\mathbf{f}\|^2 \\ &= \arg\min_{\mathbf{f}} \sum_{i<j} \|(\mathbf{Y}_{j\to i}^T * \mathbf{P}_i^T)\mathbf{f}_i - (\mathbf{P}_j^T * \mathbf{Y}_{i\to j}^T)\mathbf{f}_j\|^2 \end{aligned}, \quad (3)$$

- To exclude the trivial solution, $\hat{\mathbf{f}} = \mathbf{0}$, this needs to be augmented with a constraint

$$\mathcal{C}(\hat{\mathbf{f}}, \mathbf{f}) = 0.$$

General Framework (Cont'd)

- Typical choices for the constraint are
  1) Norm plus phase constraint (NPC):

$$\text{norm: } \mathrm{Re}\{\mathcal{C}(\hat{\mathbf{f}}, \mathbf{f})\} = ||\hat{\mathbf{f}}||^2 - c\,, \; c = ||\mathbf{f}||^2, \tag{4}$$

$$\text{phase: } \mathrm{Im}\{\mathcal{C}(\hat{\mathbf{f}}, \mathbf{f})\} = \mathrm{Im}\{\hat{\mathbf{f}}^H \mathbf{f}\} = 0. \tag{5}$$

  2) Linear constraint:

$$\mathcal{C}(\hat{\mathbf{f}}, \mathbf{f}) = \hat{\mathbf{f}}^H \mathbf{g} - c = 0\,. \tag{6}$$

  The most popular linear constraint is the First Coefficient Constraint (FCC), which is (6) with $\mathbf{g} = \mathbf{e}_1$, $c = 1$.

- We now proceed to investigate some optimal estimation methods and performance limits.

Cramer Rao Bound (CRB) - Defining an auxiliary channel

- Re-parameterize in terms of the relative calibration factors and an auxiliary internal channel.
- As $\mathbf{F}_i = \mathbf{R}_i^{-T}\mathbf{T}_i$,

$$\mathbf{Y}_{i\to j} = \mathbf{R}_j\mathbf{C}_{i\to j}\mathbf{T}_i\mathbf{P}_i + \mathbf{N}_{i\to j} = \underbrace{\mathbf{R}_j\mathbf{C}_{i\to j}\mathbf{R}_i^T}_{\mathcal{H}_{i\to j}}\mathbf{F}_i\mathbf{P}_i + \mathbf{N}_{i\to j}. \quad (7)$$

Auxiliary internal channel

- $\mathcal{H}_{i\to j}$ is a nuisance parameter and does not correspond to any physically measurable quantity.
- $\mathcal{H}_{i\to j}$ is reciprocal: $\mathcal{H}_{i\to j} = \mathcal{H}_{j\to i}^T$.

Cramer Rao Bound

- Stacking these observations into a vector

$$\begin{aligned}\mathbf{y} &= \left[\mathrm{vec}(\mathbf{Y}_{1\to 2})^T\,\mathrm{vec}(\mathbf{Y}_{2\to 1}^T)^T\,\mathrm{vec}(\mathbf{Y}_{1\to 3})^T \ldots\right]^T \\ &= \mathcal{H}(\mathbf{h},\mathbf{P})\mathbf{f} + \mathbf{n} = \mathcal{F}(\mathbf{f},\mathbf{P})\mathbf{h} + \mathbf{n},\end{aligned} \tag{8}$$

where $\mathbf{h} = \left[\mathrm{vec}(\mathcal{H}_{1\to 2})^T\,\mathrm{vec}(\mathcal{H}_{1\to 3})^T\,\mathrm{vec}(\mathcal{H}_{2\to 3})^T \ldots\right]^T$, $\mathbf{f} = [\mathbf{f}_1,\, \mathbf{f}_2,\, \ldots,\, \mathbf{f}_M]$ (Relative calibration factors for M antennas) and $\mathbf{n}$ is the corresponding noise vector.

Cramer Rao Bound (Cont'd)

- The composite matrices $\mathcal{H}$ and $\mathcal{F}$ are given by,

$$\mathcal{H}(\mathbf{h},\mathbf{P}) = \begin{bmatrix} \mathbf{P}_1^T * \mathcal{H}_{1\to 2} & 0 & 0 & \dots \\ 0 & \mathcal{H}_{1\to 2}^T * \mathbf{P}_2^T & 0 & \dots \\ \mathbf{P}_1^T * \mathcal{H}_{1\to 3} & 0 & 0 & \dots \\ 0 & 0 & \mathcal{H}_{1\to 3}^T * \mathbf{P}_3^T & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

$$\mathcal{F}(\mathbf{f},\mathbf{P}) = \begin{bmatrix} \mathbf{P}_1^T\mathbf{F}_1 \otimes \mathbf{I} & 0 & 0 & 0 & \dots \\ \mathbf{I} \otimes \mathbf{P}_2^T\mathbf{F}_2 & 0 & 0 & 0 & \dots \\ 0 & \mathbf{P}_1^T\mathbf{F}_1 \otimes \mathbf{I} & 0 & 0 & \dots \\ 0 & \mathbf{I} \otimes \mathbf{P}_3^T\mathbf{F}_3 & 0 & 0 & \dots \\ 0 & 0 & \mathbf{P}_2^T\mathbf{F}_2 \otimes \mathbf{I} & 0 & \dots \\ 0 & 0 & \mathbf{I} \otimes \mathbf{P}_3^T\mathbf{F}_3 & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{bmatrix}. \quad (9)$$

Cramer Rao Bound (Cont'd)

$$\mathbf{y} = \mathcal{H}(\mathbf{h}, \mathbf{P})\mathbf{f} + \mathbf{n} = \mathcal{F}(\mathbf{f}, \mathbf{P})\mathbf{h} + \mathbf{n},$$

- The scenario is now identical to that encountered in some blind channel estimation scenarios and hence we can take advantage of some existing tools[2] [3].
- Treating $\mathbf{h}$ and $\mathbf{f}$ as deterministic unknown parameters, $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2\mathbf{I})$, the Fisher Information Matrix (FIM) $\mathbf{J}$ for jointly estimating $\mathbf{f}$ and $\mathbf{h}$ is,

$$\mathbf{J} = \frac{1}{\sigma^2} \begin{bmatrix} \mathcal{H}^H \\ \mathcal{F}^H \end{bmatrix} \begin{bmatrix} \mathcal{H} & \mathcal{F} \end{bmatrix}. \tag{10}$$

---

[2] Carvalho and Slock, "Semi–Blind Methods for FIR Multichannel Estimation".

[3] Carvalho, Omar, and Slock, "Performance and complexity analysis of blind FIR channel identification algorithms based on deterministic maximum likelihood in SIMO systems".

Cramer Rao Bound (Cont'd)

- Real interest is in deriving the CRB for $\mathbf{f}$, in the presence of the nuisance parameters $\mathbf{h}$.
- Incorporating the effect of the constraint $\mathcal{C}$ on $\mathbf{f}$, we can derive from[4], the following constrained CRB for $\mathbf{f}$

$$\text{CRB}_{\mathbf{f}} = \sigma^2 \mathcal{V}_{\mathbf{f}} \left(\mathcal{V}_{\mathbf{f}}^H \mathcal{H}^H \mathcal{P}_{\mathcal{F}}^{\perp} \mathcal{H} \mathcal{V}_{\mathbf{f}} \right)^{-1} \mathcal{V}_{\mathbf{f}}^H \tag{11}$$

where $\mathcal{P}_{\mathcal{F}} = \mathcal{F}(\mathcal{F}^H \mathcal{F})^{\dagger} \mathcal{F}^H$ is the projection operator on the column space of matrix $\mathcal{F}$ and $\mathcal{P}_{\mathcal{F}}^{\perp} = \mathbf{I} - \mathcal{P}_{\mathcal{F}}$.

- $\dagger$ corresponds to the Moore-Penrose pseudo inverse.
- The $M \times (M-1)$ matrix $\mathcal{V}_{\mathbf{f}}$ is such that its column space spans the orthogonal complement of that of $\dfrac{\partial \mathcal{C}(f)}{\partial \mathbf{f}^*}$, i.e., $\mathcal{P}_{\mathcal{V}_{\mathbf{f}}} = \mathcal{P}^{\perp}_{\frac{\partial \mathcal{C}}{\partial \mathbf{f}^*}}$.

[4] Carvalho and Slock, *Cramér-Rao bounds for blind multichannel estimation*.

Maximum Likelihood(ML) & LS

- The log likelihood for the ML estimator may be written as $\frac{1}{\sigma^2}\left\|\mathbf{y}-\mathcal{F}(\mathbf{f},\mathbf{P})\mathbf{h}\right\|^2$.
- Optimizing w.r.t. $\mathbf{h}$ leads to $\mathbf{h}=(\mathcal{F}^H\mathcal{F})^{\dagger}\mathcal{F}^H\mathbf{y}$.
- Substituting back this estimate yields an ML estimator for $\hat{\mathbf{f}}$ minimizing

$$\mathbf{y}^H\mathcal{F}^{\perp}(\mathcal{F}^{\perp H}\mathcal{F}^{\perp})^{\dagger}\mathcal{F}^{\perp H}\mathbf{y}, \tag{12}$$

- Now, it can be shown that,

$$\mathcal{Y}(\mathbf{P})\mathbf{f}=\mathcal{F}^{\perp H}\mathbf{y}=\widetilde{\mathbf{n}}, \tag{13}$$

- Hence, LS estimator minimizes $\left\|\mathcal{Y}(\mathbf{P})\mathbf{f}\right\|^2=\mathbf{y}^H\mathcal{F}^{\perp}\mathcal{F}^{\perp H}\mathbf{y}$.

Maximum Likelihood(ML) & LS Cont'd

ML Metric

$$\mathbf{y}^H \mathcal{F}^{\perp} (\mathcal{F}^{\perp H} \mathcal{F}^{\perp})^{\dagger} \mathcal{F}^{\perp H} \mathbf{y}$$

LS Metric

$$\mathbf{y}^H \mathcal{F}^{\perp} \mathcal{F}^{\perp H} \mathbf{y}$$

- Thus, ML estimator is LS with $\mathbf{F}$ dependent weighting.
- In the case of single antenna grouping, $(\mathcal{F}^{\perp H} \mathcal{F}^{\perp})$ has a diagonal structure and is a multiple of the Identity matrix if all calibration values have equal magnitude.
- This implies that the ML techniques are more useful when the calibration factors vary over a significant range.

Variational Bayes (VB) Estimation

- In VB, a Bayesian estimate is obtained by computing an approximation (variational distribution) to the posterior distribution of the parameters $\mathbf{h}, \mathbf{f}$
- $\mathbf{h}, \mathbf{f}$ priors : $\mathbf{f} \sim \mathcal{CN}(0, \alpha^{-1}\mathbf{I_M})$, $\mathbf{h} \sim \mathcal{CN}(0, \beta^{-1}\mathbf{I_{N_h}})$.
- α, β are assumed to have themselves a uniform prior. N_h is the number of elements in $\mathbf{h}$.
- The variational distribution, is chosen to minimize the Kullback-Leibler distance between the true posterior distribution $p(\mathbf{h}, \mathbf{f}, \alpha, \beta|\mathbf{y})$ and a factored variational distribution

$$p(\mathbf{h}, \mathbf{f}, \alpha, \beta|\mathbf{y}) \approx q_{\mathbf{h}}(\mathbf{h})\, q_{\mathbf{f}}(\mathbf{f})\, q_{\alpha}(\alpha)\, q_{\beta}(\beta).$$

Variational Bayes Estimation (Cont'd)

- The factors can be obtained in an alternating fashion as,

$$\ln(q_{\psi_i}(\psi_i)) = < \ln p(\mathbf{y}, \mathbf{h}, \mathbf{f}, \alpha, \beta) >_{k \neq i} + c_i, \tag{14}$$

where ψ_i refers to the i^{th} block of $\psi = [\mathbf{h}, \mathbf{f}, \alpha, \beta]$ and $<>_{k \neq i}$ represents the expectation operator over the distributions q_{ψ_k} for all $k \neq i$. c_i is a normalizing constant.

- The log likelihood,

$$\begin{aligned} \ln p(\mathbf{y}, \mathbf{h}, \mathbf{f}, \alpha, \beta) &= \ln p(\mathbf{y}|\mathbf{h}, \mathbf{f}, \alpha, \beta) + \ln p(\mathbf{f}|\alpha) + \ln p(\mathbf{h}|\beta) \\ &= -N_y \ln \sigma^2 - \frac{1}{\sigma^2} \|\mathbf{y} - \mathcal{H}\mathbf{f}\|^2 + M \ln \alpha - \alpha \|\mathbf{f}\|^2 \\ &+ N_h \ln \beta - \beta \|\mathbf{h}\|^2 + c. \end{aligned} \tag{15}$$

Here, N_y refers to the number of elements in $\mathbf{y}$ and c is a constant.

Variational Bayes Estimation - Overall algorithm

- $\mathbf{f} \sim \mathcal{CN}(\hat{\mathbf{f}}, \mathbf{C}_{\tilde{f}\tilde{f}})$ and $\mathbf{h} \sim \mathcal{CN}(\widehat{\mathbf{h}}, \mathbf{C}_{\tilde{h}\tilde{h}})$.

Initialization: Initialize $\hat{\mathbf{f}}$ using existing calibration methods. Use this to determine $\widehat{\mathbf{h}}, <\alpha>, <\beta>$.

repeat

$$<\mathcal{H}^H\mathcal{H}>= \mathcal{H}^H(\widehat{\mathbf{h}})\mathcal{H})(\widehat{\mathbf{h}}) + <\mathcal{H}^H(\tilde{\mathbf{h}})\mathcal{H}(\tilde{\mathbf{h}})>$$

$$\hat{\mathbf{f}} = (<\mathcal{H}^H\mathcal{H}> + <\alpha> \mathbf{I})^{-1}\mathcal{H}^{\mathbf{H}}\mathbf{y}$$

$$\mathbf{C}_{\tilde{f}\tilde{f}} = (<\mathcal{H}^H\mathcal{H}> + <\alpha> \mathbf{I})^{-1}$$

$$<\mathcal{F}^H\mathcal{F}>= \mathcal{F}^H(\hat{\mathbf{f}})\mathcal{F}(\hat{\mathbf{f}}) + <\mathcal{F}^H(\tilde{\mathbf{f}})\mathcal{F}(\tilde{\mathbf{f}})>$$

$$\widehat{\mathbf{h}} = (<\mathcal{F}^H\mathcal{F}> + <\beta> \mathbf{I})^{-1}\mathcal{F}^{\mathbf{H}}\mathbf{y}$$

$$\mathbf{C}_{\tilde{h}\tilde{h}} = (<\mathcal{F}^H\mathcal{F}> + <\beta> \mathbf{I})^{-1}$$

$$<\alpha>= \frac{M}{<\|\mathbf{f}\|^2>}, \; <\|\mathbf{f}\|^2>= \hat{\mathbf{f}}^H\hat{\mathbf{f}} + tr\{\mathbf{C}_{\tilde{f}\tilde{f}}\} \; .$$

$$<\beta>= \frac{N_h}{<\|\mathbf{h}\|^2>}, \; <\|\mathbf{h}\|^2>= \widehat{\mathbf{h}}^H\widehat{\mathbf{h}} + tr\{\mathbf{C}_{\tilde{h}\tilde{h}}\}.$$

until convergence.

Variational Bayes Estimation (Cont'd)

- The log likelihood,

$$\ln p(\mathbf{y}, \mathbf{h}, \mathbf{f}, \alpha, \beta) = -N_y \ln \sigma^2 - \frac{1}{\sigma^2} \left\| \mathbf{y} - \mathcal{H}\mathbf{f} \right\|^2 + M \ln \alpha - \alpha \left\| \mathbf{f} \right\|^2 + N_h \ln \beta - \beta \left\| \mathbf{h} \right\|^2 + c. \tag{16}$$

- The parameters of the prior distributions of $\mathbf{f}$ and $\mathbf{h}$ act as regularization parameters in their estimation.
- The advantage of using VB is that it automatically computes the optimal values for these regularization parameters too!!

Expectation Consistent Variational Bayes Estimation

- For single antenna grouping $(G = M)$, $\mathbf{C}_{\tilde{f}\tilde{f}}$ and $\mathbf{C}_{\tilde{h}\tilde{h}}$ are diagonal and $<\mathcal{F}^H(\tilde{\mathbf{f}})\mathcal{F}(\tilde{\mathbf{f}})>$, $<\mathcal{H}^H(\tilde{\mathbf{h}})\mathcal{H}(\tilde{\mathbf{h}})>$ can be computed easily.
- However, when $G < M$, these matrices are block diagonal.
- To simplify the computation, we propose to approximate them with multiples of the Identity matrix,

$$\begin{aligned} \mathbf{C}_{\tilde{f}\tilde{f}} &\approx \frac{tr\{(<\mathcal{H}^H\mathcal{H}> + <\alpha>\mathbf{I})^{-\mathbf{1}}\}}{M}\mathbf{I}_M \\ \mathbf{C}_{\tilde{h}\tilde{h}} &\approx \frac{tr\{(<\mathcal{F}^H\mathcal{F}> + <\beta>\mathbf{I})^{-\mathbf{1}}\}}{N_h}\mathbf{I}_{N_h}. \end{aligned} \tag{17}$$

- We call this approach EC-VB (Expectation consistent[5] VB).

[5] Opper and Winther, "Expectation Consistent Approximate Inference".

VB and Alternating Maximum Likelihood (AML)

- By forcing the matrices $\mathbf{C}_{\tilde{f}\tilde{f}}$, $\mathbf{C}_{\tilde{h}\tilde{h}}$ to zero and α, β to zero, this algorithm reduces to the AML algorithm[6] [7].
- AML iteratively maximizes the likelihood by alternating between the desired parameters $\mathbf{f}$ and the nuisance parameters $\mathbf{h}$ in a deterministic (non-Bayesian) setting.

[6] Carvalho and Slock, "Semi–Blind Methods for FIR Multichannel Estimation".

[7] Carvalho, Omar, and Slock, "Performance and complexity analysis of blind FIR channel identification algorithms based on deterministic maximum likelihood in SIMO systems".

Simulation setup

- We assess numerically the performance of various calibration algorithms and also compare them against their CRBs.
- The Tx and Rx calibration parameters for the BS antennas are assumed to have random phases uniformly distributed over $[-\pi, \pi]$ and amplitudes uniformly distributed in the range $[1-\delta, 1+\delta]$.
- SNR is defined as the ratio of the average received signal power across channel realizations at an antenna and the noise power at that antenna.
- Transmission happens from one antenna at a time $(G = M)$.

Mean square error performance of the algorithms.

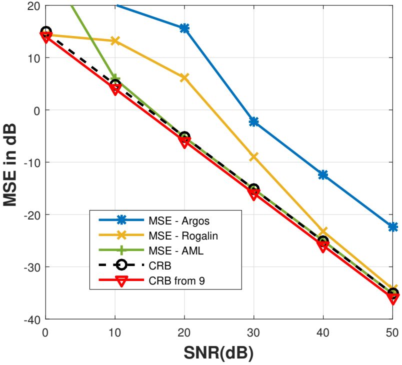


Figure: Comparison of single antenna transmit schemes with the CRB ($G = M = 16$).

- Curves generated over one realization of an i.i.d. Rayleigh channel. The first coefficient known constraint is used.
- Rogalin [8]method performs the LS as was presented earlier and improves over Argos by using all the bi-directional received data.
- AML outperforms the Rogalin performance at low SNR.

[8]Rogalin et al., "Scalable synchronization and reciprocity calibration for distributed multiuser MIMO".

CRB vs MSE.

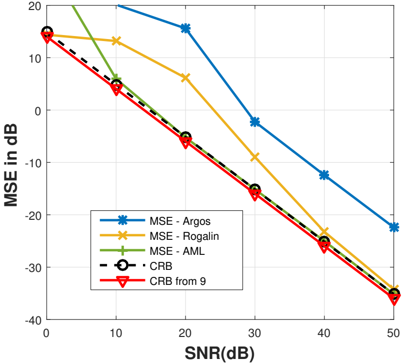


Figure: Comparison of single antenna transmit schemes with the CRB ($G = M = 16$).

- AML curve overlaps with the CRB at higher SNRs.
- Also plotted is the CRB as given in[9] assuming the internal propagation channel is fully known (the mean is known and the variance is negligible) and a (small) underestimation of the MSE can be observed as expected.

[9] Joao Vieira et al. "Reciprocity Calibration for Massive MIMO: Proposal, Modeling and Validation". In: *IEEE Trans. Wireless Commun.* (2017)

Convergence of the Optimal algorithms.

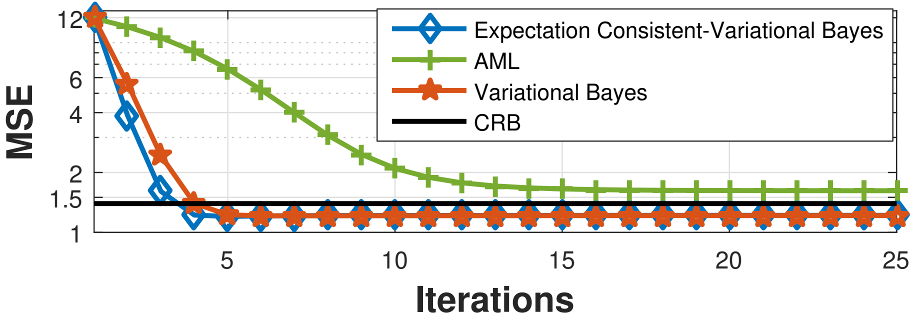


Figure: Convergence of the various iterative schemes for $M = 16$ and single antenna transmission.

Conclusion

- A simple and elegant derivation of the CRB has been presented for a general calibration framework.
- An optimal estimation algorithm based on VB is also introduced along with a variant.
- All these techniques have been compared via simulations in terms of both MSE performance and speed of convergence.

Thank You !!