Mismatched Estimation in Large Linear Systems

Yanting Ma[†], Dror Baron[†], and Ahmad Beirami[‡]

[†] North Carolina State University

[‡] Massachusetts Institute of Technology & Duke University

HONG KONG 2015

Supported by NSF & ARO

Motivation

- True prior → minimum mean square error (MMSE)

- True prior unknown in practical problems

  → use postulated prior different from true prior

  → mismatch in prior

- Mismatched estimation

  → mean square error (MSE) bigger than MMSE

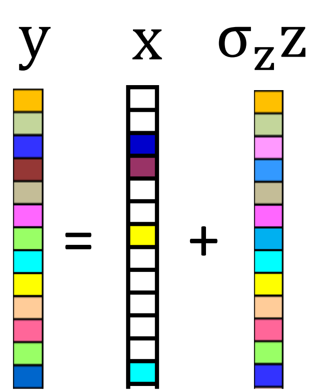
  → excess mean square error (EMSE)

Motivation

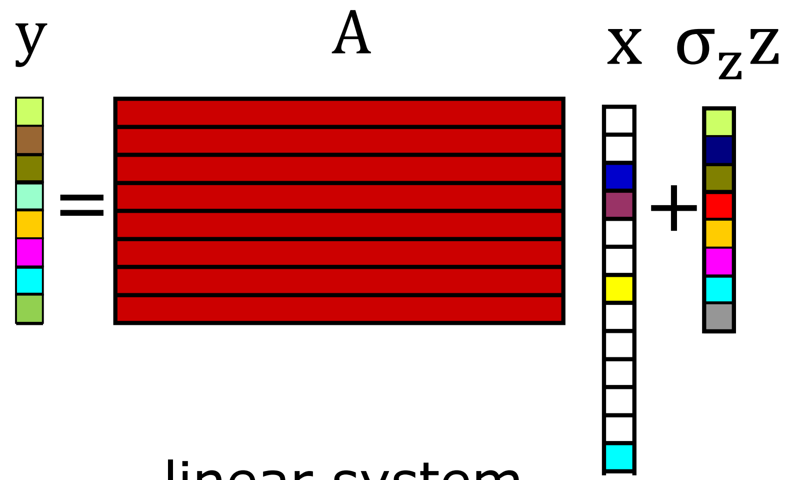
- Mismatched estimation in scalar channels [Verdú 2010]

- Many practical problems are modeled by linear systems (e.g., medical imaging, compressed sensing)

- *Extension to large linear systems*

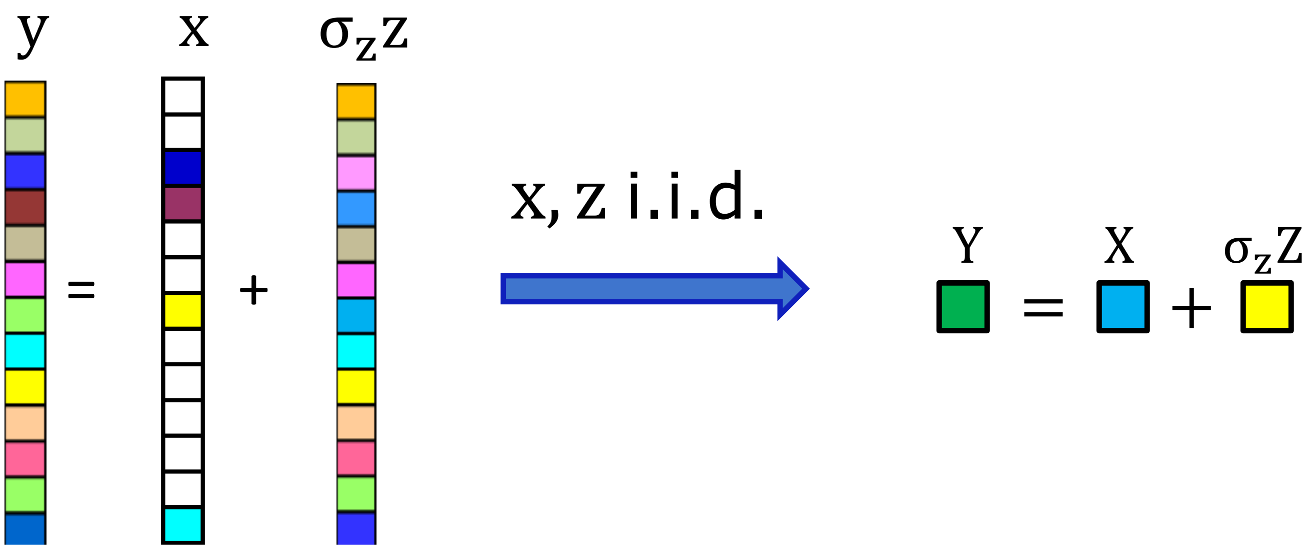


scalar channel



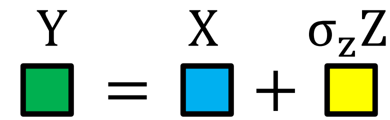
linear system

Simplification



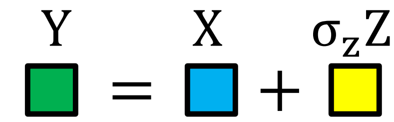
Mismatch in Scalar Channels [Verdú 2010]

- Additive white Gaussian noise channel:
  $Y = X + \sigma_z Z \in \mathbb{R};\qquad X \sim P, \quad Z \sim N(0,1)$



Mismatch in Scalar Channels [Verdú 2010]

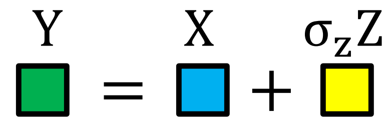
- Additive white Gaussian noise channel:
  $\mathrm{Y} = \mathrm{X} + \sigma_z \mathrm{Z} \in \mathbb{R}$; $\mathrm{X} \sim \mathrm{P}$, $\mathrm{Z} \sim \mathrm{N}(0,1)$



- *True prior* P:
  Conditional expectation: $\widehat{\mathrm{X}}_{\mathrm{P}}(\mathrm{Y}) = \mathbb{E}_{\mathrm{P}}[\mathrm{X}|\mathrm{Y}]$
  Mean square error (MSE): $\mathrm{MSE}_{\mathrm{P}}(\sigma_{\mathrm{z}}^2) = \mathbb{E}\left[\left(\widehat{\mathrm{X}}_{\mathrm{P}}(\mathrm{Y}) - \mathrm{X}\right)^2\right]$
  → *minimum* MSE (MMSE): $\mathrm{MMSE}(\sigma_{\mathrm{z}}^2) = \mathrm{MSE}_{\mathrm{P}}(\sigma_{\mathrm{z}}^2)$

Mismatch in Scalar Channels [Verdú 2010]

- Additive white Gaussian noise channel:
  $\mathrm{Y} = \mathrm{X} + \sigma_z \mathrm{Z} \in \mathbb{R};\qquad \mathrm{X} \sim \mathrm{P},\quad \mathrm{Z} \sim \mathrm{N}(0,1)$



- *True prior* P:
  Conditional expectation: $\widehat{\mathrm{X}}_{\mathrm{P}}(\mathrm{Y}) = \mathbb{E}_{\mathrm{P}}[\mathrm{X}|\mathrm{Y}]$
  Mean square error (MSE): $\mathrm{MSE}_{\mathrm{P}}(\sigma_{\mathrm{z}}^2) = \mathbb{E}\left[\left(\widehat{\mathrm{X}}_{\mathrm{P}}(\mathrm{Y}) - \mathrm{X}\right)^2\right]$
  → *minimum* MSE (MMSE): $\mathrm{MMSE}(\sigma_{\mathrm{z}}^2) = \mathrm{MSE}_{\mathrm{P}}(\sigma_{\mathrm{z}}^2)$

- *Mismatched prior* Q:
  MSE: $\mathrm{MSE}_{\mathrm{Q}}(\sigma_{\mathrm{z}}^2) = \mathbb{E}\left[\left(\widehat{\mathrm{X}}_{\mathrm{Q}}(\mathrm{Y}) - \mathrm{X}\right)^2\right]$
  Bigger than MMSE

Mismatch in Scalar Channels [Verdú 2010]

- Excess MSE of scalar channel (S-EMSE):

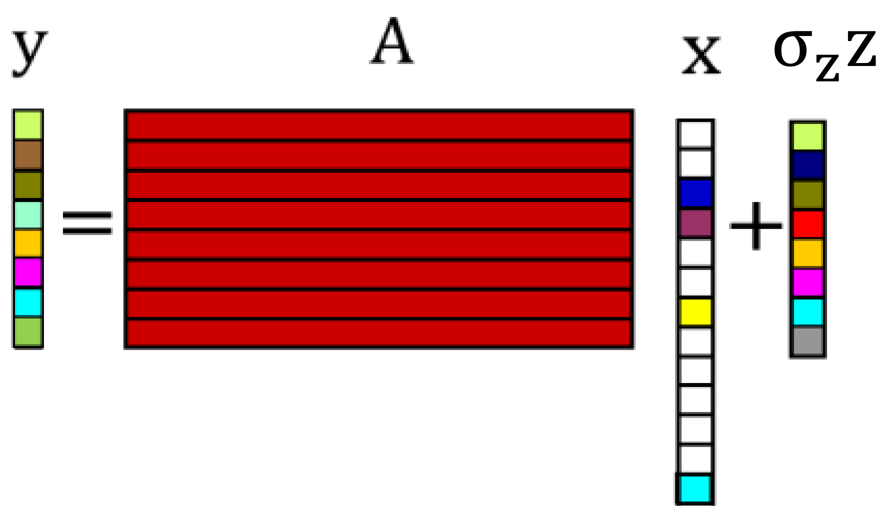
  $\text{S-EMSE}(\sigma_z^2) = \mathrm{MSE_Q}(\sigma_z^2) - \mathrm{MMSE}(\sigma_z^2)$

- $\mathrm{D}(\mathrm{P}||\mathrm{Q}) = \frac{1}{2}\int_0^\infty \text{S-EMSE}(1/\gamma)\, d\gamma, \quad \gamma = 1/\sigma_z^2$

  relative entropy

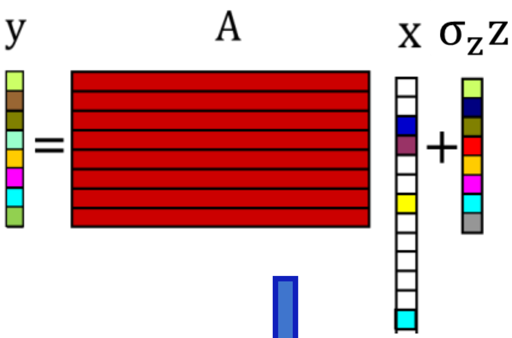
Linear System

- Linear system: $y = Ax + \sigma_z z \in \mathbb{R}^M$
- Input signal: $x \in \mathbb{R}^N$, $x_j \sim P$
- Noise: $z_i \sim N(0,1)$
- Measurement matrix: $A \in \mathbb{R}^{M \times N}, \mathbb{E}[A_{ij}] = 0, \mathrm{Var}[A_{ij}] = \frac{1}{M}$



Decoupling [Tanaka 2002, Guo & Verdú 2005,...]

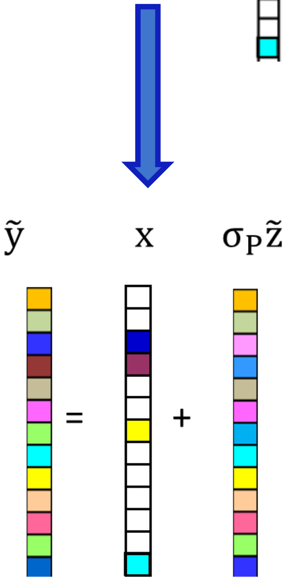
Linear system: $y = Ax + \sigma_z z$



decoupling

$N \to \infty, \frac{M}{N} \to \delta$

large system limit



Scalar channel: $\tilde{y}_i = x_i + \sigma_P \tilde{z}_i$

$x_i \sim P, \quad \tilde{z}_i \sim N(0,1)$

Fixed point equation: $\delta \cdot \left(\sigma_P^2 - \sigma_z^2\right) = \mathrm{MSE}_P\left(\sigma_P^2\right)$

Problem Statement

MMSE in Large Linear Systems

- Focusing on large linear systems

- *True prior* P:

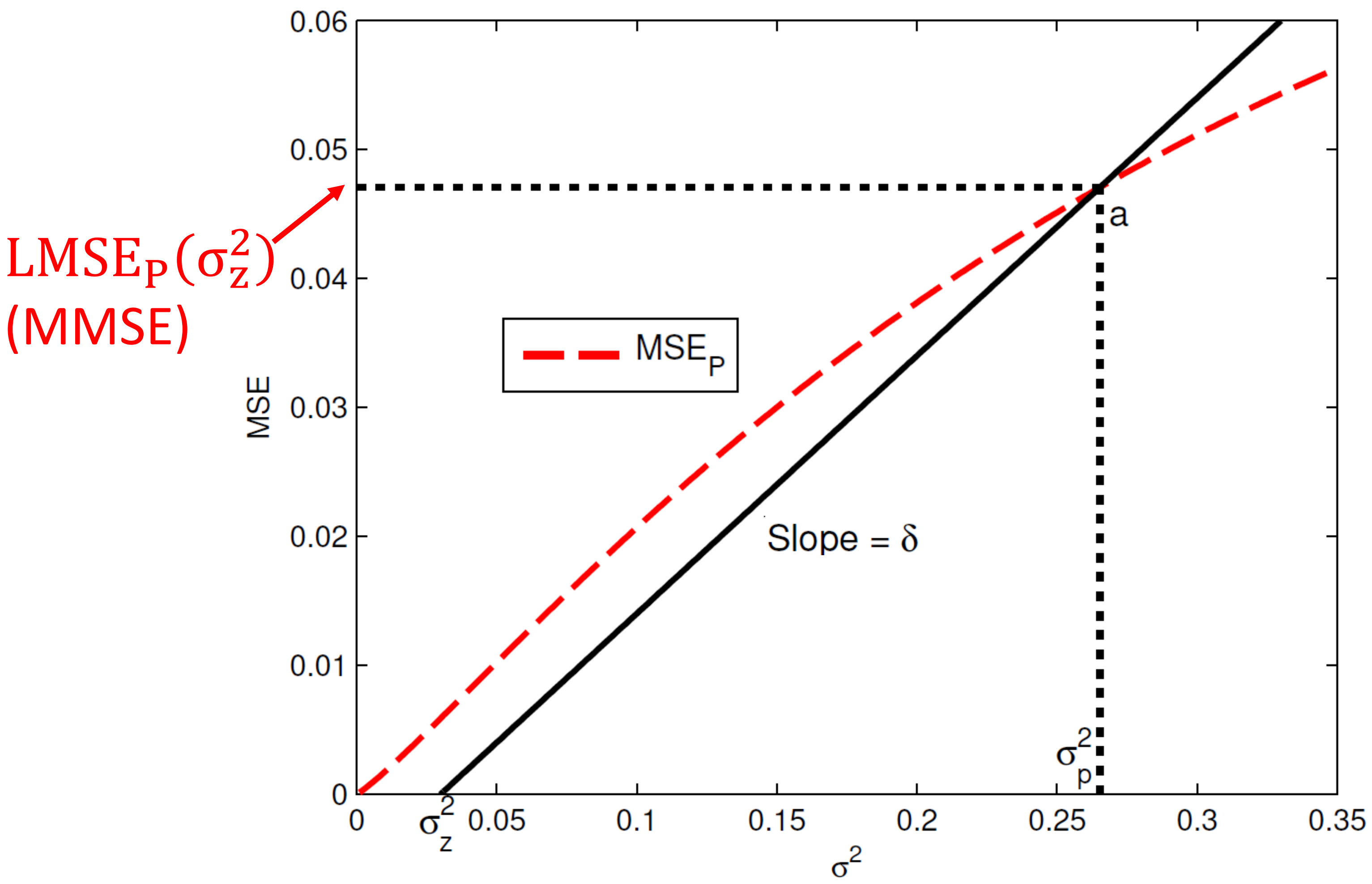
  Conditional expectation: $\hat{\mathrm{x}}_{\mathrm{P}}(\mathrm{y},\mathrm{A}) = \mathbb{E}_{\mathrm{P}}[\mathrm{x}|\mathrm{y},\mathrm{A}]$

  MSE: $\mathrm{LMSE}_{\mathrm{P}}(\sigma_z^2) = \lim_{N\to\infty} \frac{1}{N}\mathbb{E}[||\hat{\mathrm{x}}_{\mathrm{P}}(\mathrm{y},\mathrm{A}) - \mathrm{x}||_2^2] = \mathrm{MMSE}(\sigma_z^2)$

  linear system

LMSE via Decoupling

[Bernoulli-Gaussian input]



Fixed point equation: $\delta \cdot (\sigma^2 - \sigma_z^2) = \mathrm{MSE_P}(\sigma^2)$ with solution σ_P^2

Mismatch in Large Linear Systems

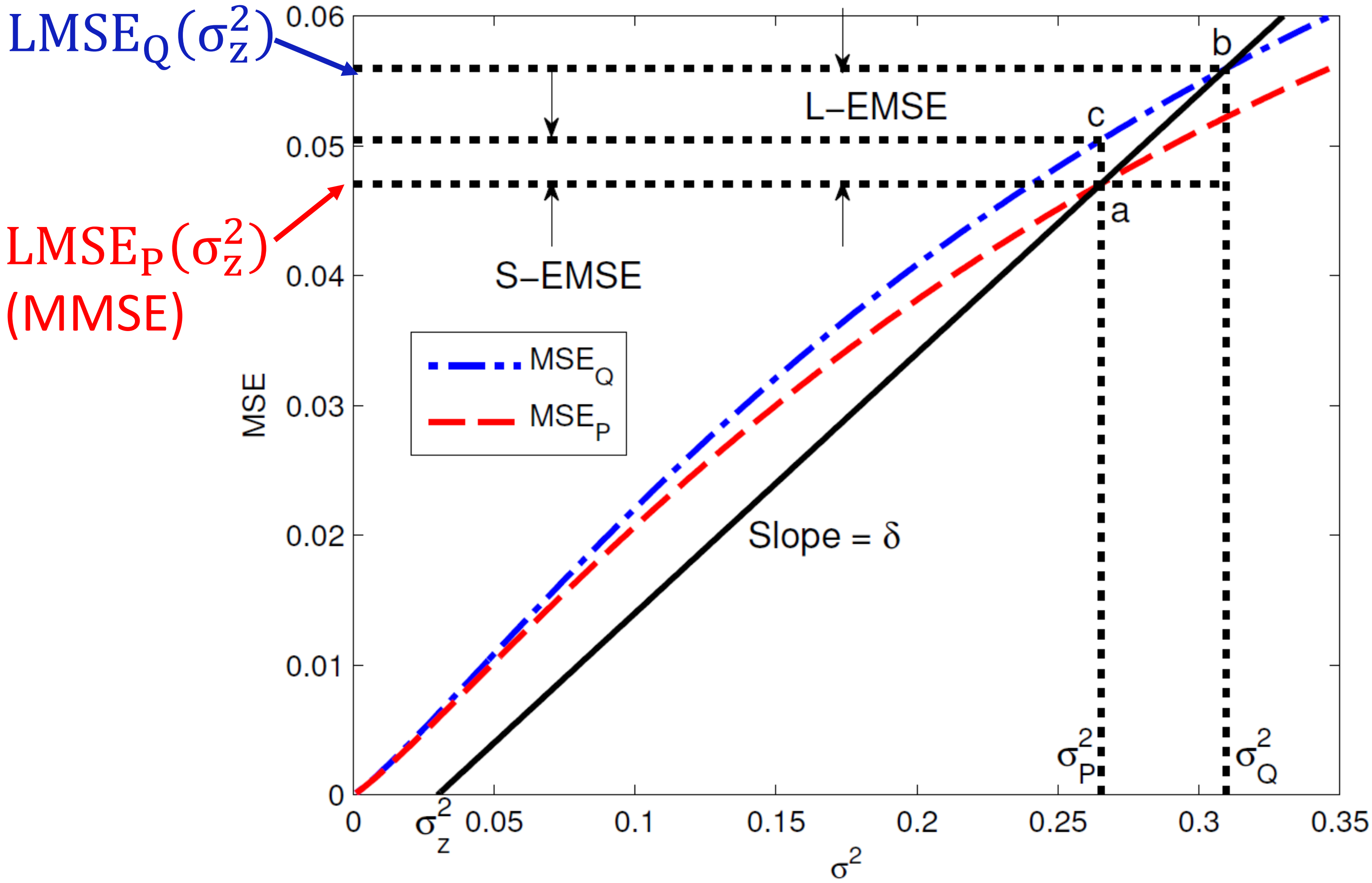
- *Mismatched prior* Q:

MSE: $\mathrm{LMSE_Q}(\sigma_z^2) = \lim_{\mathrm{N}\to\infty} \frac{1}{\mathrm{N}} \mathbb{E}\left[||\hat{\mathrm{x}}_\mathrm{Q}(\mathrm{y},\mathrm{A}) - \mathrm{x}||_2^2\right]$

- Excess MSE of linear system:

$\text{L-EMSE}(\sigma_\mathrm{z}^2) = \mathrm{LMSE_Q}(\sigma_\mathrm{z}^2) - \mathrm{MMSE}(\sigma_\mathrm{z}^2)$

Mismatch in Large Linear Systems

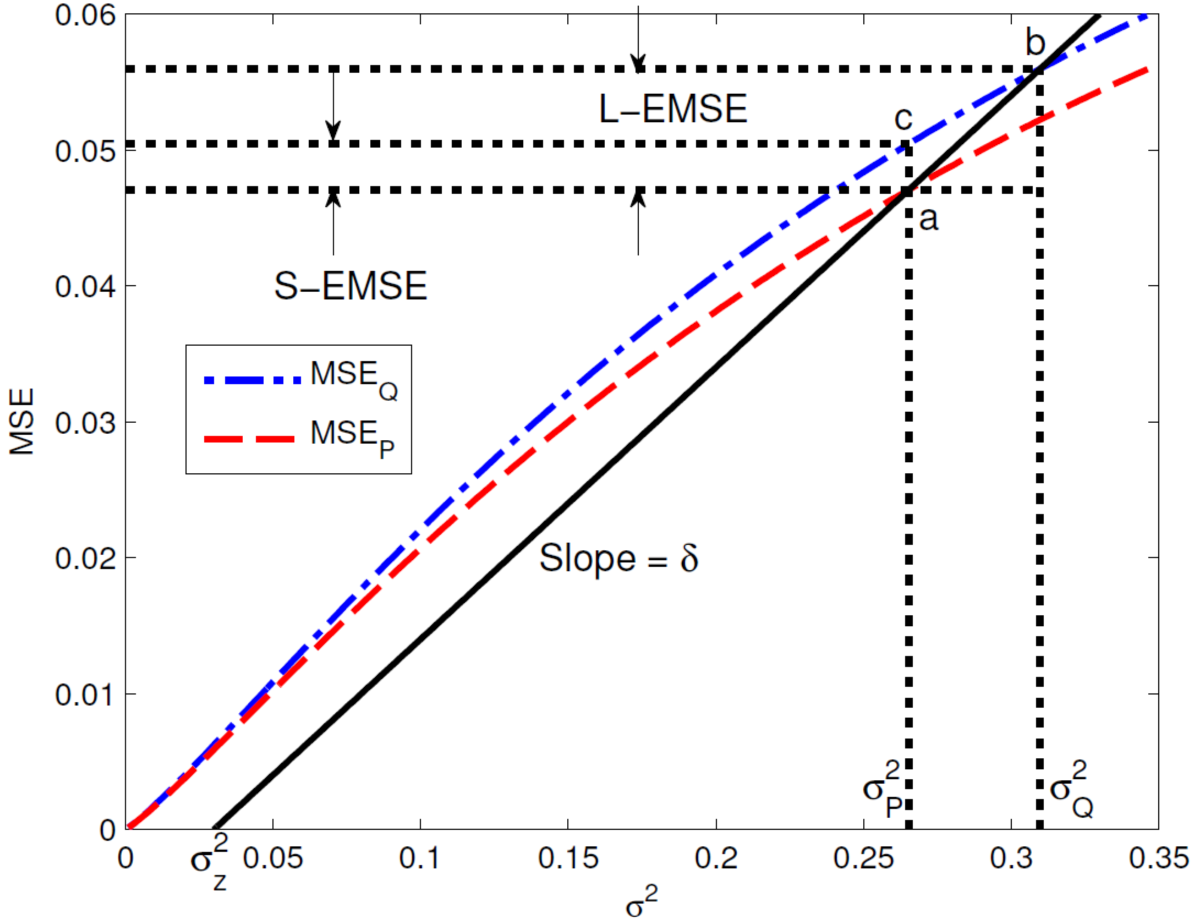


Goal: relate L-EMSE(σ_z^2) to S-EMSE(σ_P^2)

Main Results

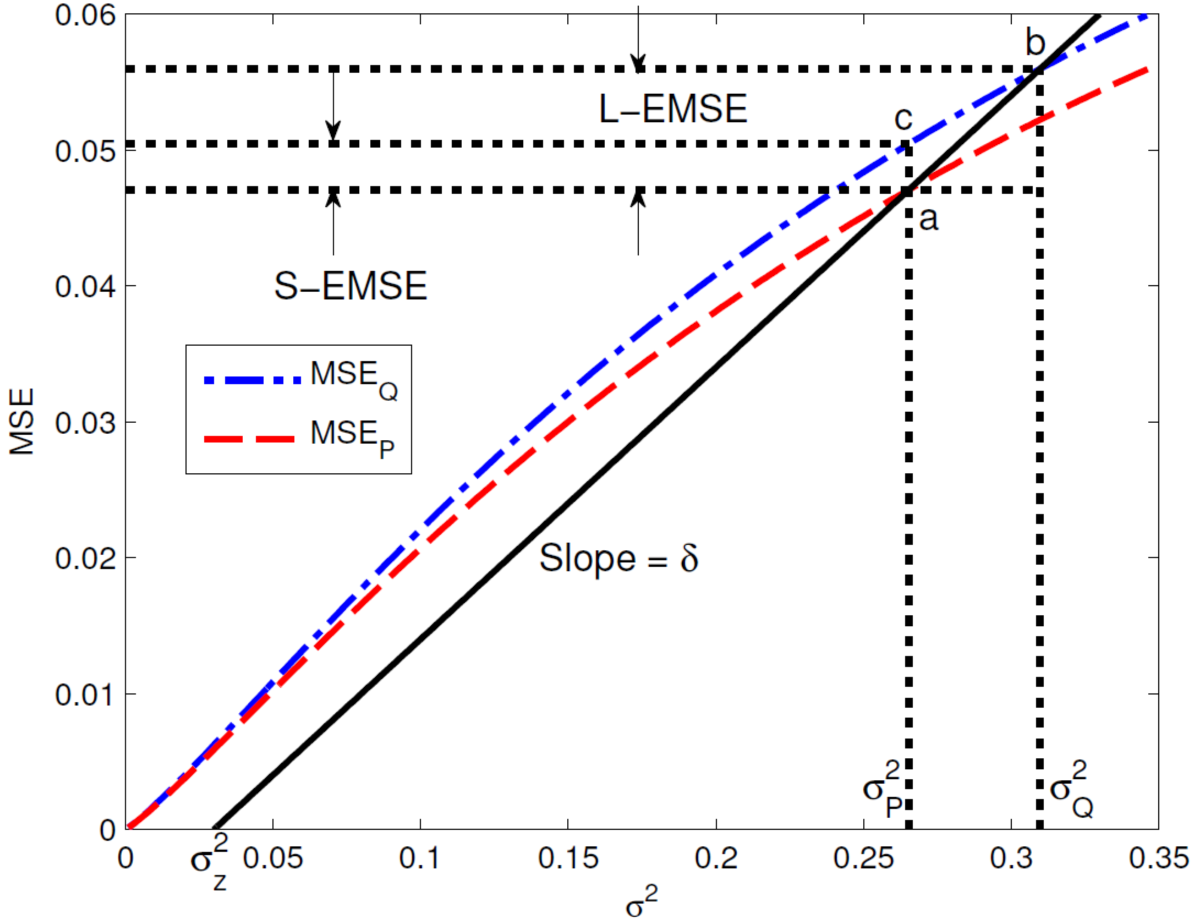
Main Result

(derivation in paper)



$$\text{L-EMSE}(\sigma_Z^2) = \text{S-EMSE}(\sigma_P^2) + \int_{\sigma_P^2}^{\sigma_P^2+\frac{1}{\delta}\text{L-EMSE}(\sigma_Z^2)} \frac{d}{d\sigma^2}\text{MSE}_Q(\sigma^2)d\sigma^2$$

Main Result
(derivation in paper)



Taylor expansion

$$\text{L-EMSE}(\sigma_Z^2) = \text{S-EMSE}(\sigma_P^2) + \int_{\sigma_P^2}^{\sigma_P^2+\frac{1}{\delta}\text{L-EMSE}(\sigma_Z^2)} \frac{d}{d\sigma^2}\text{MSE}_Q(\sigma^2)\, d\sigma^2$$

Taylor Approximations

- First order approximation:

$$\text{L-EMSE}(\sigma_z^2) = \frac{\delta}{\delta - \alpha}\text{S-EMSE}\left(\sigma_p^2\right) + o(\Delta)$$

- Second order approximation:

$$\text{L-EMSE}(\sigma_z^2) = \frac{\delta}{\delta - \alpha}\text{S-EMSE}\left(\sigma_p^2\right)\left[1 + \frac{1}{2}\frac{\beta}{(\delta - \alpha)^2}\text{S-EMSE}\left(\sigma_p^2\right)\right] + o(\Delta^2)$$

$\Delta = \sigma_Q^2 - \sigma_P^2$

α: first derivative of $\mathrm{MSE}_Q(\sigma^2)$ at σ_P^2

β: second derivative of $\mathrm{MSE}_Q(\sigma^2)$ at σ_P^2

Taylor Approximations

- First order approximation:

$$\text{L-EMSE}(\sigma_z^2) = \frac{\delta}{\delta-\alpha}\text{S-EMSE}\left(\sigma_p^2\right) + o(\Delta)$$

- Second order approximation:

$$\text{L–EMSE}(\sigma_z^2) = \frac{\delta}{\delta-\alpha}\text{S–EMSE}\left(\sigma_p^2\right)\left[1 + \frac{1}{2}\frac{\beta}{(\delta-\alpha)^2}\text{S–EMSE}\left(\sigma_p^2\right)\right] + o(\Delta^2)$$

$\Delta = \sigma_Q^2 - \sigma_P^2$

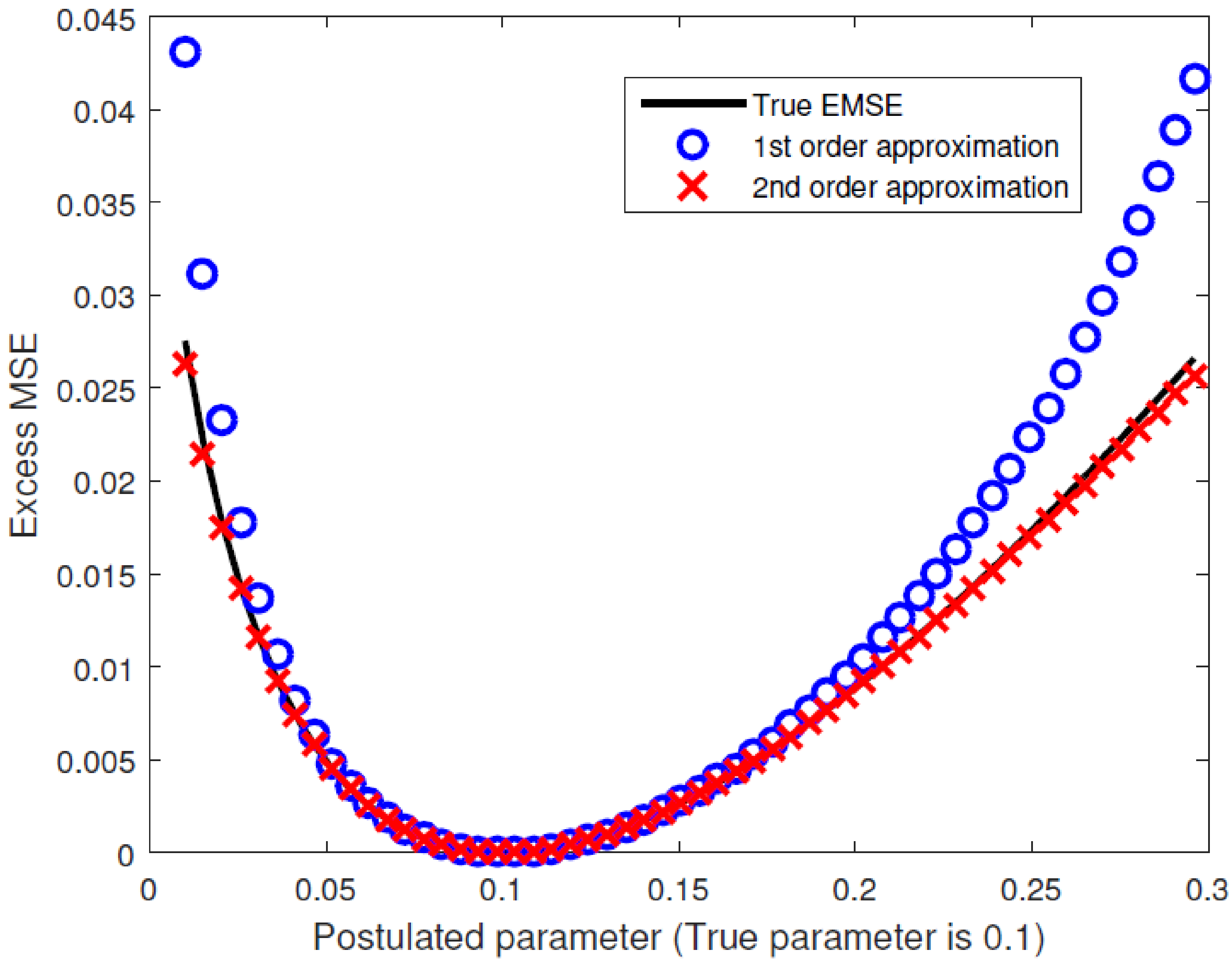
α: first derivative of $\text{MSE}_Q(\sigma^2)$ at σ_P^2

β: second derivative of $\text{MSE}_Q(\sigma^2)$ at σ_P^2

Numerical Examples

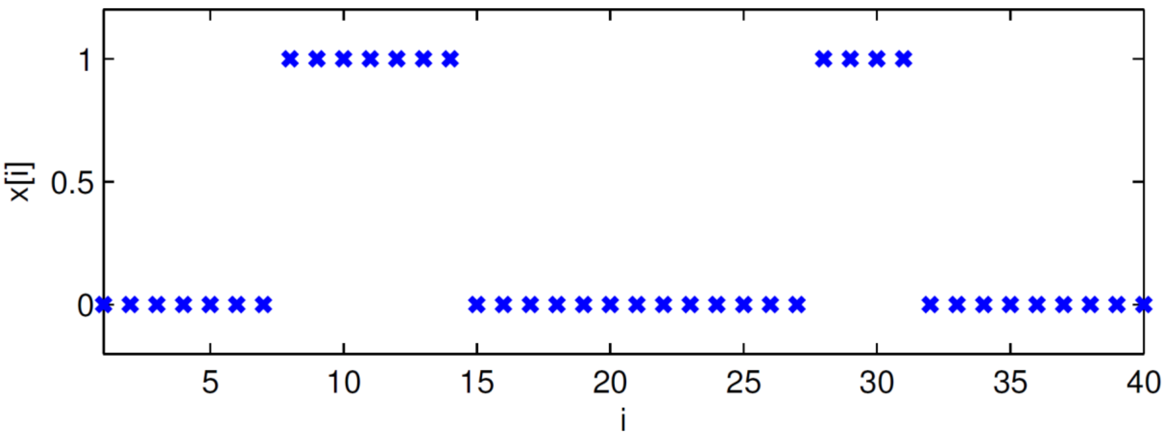
Example 1: Bernoulli-Gaussian input

- $f_X(x) = \theta \cdot N(0,1) + (1-\theta) \cdot \delta_0(x)$



Example 2: Markov Source

- Two state Markov source with state space {0,1}

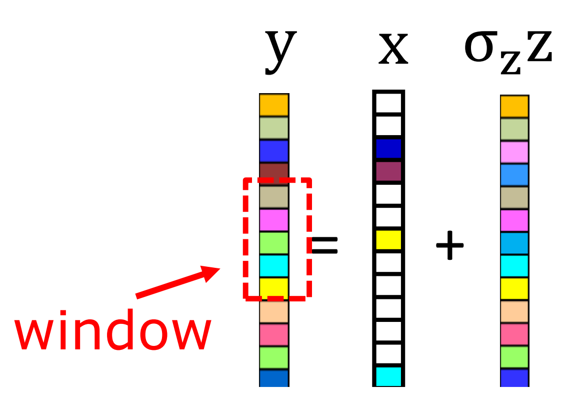


- For *non-i.i.d*. input (e.g., Markov source) decoupling principal not well-understood

- Alternative tool: state evolution (SE) of approximate message passing (AMP) [Bayati & Montanari 2011]

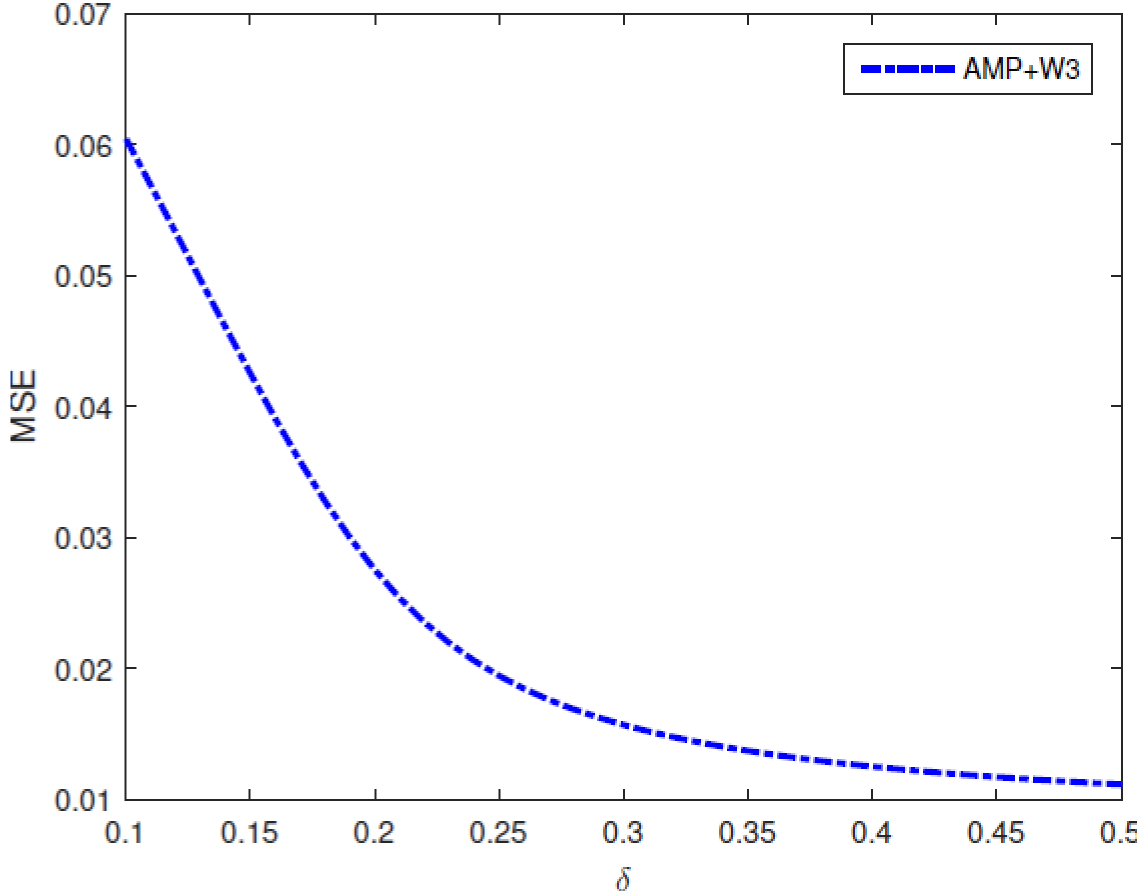
Approximate Message Passing [Donoho *et al.* 2009]

- AMP solves linear system by iterative decoupling
- Decoupled scalar channel characterized by state evolution (SE)
- SE for AMP with $\mathbb{E}\left[x_i | \tilde{y}_i^t\right]$ coincides with fixed point equation for information theoretic decoupling
- For *non-i.i.d.* input $\mathbb{E}[x_i | \text{window}_i]$ better than $\mathbb{E}\left[x_i | \tilde{y}_i^t\right]$
- SE for AMP with $\mathbb{E}[x_i | \text{window}_i]$ not rigorous but numerically verified [Ma et al. 2014]



Example 2: Markov Source

- Two state Markov source with state space {0,1}
- W2(3): $\mathbb{E}[x_i|\text{window}_i]$ with window-size 2(3)

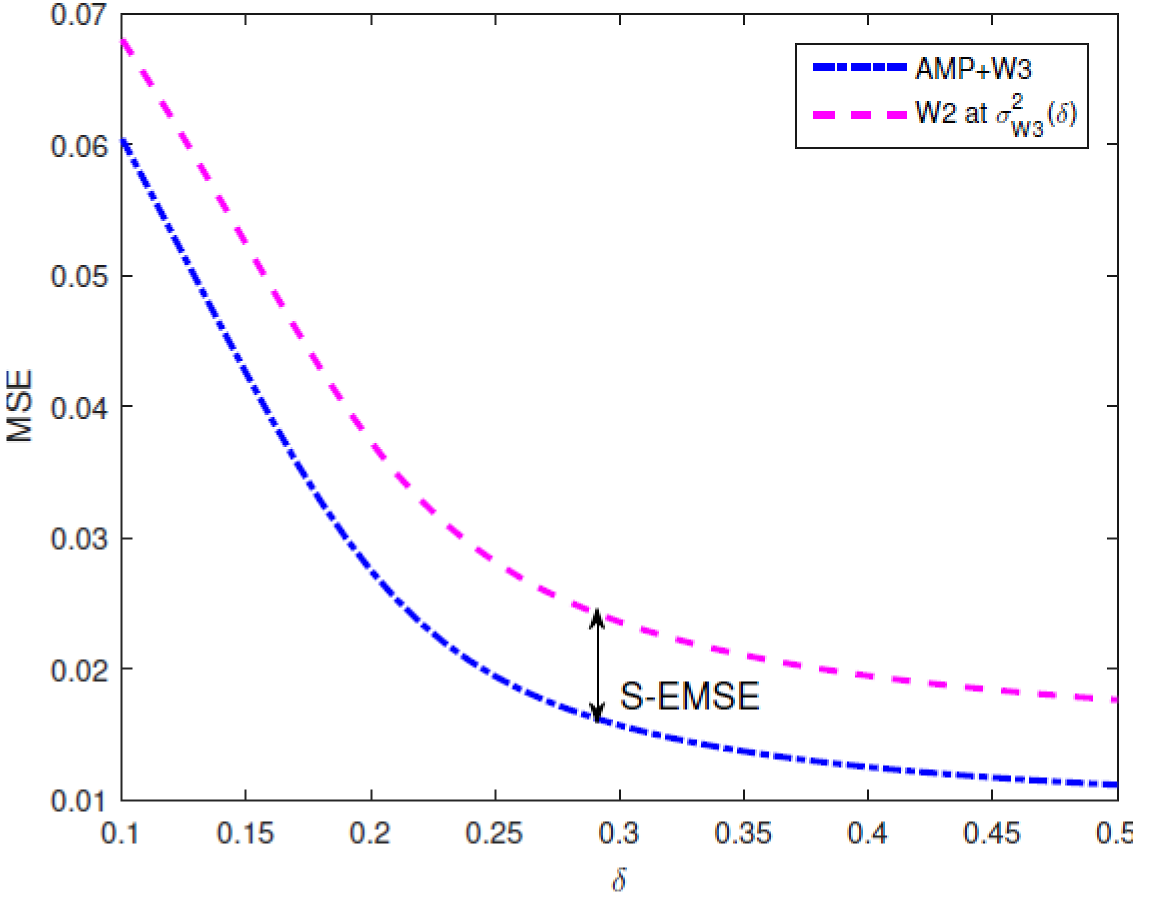


$$\delta \cdot \left(\sigma_{W3}^2 - \sigma_z^2\right) = \text{MSE}_{W3}\left(\sigma_{W3}^2\right)$$

$$Y = X + \sigma_{W3} Z$$

Example 2: Markov Source

- Two state Markov source with state space {0,1}
- W2(3): $\mathbb{E}[\mathrm{x_i}|\mathrm{window_i}]$ with window-size 2(3)

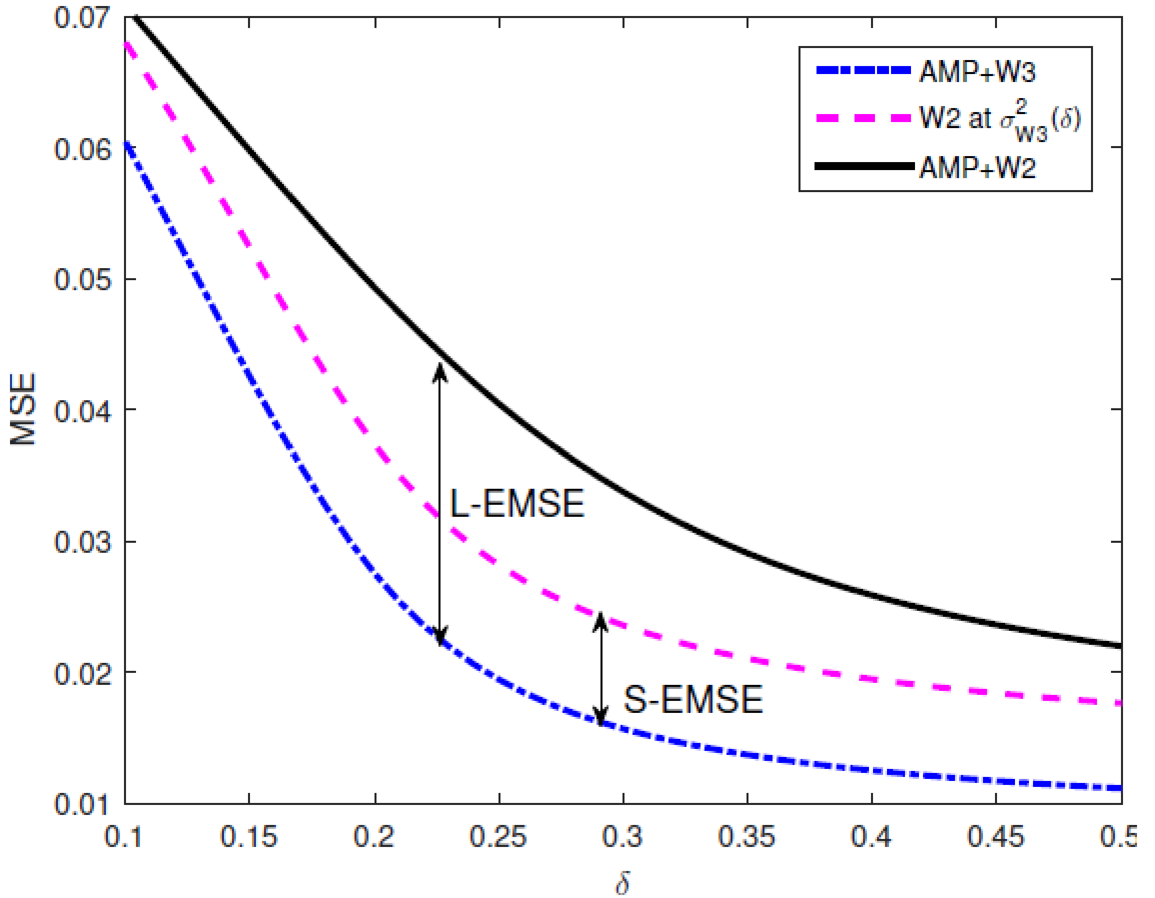


$$\delta \cdot \left(\sigma_{W3}^2 - \sigma_z^2\right) = \mathrm{MSE_{W3}}\left(\sigma_{W3}^2\right)$$

Apply W2 to → $\mathrm{Y} = \mathrm{X} + \sigma_{\mathrm{W3}}\mathrm{Z}$

Example 2: Markov Source

- Two state Markov source with state space {0,1}
- W2(3): $\mathbb{E}[x_i|\text{window}_i]$ with window-size 2(3)

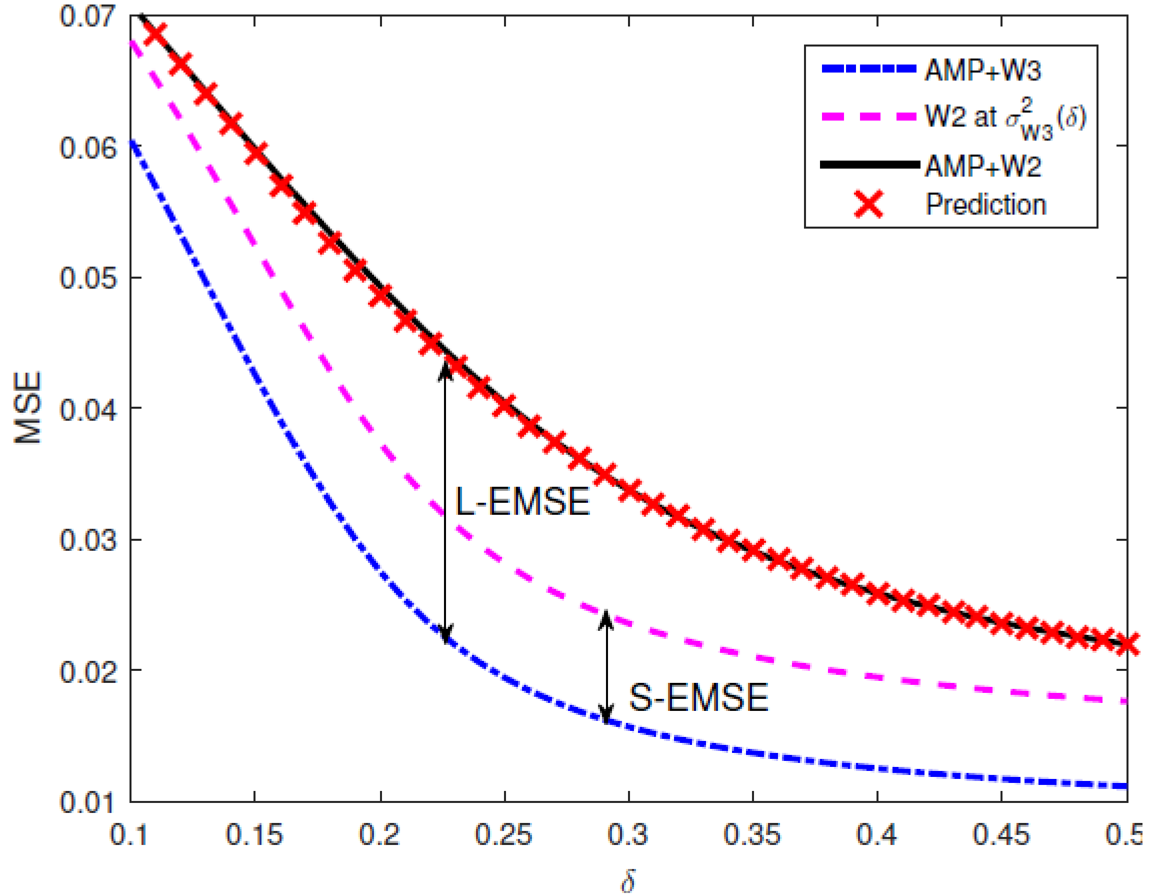


$$\delta \cdot \left(\sigma^2_{W3} - \sigma^2_z\right) = \mathrm{MSE}_{W3}\left(\sigma^2_{W3}\right)$$

$$Y = X + \sigma_{W3} Z$$

Example 2: Markov Source

- Two state Markov source with state space {0,1}
- W2(3): $\mathbb{E}[x_i|\text{window}_i]$ with window-size 2(3)



$$\delta \cdot \left(\sigma_{W3}^2 - \sigma_z^2\right) = \mathrm{MSE}_{W3}\left(\sigma_{W3}^2\right)$$

$$Y = X + \sigma_{W3} Z$$

Discussion

Summary

- Provided expression relating L-EMSE to S-EMSE
- Two Taylor approximations derived and accuracy verified numerically

Next step

- How does parameter estimation error affect EMSE in linear systems?

Thank you!