ulm university universität ulm

Variants of VAMP: Variable Separation and Individual Variances

Carmen Sippel   Robert F.H. Fischer

NACHRICHTENTECHNIK
Universität Ulm

Project:
Iterative Signal Recovery Algorithms
—A Unified View of Turbo and Message-Passing Approaches—

Project Details

Objectives:

- Unified exposition of turbo processing and message-passing philosophies
- Examination of similarities/differences, advantages/disadvantages
- Engineering perspective: finite-length performance rather than asymptotics
  - Performance
  - Complexity

Project Start: February, 2019

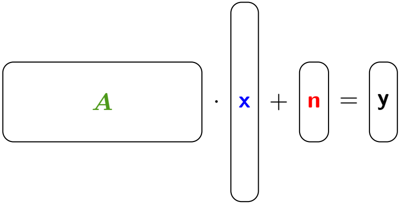
So Far:

- Design of algorithms based on Expectation-Consistent (EC) approximate inference
  - Variable separation of the problem
  - Use of individual variances

System Model for Compressed Sensing [CRT'06, Don'06]

Scenario:

- Noisy compressed measurements $\mathbf{y}$ given



Task:

Estimation of signal $\mathbf{x}$

Assumptions:

- Sensing matrix $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ known
- Noise $\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \boldsymbol{I}_M)$

Random variables in sans-serif font, e.g., $\mathbf{x}$, $\mathbf{n}$, realizations in serif font, e.g., $\boldsymbol{A}$.

Scenario:

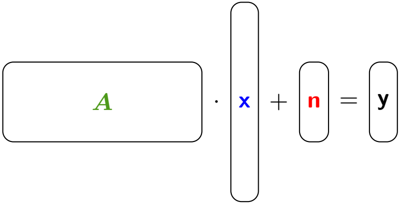
- Noisy compressed measurements **y** given
- Pdf of signal **x** known

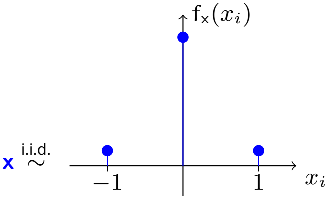
Task:

Estimation of signal **x**

Assumptions:

- Sensing matrix $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ known
- Noise **n** $\sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \boldsymbol{I}_M)$
- Pdf of **x** separable





Random variables in sans-serif font, e.g., **x**, **n**, realizations in serif font, e.g., $\boldsymbol{A}$.

Scenario:

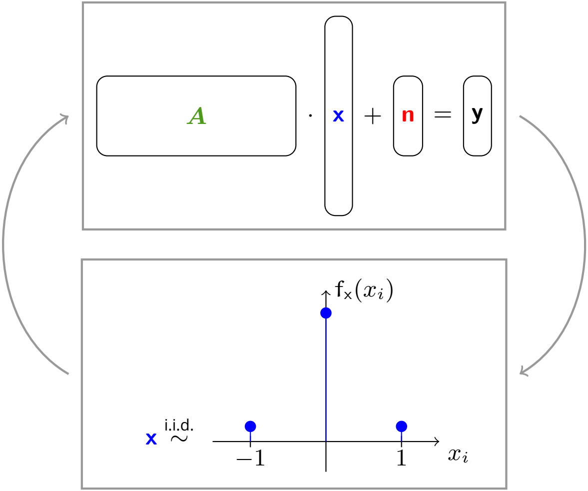
- Noisy compressed measurements **y** given
- Pdf of signal **x** known

Task:

Estimation of signal **x**

Assumptions:

- Sensing matrix $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ known
- Noise **n** $\sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \boldsymbol{I}_M)$
- Pdf of **x** separable



Random variables in sans-serif font, e.g., **x**, **n**, realizations in serif font, e.g., $\boldsymbol{A}$.

Task:

Estimation, i.e., calculation of

- First order moments (means): $\mathrm{E}\{\mathsf{x}_i \mid \boldsymbol{y}\}$
- Second order moments (reliabilty, here averaged): $\frac{1}{N}\sum_{i=1}^{N}\mathrm{E}\{\mathsf{x}_i^2 \mid \boldsymbol{y}\}$

Compact Representation:

Defining (as in VAMP [RSF'19])

$$\boldsymbol{g}(\boldsymbol{x}) = [x_1, \ldots, x_N, -\frac{1}{2}\sum_{i=1}^{N} x_i^2]^\top$$ [FSR+16]

yields a compact formula

$$\mathrm{E}_{\mathsf{x}}\{\boldsymbol{g}(\mathbf{x}) \mid \boldsymbol{y}\} = \int \boldsymbol{g}(\boldsymbol{x})\, \mathrm{f}_{\mathsf{x}}(\boldsymbol{x} \mid \boldsymbol{y})\, \mathrm{d}\boldsymbol{x}$$

Idea:

Partition of infeasible problem into subproblems:

$$\mathrm{E}_{\mathsf{x}}\{\boldsymbol{g}(\mathbf{x}) \mid \boldsymbol{y}\} = \int \boldsymbol{g}(\boldsymbol{x}) \, \tfrac{1}{Z} f_{\mathrm{A}}(\boldsymbol{x} \mid \boldsymbol{y}) \cdot f_{\mathrm{B}}(\boldsymbol{x} \mid \boldsymbol{y}) \, \mathrm{d}\boldsymbol{x}$$

$$\mathrm{f}_{\mathsf{x}}(\mathbf{x} \mid \boldsymbol{y}) = \tfrac{1}{Z} \; f_{\mathrm{A}}(\boldsymbol{x} \mid \boldsymbol{y}) \quad \cdot \quad f_{\mathrm{B}}(\boldsymbol{x} \mid \boldsymbol{y})$$

Idea:

Partition of infeasible problem into subproblems:

$$\mathrm{E}_{\mathsf{x},\mathrm{A}}\{\boldsymbol{g}(\mathbf{x}) \mid \boldsymbol{y}\} = \int \boldsymbol{g}(\boldsymbol{x})\, \frac{1}{Z_\mathrm{A}} f_\mathrm{A}(\boldsymbol{x} \mid \boldsymbol{y}) \cdot \exp(\boldsymbol{\theta}_\mathrm{A}^\top \boldsymbol{g}(\boldsymbol{x}))\; \mathrm{d}\boldsymbol{x}$$

$$\mathrm{f}_{\mathsf{x}}(\mathbf{x} \mid \boldsymbol{y}) = \frac{1}{Z}\, f_\mathrm{A}(\boldsymbol{x} \mid \boldsymbol{y}) \cdot f_\mathrm{B}(\boldsymbol{x} \mid \boldsymbol{y})$$

$$\mathrm{E}_{\mathsf{x},\mathrm{B}}\{\boldsymbol{g}(\mathbf{x}) \mid \boldsymbol{y}\} = \int \boldsymbol{g}(\boldsymbol{x})\, \frac{1}{Z_\mathrm{B}} \exp(\boldsymbol{\theta}_\mathrm{B}^\top \boldsymbol{g}(\boldsymbol{x})) \cdot f_\mathrm{B}(\boldsymbol{x} \mid \boldsymbol{y})\; \mathrm{d}\boldsymbol{x}$$

Replacement of "other" part by member of exponential family [Bro'86, MV'18]

Usual Separation for CS:

$$\mathrm{E}_{\mathsf{x},\mathrm{A}}\{\boldsymbol{g}(\mathbf{x}) \mid \boldsymbol{y}\} = \int \boldsymbol{g}(\boldsymbol{x}) \, \frac{1}{Z_\mathrm{A}} \exp\!\left(-\frac{1}{2\sigma_n^2}\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} + \frac{1}{\sigma_n^2}\boldsymbol{A}^\top \boldsymbol{y}\right) \exp(\boldsymbol{\theta}_\mathrm{A}^\top \boldsymbol{g}(\boldsymbol{x})) \, \mathrm{d}\boldsymbol{x}$$

$$\mathrm{f}_{\mathsf{x}}(\mathbf{x} \mid \boldsymbol{y}) = \frac{1}{Z} \, f_\mathrm{A}(\boldsymbol{x} \mid \boldsymbol{y}) \cdot f_\mathrm{B}(\boldsymbol{x} \mid \boldsymbol{y})$$

$$\mathrm{E}_{\mathsf{x},\mathrm{B}}\{\boldsymbol{g}(\mathbf{x}) \mid \boldsymbol{y}\} = \int \boldsymbol{g}(\boldsymbol{x}) \, \frac{1}{Z_\mathrm{B}} \exp(\boldsymbol{\theta}_\mathrm{B}^\top \boldsymbol{g}(\boldsymbol{x})) \prod_{i=1}^{N} \mathrm{f}_{\mathsf{x}}(x_i) \, \mathrm{d}\boldsymbol{x}$$

Usual Separation for CS:

joint, linear

$$\mathrm{E}_{\mathsf{x},\mathrm{A}}\{\boldsymbol{g}(\mathbf{x}) \mid \boldsymbol{y}\} = \int \boldsymbol{g}(\boldsymbol{x}) \, \frac{1}{Z_\mathrm{A}} \exp(-\tfrac{1}{2\sigma_n^2}\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} + \tfrac{1}{\sigma_n^2}\boldsymbol{A}^\top \boldsymbol{y}) \exp(\boldsymbol{\theta}_\mathrm{A}^\top \boldsymbol{g}(\boldsymbol{x})) \, \mathrm{d}\boldsymbol{x}$$

$$\mathsf{f}_\mathsf{x}(\mathbf{x} \mid \boldsymbol{y}) = \tfrac{1}{Z} \exp(-\tfrac{1}{2\sigma_n^2}\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} + \tfrac{1}{\sigma_n^2}\boldsymbol{A}^\top \boldsymbol{y}) \cdot \prod_{i=1}^N \mathsf{f}_\mathsf{x}(x_i)$$

$$\mathrm{E}_{\mathsf{x},\mathrm{B}}\{\boldsymbol{g}(\mathbf{x}) \mid \boldsymbol{y}\} = \int \boldsymbol{g}(\boldsymbol{x}) \, \frac{1}{Z_\mathrm{B}} \exp(\boldsymbol{\theta}_\mathrm{B}^\top \boldsymbol{g}(\boldsymbol{x})) \prod_{i=1}^N \mathsf{f}_\mathsf{x}(x_i) \, \mathrm{d}\boldsymbol{x}$$

separable, non-linear

NACHRICHTENTECHNIK Universität Ulm

Usual Separation for CS:

$$\mathrm{E}_{\mathsf{x},\mathrm{A}}\{\boldsymbol{g}(\mathbf{x}) \mid \boldsymbol{y}\} = \int \boldsymbol{g}(\boldsymbol{x}) \, \frac{1}{Z_\mathrm{A}} \exp(-\tfrac{1}{2\sigma_n^2}\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x} + \tfrac{1}{\sigma_n^2}\boldsymbol{A}^\top\boldsymbol{y}) \exp(\boldsymbol{\theta}_\mathrm{A}^\top\boldsymbol{g}(\boldsymbol{x})) \, \mathrm{d}\boldsymbol{x}$$

↑

$$\mathsf{f}_\mathsf{x}(\mathbf{x} \mid \boldsymbol{y}) = \tfrac{1}{Z} \exp(-\tfrac{1}{2\sigma_n^2}\boldsymbol{x}^\top\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{x} + \tfrac{1}{\sigma_n^2}\boldsymbol{A}^\top\boldsymbol{y}) \cdot \prod_{i=1}^{N} \mathsf{f}_\mathsf{x}(x_i)$$

↓

$$\mathrm{E}_{\mathsf{x},\mathrm{B}}\{\boldsymbol{g}(\mathbf{x}) \mid \boldsymbol{y}\} = \int \boldsymbol{g}(\boldsymbol{x}) \, \frac{1}{Z_\mathrm{B}} \exp(\boldsymbol{\theta}_\mathrm{B}^\top\boldsymbol{g}(\boldsymbol{x})) \, \prod_{i=1}^{N} \mathsf{f}_\mathsf{x}(x_i) \, \mathrm{d}\boldsymbol{x}$$

Natural Parameters:

[Bro'86]

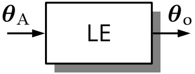
$$\boldsymbol{g}(\boldsymbol{x}) = [x_1, \ldots, x_N, -\tfrac{1}{2}\textstyle\sum_{i=1}^{N} x_i^2]^\top$$

$$\boldsymbol{\theta} = [\lambda_1, \ldots, \lambda_N, \quad \Lambda \quad]^\top = [\boldsymbol{\lambda}^\top, \Lambda]^\top, \Lambda = 1/\sigma_\mathrm{avg}^2, \boldsymbol{\lambda} = \boldsymbol{m}/\sigma_\mathrm{avg}^2$$

Source Parameters:

$$\sigma_\mathrm{avg}^2 = 1/\Lambda, \quad \boldsymbol{m} = [m_1, \ldots, m_N]^\top = \boldsymbol{\lambda}/\Lambda$$

Crossover Between Subproblems:



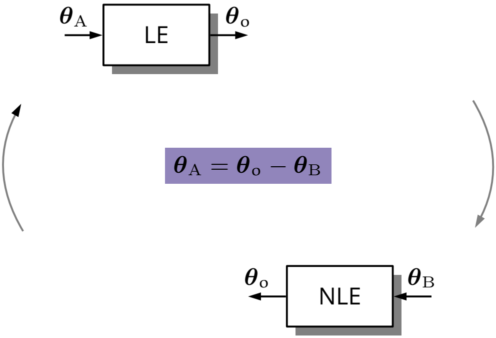
$$\boldsymbol{\theta}_\mathrm{A} = \boldsymbol{\theta}_\mathrm{o} - \boldsymbol{\theta}_\mathrm{B}$$

$$\mathsf{f}_\mathsf{x}(\mathbf{x} \mid \boldsymbol{y}) = \frac{1}{Z} \exp\left(-\frac{1}{2\sigma_n^2}\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} + \frac{1}{\sigma_n^2}\boldsymbol{A}^\top \boldsymbol{y}\right) \cdot \prod_{i=1}^{N} \mathsf{f}_\mathsf{x}(x_i)$$

$$\boldsymbol{\theta} = [\boldsymbol{\lambda},\, \Lambda], \quad \Lambda = 1/\sigma_\mathrm{avg}^2, \quad \lambda_i = m_i/\sigma_\mathrm{avg}^2 \quad \forall i \in \{1, \ldots, N\}$$

Crossover Between Subproblems:

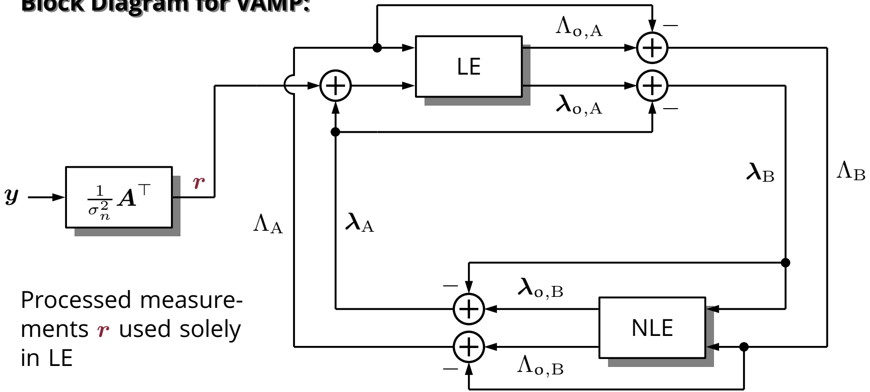
Iterative Manner



$$\mathsf{f}_\mathsf{x}(\mathbf{x} \mid \boldsymbol{y}) = \frac{1}{Z} \exp\left(-\frac{1}{2\sigma_n^2} \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A} \boldsymbol{x} + \frac{1}{\sigma_n^2} \boldsymbol{A}^\top \boldsymbol{y}\right) \cdot \prod_{i=1}^{N} \mathsf{f}_\mathsf{x}(x_i)$$

$$\boldsymbol{\theta} = [\boldsymbol{\lambda},\ \Lambda], \quad \Lambda = 1/\sigma_\mathrm{avg}^2, \quad \lambda_i = m_i/\sigma_\mathrm{avg}^2 \quad \forall i \in \{1, \ldots, N\}$$

Block Diagram for VAMP:



Processed measurements $\boldsymbol{r}$ used solely in LE

$$\mathrm{f}_{\mathbf{x}}(\mathbf{x} \mid \boldsymbol{y}) = \frac{1}{Z} \exp\left(-\frac{1}{2\sigma_n^2} \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A} \boldsymbol{x} + \frac{1}{\sigma_n^2} \boldsymbol{A}^\top \boldsymbol{y}\right) \cdot \prod_{i=1}^{N} \mathrm{f}_{\mathsf{x}}(x_i)$$

$$\boldsymbol{\theta} = [\boldsymbol{\lambda}, \Lambda], \quad \Lambda = 1/\sigma_{\text{avg}}^2, \quad \lambda_i = m_i/\sigma_{\text{avg}}^2 \quad \forall i \in \{1, \ldots, N\}$$

So Far: Usual Separation $\rightsquigarrow$ VAMP [RSF'19]

$$\mathsf{f}_{\mathsf{x}}(\mathbf{x} \mid \boldsymbol{y}) = \frac{1}{Z} \exp(-\frac{1}{2\sigma_n^2} \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x}) \cdot \exp(\boldsymbol{r}) \cdot \prod_{i=1}^N \mathsf{f}_{\mathsf{x}}(x_i)$$

with $\boldsymbol{r} = [r_1, \ldots, r_N]^\top = \frac{1}{\sigma_n^2} \boldsymbol{A}^\top \boldsymbol{y}$

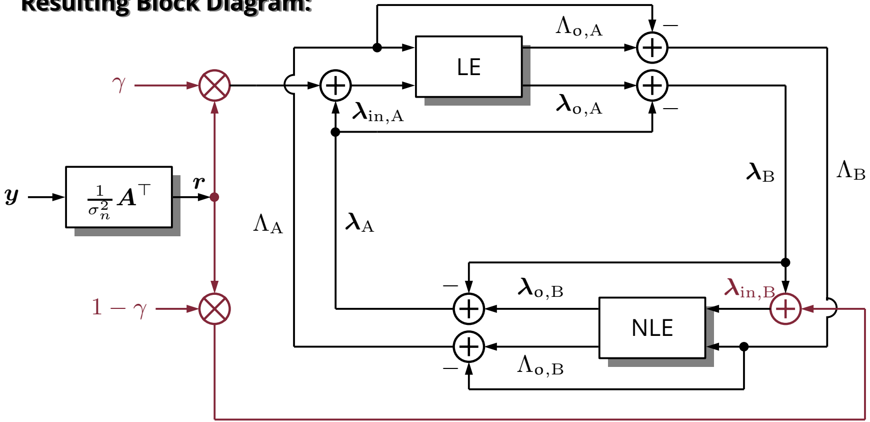
New: Variable Separation $\rightsquigarrow$ more general and flexible [SF'20]

$$\begin{aligned} \mathsf{f}_{\mathsf{x}}(\mathbf{x} \mid \boldsymbol{y}) &= \frac{1}{Z} \exp(-\frac{1}{2\sigma_n^2} \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x}) \cdot \exp(\frac{1}{\sigma_n^2} \boldsymbol{A}^\top \boldsymbol{y}) \cdot \prod_{i=1}^N \mathsf{f}_{\mathsf{x}}(x_i) \\ &= \frac{1}{Z} \exp(-\frac{1}{2\sigma_n^2} \boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x}) \cdot \exp(\gamma \boldsymbol{r}) \cdot \prod_{i=1}^N \mathsf{f}_{\mathsf{x}}(x_i) \exp((1-\gamma) r_i) \end{aligned}$$

with separation parameter: $\gamma \in [0, 1]$

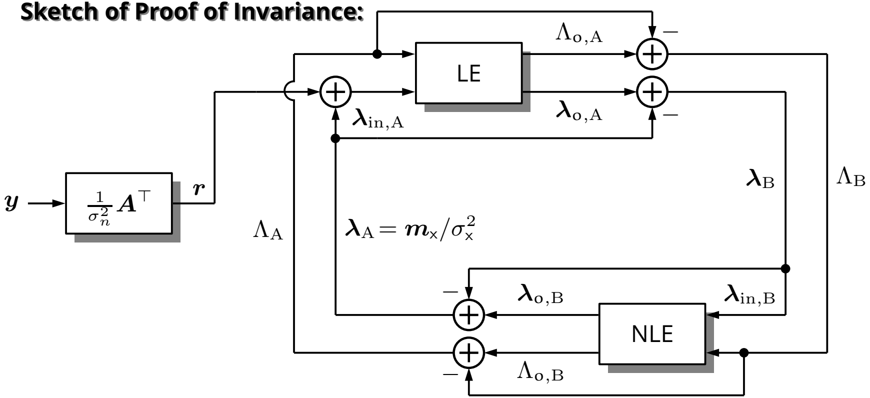
Application of EC framework directly leads to respective estimators

Resulting Block Diagram:



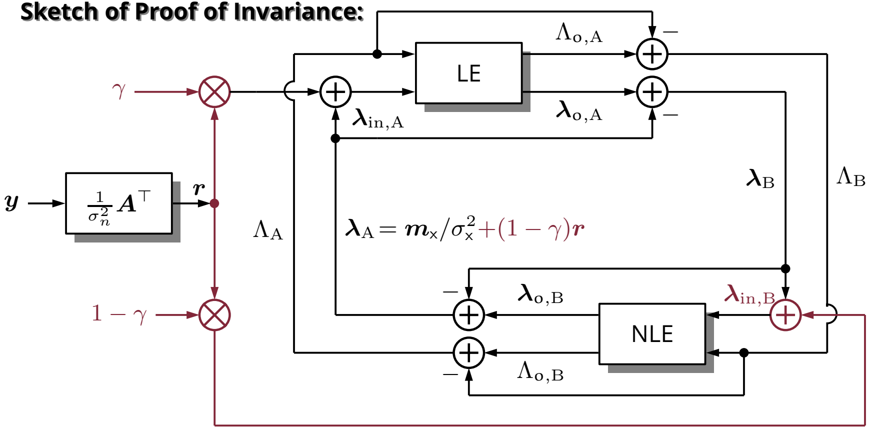
How does the performance change?

Sketch of Proof of Invariance:



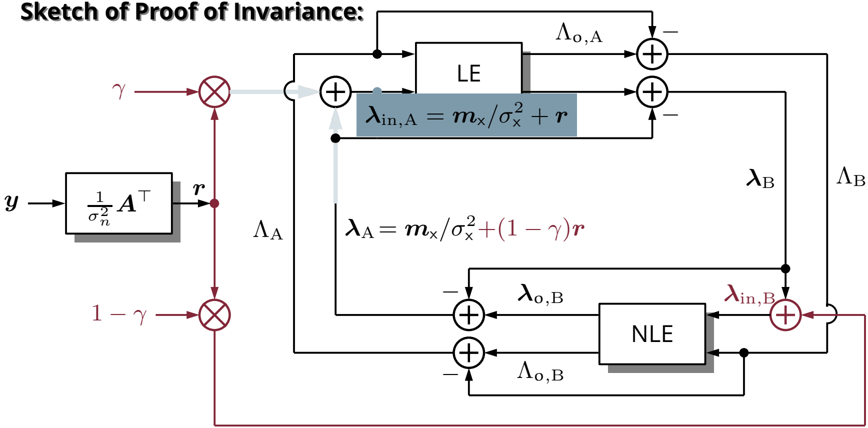
Usual initialization: $\Lambda_{\text{A}} = 1/\sigma_{\mathsf{x}}^2$, $\quad \boldsymbol{\lambda}_{\text{A}} = \boldsymbol{m}_{\mathsf{x}}/\sigma_{\mathsf{x}}^2 = \boldsymbol{0}$

Sketch of Proof of Invariance:



New initialization: $\Lambda_{\mathrm{A}} = 1/\sigma_{\mathsf{x}}^2$, $\quad \boldsymbol{\lambda}_{\mathrm{A}} = \boldsymbol{m}_{\mathsf{x}}/\sigma_{\mathsf{x}}^2 \; +(1-\gamma)\boldsymbol{r}$

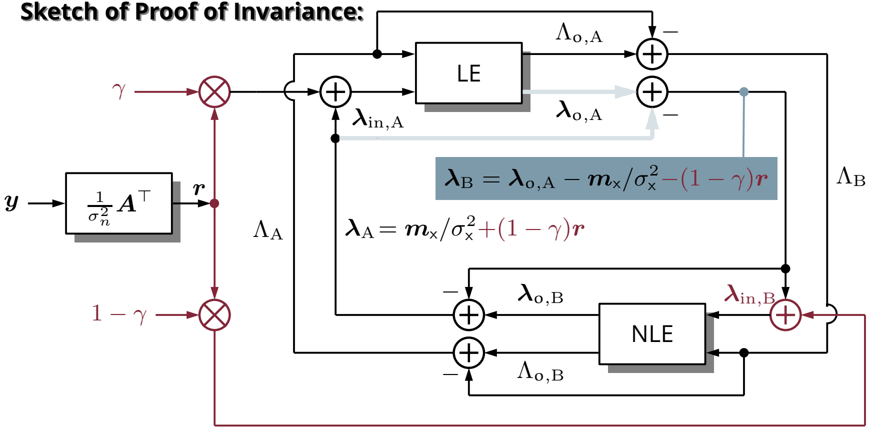
Sketch of Proof of Invariance:



New initialization: $\Lambda_{\mathrm{A}} = 1/\sigma_{\mathsf{x}}^2, \quad \boldsymbol{\lambda}_{\mathrm{A}} = \boldsymbol{m}_{\mathsf{x}}/\sigma_{\mathsf{x}}^2 + (1-\gamma)\boldsymbol{r}$

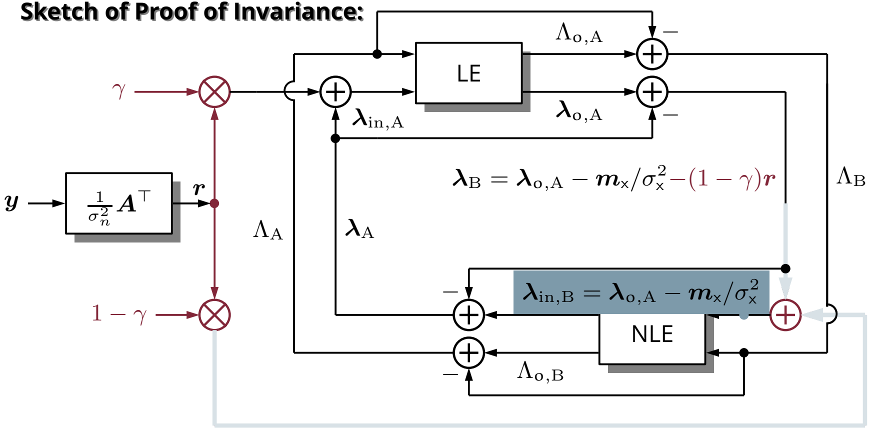
NACHRICHTENTECHNIK Universität Ulm

Sketch of Proof of Invariance:



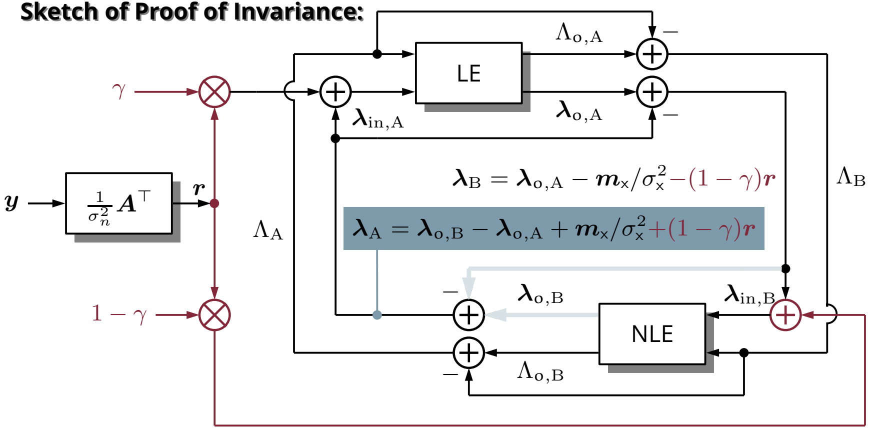
New initialization: $\Lambda_{\mathrm{A}} = 1/\sigma_{\mathsf{x}}^2$, $\quad \boldsymbol{\lambda}_{\mathrm{A}} = \boldsymbol{m}_{\mathsf{x}}/\sigma_{\mathsf{x}}^2 + (1-\gamma)\boldsymbol{r}$

Sketch of Proof of Invariance:



New initialization: $\Lambda_{\mathrm{A}} = 1/\sigma_{\mathrm{x}}^2$, $\quad \boldsymbol{\lambda}_{\mathrm{A}} = \boldsymbol{m}_{\mathrm{x}}/\sigma_{\mathrm{x}}^2 + (1-\gamma)\boldsymbol{r}$

Sketch of Proof of Invariance:



New initialization: $\Lambda_{\mathrm{A}} = 1/\sigma_{\mathrm{x}}^2, \quad \boldsymbol{\lambda}_{\mathrm{A}} = \boldsymbol{m}_{\mathrm{x}}/\sigma_{\mathrm{x}}^2 + (1-\gamma)\boldsymbol{r}$

Sketch of Proof of Invariance:

- No change in signals $\boldsymbol{\lambda}_{\mathrm{o,A}}$, $\boldsymbol{\lambda}_{\mathrm{o,B}}$, $\Lambda_{\mathrm{o,A}}$, and $\Lambda_{\mathrm{o,B}}$

  $\Rightarrow$ No change in performance

- Starting point $\boldsymbol{\lambda}_{\mathrm{A}} = \boldsymbol{m}_{\mathsf{x}}/\sigma_{\mathsf{x}}^2$ is optimal for usual separation

  $\Rightarrow$ $\boldsymbol{\lambda}_{\mathrm{A}} = \boldsymbol{m}_{\mathsf{x}}/\sigma_{\mathsf{x}}^2 + (1-\gamma)\boldsymbol{r}$ optimal for variable separation

NACHRICHTENTECHNIK Universität Ulm

Setting and Parameters:

- Gaussian sensing matrix $\boldsymbol{A} \in \mathbb{R}^{M \times N}$ with unit columns and
  - $N = 258$,
  - $M = 129$.

$$\mathbf{y} = \boldsymbol{A}\mathbf{x} + \mathbf{n}$$

- Signal-to-noise-ratio: $10 \log_{10}(1/\sigma_n^2) \;=\; 17\mathrm{dB}$

$$\mathbf{n} \sim \mathcal{N}(\mathbf{0}, \sigma_n^2 \boldsymbol{I}_M)$$

- Pdf of signal $\mathbf{x}$ with sparsity $s = 12$:
  - Discrete prior

$$\mathrm{f}_{\mathrm{x}}(x) = \frac{s}{N}\delta(x+1) + \frac{N-s}{N}\delta(x) + \frac{s}{N}\delta(x-1)$$

  - Bernoulli-Gaussian prior

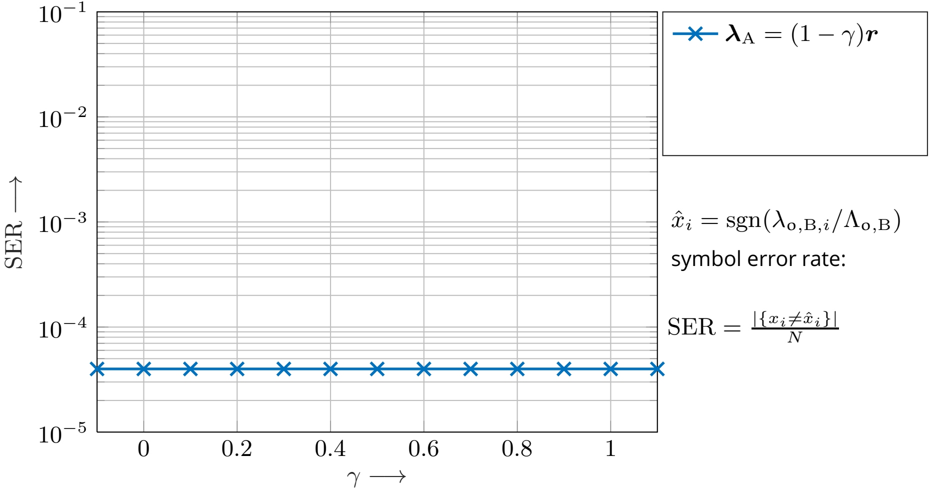
$$\mathrm{f}_{\mathbf{x}}(\boldsymbol{x}) = \prod_{i=1}^{N} \mathrm{f}_{\mathrm{x}}(x_i)$$

$$\mathrm{f}_{\mathrm{x}}(x) = \frac{N-s}{N}\delta(x) + \frac{s}{N}\frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}$$

$\Rightarrow \boldsymbol{m}_{\mathrm{x}} = \mathbf{0}$, $\Lambda_{\mathrm{A}} = 1/\sigma_{\mathrm{x}}^2 = N/s$

- 20 iterations performed
- $N - s$ smallest (in magnitude) values are rounded to 0

Performance for Different Starting Points—Discrete Prior:

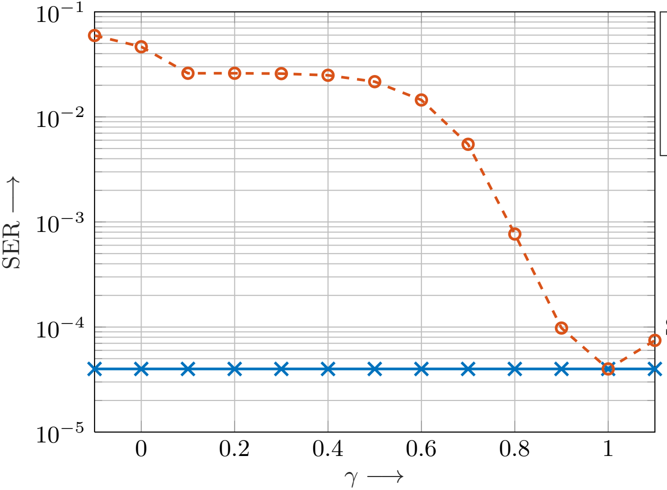


$\hat{x}_i = \mathrm{sgn}(\lambda_{\mathrm{o,B},i}/\Lambda_{\mathrm{o,B}})$

symbol error rate:

$$\mathrm{SER} = \frac{|\{x_i \neq \hat{x}_i\}|}{N}$$

Performance for Different Starting Points—Discrete Prior:

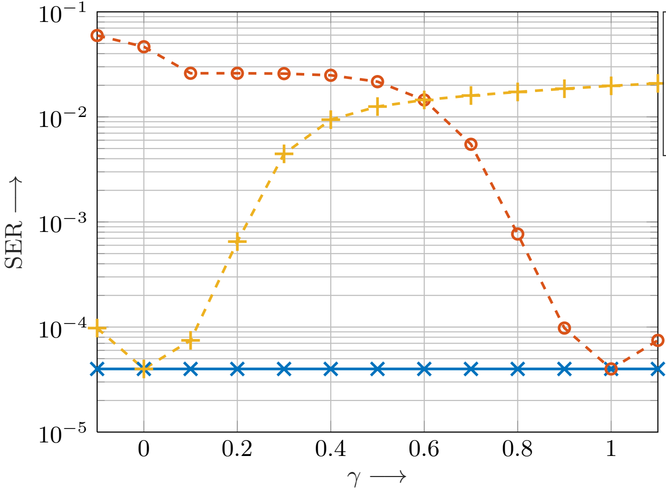


$\hat{x}_i = \mathrm{sgn}(\lambda_{\mathrm{o,B},i}/\Lambda_{\mathrm{o,B}})$

symbol error rate:

$$\mathrm{SER} = \frac{|\{x_i \neq \hat{x}_i\}|}{N}$$

Performance for Different Starting Points—Discrete Prior:

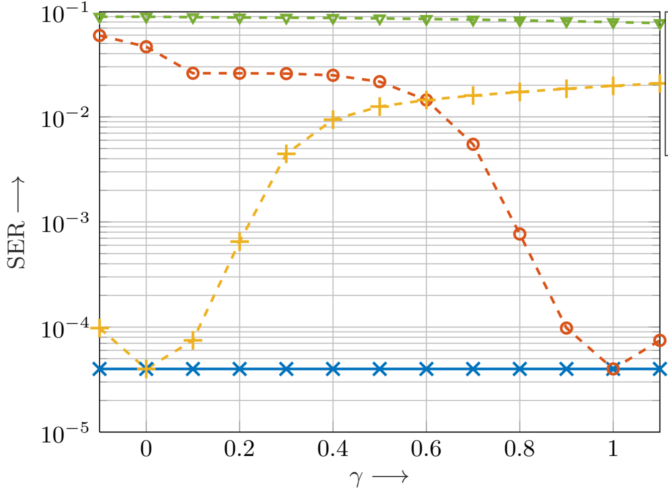


$\hat{x}_i = \mathrm{sgn}(\lambda_{\mathrm{o,B},i}/\Lambda_{\mathrm{o,B}})$

symbol error rate:

$$\mathrm{SER} = \frac{|\{x_i \neq \hat{x}_i\}|}{N}$$

Performance for Different Starting Points—Discrete Prior:

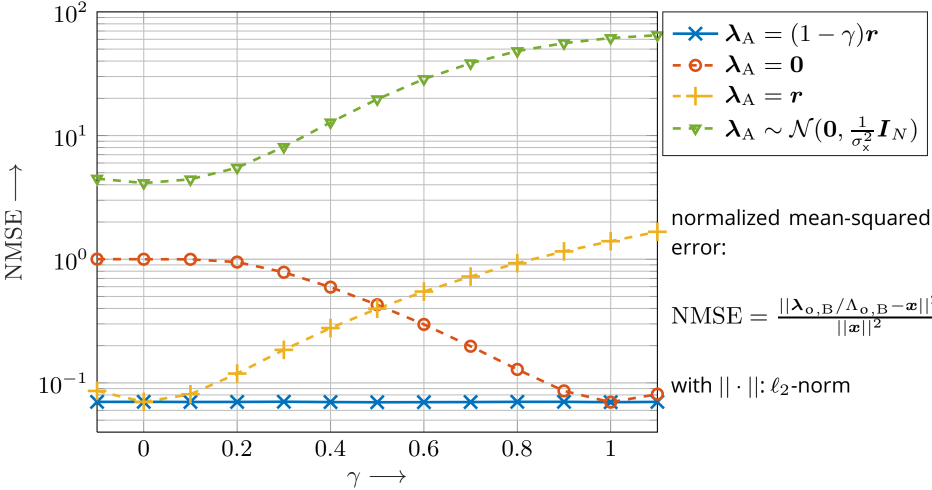


$$\hat{x}_i = \operatorname{sgn}(\lambda_{\mathrm{o,B},i}/\Lambda_{\mathrm{o,B}})$$

symbol error rate:

$$\mathrm{SER} = \frac{|\{x_i \neq \hat{x}_i\}|}{N}$$

Performance for Different Starting Points—Bernoulli-Gaussian Prior:



normalized mean-squared error:

$$\mathrm{NMSE} = \frac{||\boldsymbol{\lambda}_{\mathrm{o,B}}/\Lambda_{\mathrm{o,B}} - \boldsymbol{x}||^2}{||\boldsymbol{x}||^2}$$

with $||\cdot||$: ℓ_2-norm

VAMP with average variance [RSF'19]

$$f_{\mathrm{A}}(\boldsymbol{x} \mid \boldsymbol{y}) = \exp(-\tfrac{1}{2\sigma_n^2}\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{r}(\boldsymbol{y}))$$

$$f_{\mathrm{B}}(\boldsymbol{x} \mid \boldsymbol{y}) = \prod_{i=1}^{N} \mathsf{f_x}(x_i)$$

. . . with average variance and variable separation [SF'20]

$$f_{\mathrm{A}}(\boldsymbol{x} \mid \boldsymbol{y}) = \exp(-\tfrac{1}{2\sigma_n^2}\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} + \gamma\boldsymbol{r}(\boldsymbol{y}))$$

$$f_{\mathrm{B}}(\boldsymbol{x} \mid \boldsymbol{y}) = \prod_{i=1}^{N} \mathsf{f_x}(x_i) \cdot \exp((1-\gamma)\boldsymbol{r}(\boldsymbol{y}))$$

$$\boldsymbol{g}(\boldsymbol{x}) = [x_1, \ldots, x_N, \ -\tfrac{1}{2}\textstyle\sum_{i=1}^{N} x_i^2]^\top$$

$$\boldsymbol{\theta} = [\lambda_1, \ldots, \lambda_N, \quad \Lambda \quad]^\top = [\boldsymbol{\lambda}^\top, \Lambda]^\top$$

NACHRICHTENTECHNIK Universität Ulm

VAMP with average variance [RSF'19]

$$f_{\mathrm{A}}(\boldsymbol{x} \mid \boldsymbol{y}) = \exp(-\tfrac{1}{2\sigma_n^2}\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{r}(\boldsymbol{y}))$$

$$f_{\mathrm{B}}(\boldsymbol{x} \mid \boldsymbol{y}) = \textstyle\prod_{i=1}^N \mathsf{f}_{\mathsf{x}}(x_i)$$

... with average variance and variable separation [SF'20]

$$f_{\mathrm{A}}(\boldsymbol{x} \mid \boldsymbol{y}) = \exp(-\tfrac{1}{2\sigma_n^2}\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} + \gamma\boldsymbol{r}(\boldsymbol{y}))$$

$$f_{\mathrm{B}}(\boldsymbol{x} \mid \boldsymbol{y}) = \textstyle\prod_{i=1}^N \mathsf{f}_{\mathsf{x}}(x_i) \cdot \exp((1-\gamma)\boldsymbol{r}(\boldsymbol{y}))$$

$$\boldsymbol{g}(\boldsymbol{x}) = [x_1, \ldots, x_N, \ -\tfrac{1}{2}\textstyle\sum_{i=1}^N x_i^2]^\top$$

$$\boldsymbol{\theta} = [\lambda_1, \ldots, \lambda_N, \quad \Lambda \quad]^\top = [\boldsymbol{\lambda}^\top, \Lambda]^\top$$

... with individual variances

$$f_{\mathrm{A}}(\boldsymbol{x} \mid \boldsymbol{y}) = \exp(-\tfrac{1}{2\sigma_n^2}\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} + \boldsymbol{r}(\boldsymbol{y}))$$

$$f_{\mathrm{B}}(\boldsymbol{x} \mid \boldsymbol{y}) = \textstyle\prod_{i=1}^N \mathsf{f}_{\mathsf{x}}(x_i)$$

...with individual variances and variable separation [here]

$$f_{\mathrm{A}}(\boldsymbol{x} \mid \boldsymbol{y}) = \exp(-\tfrac{1}{2\sigma_n^2}\boldsymbol{x}^\top \boldsymbol{A}^\top \boldsymbol{A}\boldsymbol{x} + \gamma\boldsymbol{r}(\boldsymbol{y}))$$

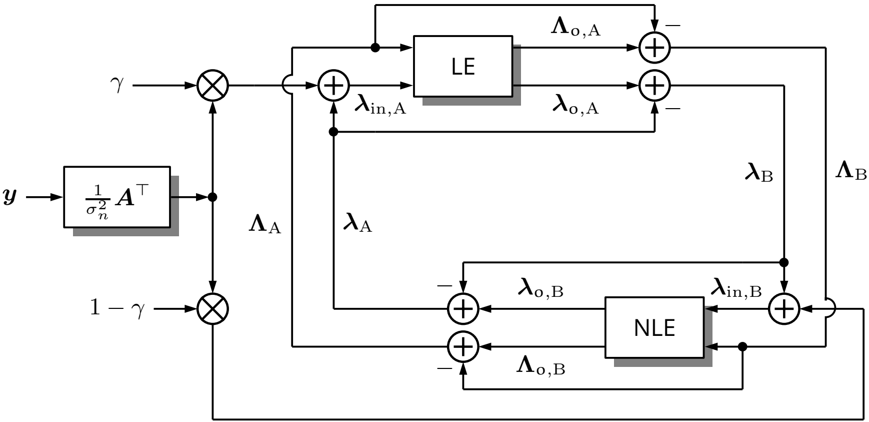
$$f_{\mathrm{B}}(\boldsymbol{x} \mid \boldsymbol{y}) = \textstyle\prod_{i=1}^N \mathsf{f}_{\mathsf{x}}(x_i) \cdot \exp((1-\gamma)\boldsymbol{r}(\boldsymbol{y}))$$

$$\boldsymbol{g}(\boldsymbol{x}) = \left[x_1, \ldots, x_N, -\frac{x_1^2}{2}, \ldots, -\frac{x_N^2}{2}\right]^\top$$

$$\boldsymbol{\theta} = [\lambda_1, \ldots, \lambda_N, \ \Lambda_1, \ldots, \ \Lambda_N]^\top = [\boldsymbol{\lambda}^\top, \boldsymbol{\Lambda}^\top]^\top$$

NACHRICHTENTECHNIK Universität Ulm

Individual Variances and Variable Separation—Straight-Forward Derivation:



Disadvantage: Worse performance than average variance case

Parameter Connections at LE:

- Before LE: $\Lambda_{\mathrm{A},i} = 1/\varsigma^2_{\mathrm{A},i}$, $\lambda_{\mathrm{A},i} = m_{\mathrm{A},i}/\varsigma^2_{\mathrm{A},i}$
- After LE: $\Lambda_{\mathrm{o,A},i} = 1/\sigma^2_{\mathrm{o,A},i}$, $\lambda_{\mathrm{o,A},i} = m_{\mathrm{o,A},i}/\sigma^2_{\mathrm{o,A},i}$

Crossover: $\forall\, i \in \{1, \ldots, N\}$

$$\Lambda_{\mathrm{B},i} = 1/\sigma^2_{\mathrm{B},i} = \Lambda_{\mathrm{o,A},i} - \Lambda_{\mathrm{A},i}, \qquad \sigma^2_{\mathrm{B},i} = \left(\frac{1}{\sigma^2_{\mathrm{o,A},i}} - \frac{1}{\varsigma^2_{\mathrm{A},i}} \right)^{-1},$$

$$\lambda_{\mathrm{B},i} = m_{\mathrm{B},i}/\sigma^2_{\mathrm{B},i} = \lambda_{\mathrm{o,A},i} - \lambda_{\mathrm{A},i}, \qquad m_{\mathrm{B},i} = \sigma^2_{\mathrm{B},i} \left(\frac{m_{\mathrm{o,A},i}}{\sigma^2_{\mathrm{o,A},i}} - \frac{m_{\mathrm{A},i}}{\varsigma^2_{\mathrm{A},i}} \right)$$

Procedure known as

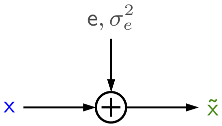
- calculation of extrinsics (turbo algorithms)
- “Onsager” correction (VAMP) [RSF’19]
- bias compensation (signal processing) [SF’18]

Scalar Estimation Problem:

- Observation
- Estimation noise

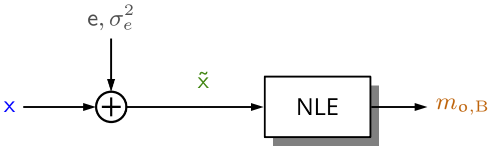
$$\tilde{x} = x + e$$

$$e \sim \mathcal{N}(0, \sigma_e^2)$$



Scalar Estimation Problem:

- Observation $\tilde{\mathsf{x}} = \mathsf{x} + \mathsf{e}$
- Estimation noise $\mathsf{e} \sim \mathcal{N}(0, \sigma_e^2)$

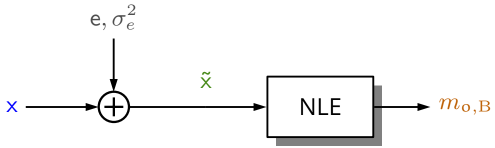


MMSE Estimates: (Estimation error: $\mathsf{e}_{\mathrm{b}} = \mathsf{x} - m_{\mathrm{o,B}}$)

- Conditional mean $m_{\mathrm{o,B}}(\tilde{x}) = \mathrm{E}_{\mathsf{x}}\{\mathsf{x} \mid \tilde{x}\}$
- Conditional variance $\varsigma^2_{\mathrm{o,B}}(\tilde{x}) = \mathrm{E}_{\mathsf{x}}\left\{(\mathsf{x} - m_{\mathrm{o,B}})^2 \mid \tilde{x}\right\}$
- Mean-squared error (MSE) $\sigma^2_{\mathrm{b}} = \mathrm{E}_{\tilde{\mathsf{x}}}\left\{\varsigma^2_{\mathrm{o,B}}(\tilde{\mathsf{x}})\right\}$

Scalar Estimation Problem:

- Observation $\tilde{x} = x + e$
- Estimation noise $e \sim \mathcal{N}(0, \sigma_e^2)$



MMSE Estimates: (Estimation error: $e_b = x - m_{o,B}$)

- Conditional mean $m_{o,B}(\tilde{x}) = E_x\{x \mid \tilde{x}\}$
- Conditional variance $\varsigma_{o,B}^2(\tilde{x}) = E_x\{(x - m_{o,B})^2 \mid \tilde{x}\}$
- Mean-squared error (MSE) $\sigma_b^2 = E_{\tilde{x}}\{\varsigma_{o,B}^2(\tilde{x})\}$

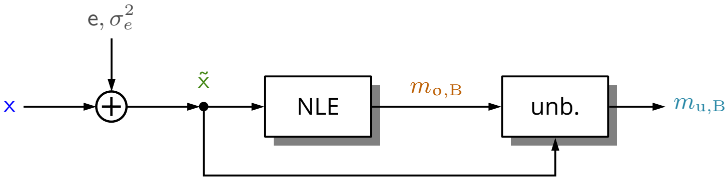
Problem: By definition of MMSE estimates: $e_b \perp \tilde{x}$ and $e_b \perp m_{o,B}(\tilde{x})$

$\Rightarrow$ Parts of signal x are accounted to e_b

$\Rightarrow$ Remove bias via scaling $\rightsquigarrow$ unbiasing

Scalar Estimation Problem:

- Observation $\tilde{x} = x + e$
- Estimation noise $e \sim \mathcal{N}(0, \sigma_e^2)$



MMSE Estimates: (Estimation error: $e_b = x - m_{o,B}$)

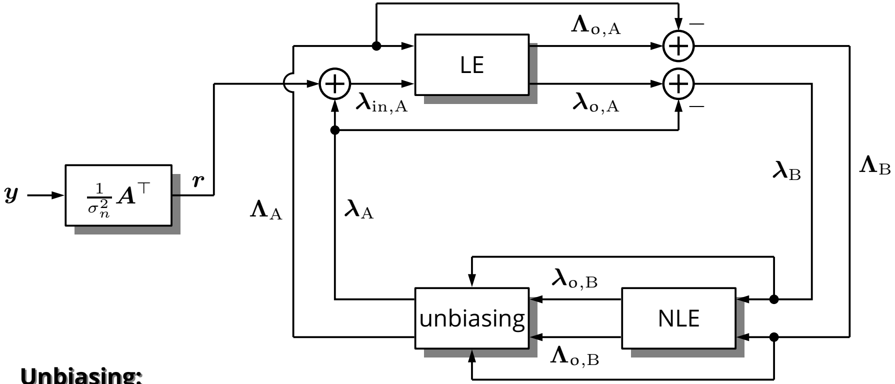
- Conditional mean $m_{o,B}(\tilde{x}) = E_x\{x \mid \tilde{x}\}$
- Conditional variance $\varsigma^2_{o,B}(\tilde{x}) = E_x\{(x - m_{o,B})^2 \mid \tilde{x}\}$
- Mean-squared error (MSE) $\sigma_b^2 = E_{\tilde{x}}\{\varsigma^2_{o,B}(\tilde{x})\}$

Unbiasing: [SF'18, FSG'20]

- Unb. conditional variance $\varsigma^2_{u,B}(y) = \varsigma^2_{o,B} + (\sigma_b^2/(\sigma_e^2 - \sigma_b^2))^2 (m_{o,B}(\tilde{x}) - \tilde{x})^2$
- Unb. estimate $m_{u,B} = \left(1/\sigma_b^2 - 1/\sigma_e^2\right)^{-1} (m_{o,B}(\tilde{x})/\sigma_b^2 - \tilde{x}/\sigma_e^2)$

NACHRICHTENTECHNIK Universität Ulm

EC with Individual Variances—Improved by MMSE Estimation: [FSG'20]

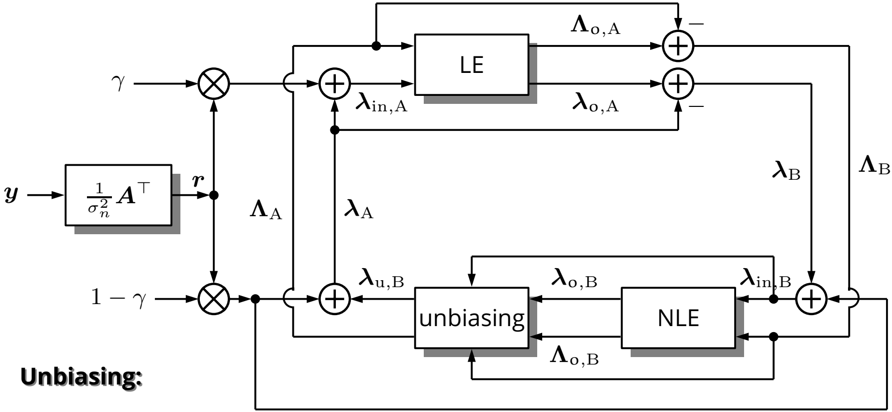


Unbiasing:

$$\Lambda_{A,i} = \left(\frac{1}{\Lambda_{o,B,i}} + \left(\frac{\Lambda_{B,i}}{\Lambda_{b,i} - \Lambda_{B,i}} \left(\frac{\lambda_{b,i}}{\Lambda_{b,i}} - \frac{\lambda_{B,i}}{\Lambda_{B,i}} \right) \right)^2 \right)^{-1}$$

$$\lambda_{A,i} = \frac{\Lambda_{A,i}}{\Lambda_{b,i} - \Lambda_{B,i}} (\lambda_{b,i} - \lambda_{B,i})$$

Combination—Improved by MMSE Estimation:

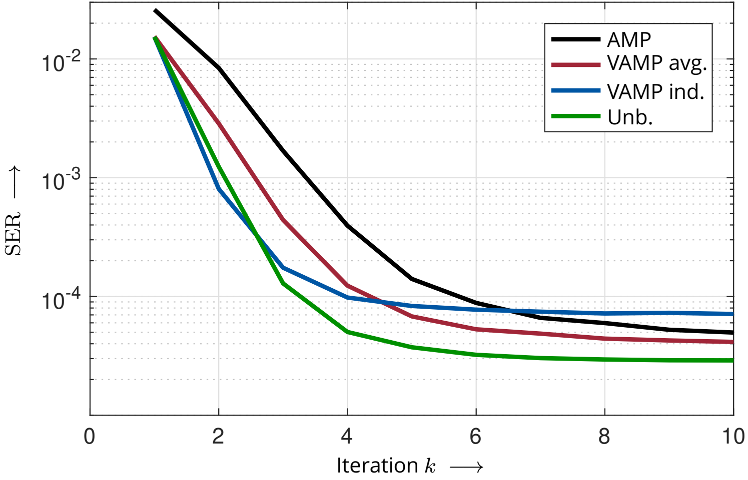


Unbiasing:

$$\Lambda_{\mathrm{A},i} = \left(\frac{1}{\Lambda_{\mathrm{o,B},i}} + \left(\frac{\Lambda_{\mathrm{B},i}}{\Lambda_{\mathrm{b},i} - \Lambda_{\mathrm{B},i}} \left(\frac{\lambda_{\mathrm{b},i}}{\Lambda_{\mathrm{b},i}} - \frac{\lambda_{\mathrm{in,B},i}}{\Lambda_{\mathrm{B},i}} \right) \right)^2 \right)^{-1}$$

$$\lambda_{\mathrm{A},i} = \frac{\Lambda_{\mathrm{A},i}}{\Lambda_{\mathrm{b},i} - \Lambda_{\mathrm{B},i}} (\lambda_{\mathrm{b},i} - \lambda_{\mathrm{in,B},i}) + (1-\gamma) r_i$$

SER over Iteration: $10\log_{10}(1/\sigma_n^2) \widehat{=} 17\,\text{dB}$, $N = 258$, $s = 12$, discrete prior



Summary:

- Review of EC for Compressed Sensing
- Introduction of a *variable separation*
- Proof of *invariance of performance* with optimal starting point
- Presentation of correct treatment of *individual variances*

Conclusion:

- Variable separation gives a *new view point* on iterative algorithms for Compressed Sensing
- Individual variances show *improved performance* compared to average variances in recovery algorithms for Compressed Sensing if treated correctly

Outlook:

- Analysis of iterative algorithms based on variable separation / individual variances
- Complexity analysis

[Fis'19] R. F. H. Fischer
Expectation-Consistent Approximate Inference with Vector-Valued Diagonalization—An Unbiasing Interpretation.
Invited Talk at *Workshop on Mathematical Data Science (MDS)*, Dürnstein, Austria, 2019.

[FSG'20] R. F. H. Fischer, C. Sippel, and N. Görtz.
VAMP with Vector-Valued Diagonalization.
Accepted for *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Barcelona, Spain, 2020.

[SF'20] C. Sippel and R. F. H. Fischer.
On the Invariance of Recovery Algorithms for Compressed Sensing based on Expectation-Consistent Approximate Inference.
Accepted for *24th International ITG Workshop on Smart Antennas (WSA 2020)*, Hamburg, Germany, February 2020.

References

[Bro'86] L. D. Brown.
Fundamentals of Statistical Exponential Families: with Applications in Statistical Decision Theory.
Institute of Mathematical Statistics, 1986.

[CRT'06] E. J. Candès, J. Romberg, and T. Tao.
Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information.
In *IEEE Transactions on Information Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[Don'06] D. L. Donoho.
Compressed Sensing.
In *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[FSR+16] A. Fletcher, M. Sahraee-Ardakan, S. Rangan, and P. Schniter.
Expectation Consistent Approximate Inference: Generalizations and Convergence.
In *IEEE International Symposium on Information Theory (ISIT)*, pp. 190–194, 2016.

[MV'18] P. Moulin and V. Veeravalli.
Statistical Inference for Engineers and Data Scientists.
Cambridge University Press, 2018.

[OW'05] M. Opper and O. Winther.
Expectation Consistent Approximate Inference.
In *Journal of Machine Learning Research*, vol. 6, pp. 2177–2204, 2005.

[RSF'19] S. Rangan, P. Schniter, and A. K. Fletcher.
Vector Approximate Message Passing.
In *IEEE Transactions on Information Theory*, vol. 65, no. 10, pp. 6664–6684, 2019.

[SF'18] S. Sparrer and R. F. H. Fischer.
Bias Compensation in Iterative Soft-Feedback Algorithms With Application to (Discrete) Compressed Sensing.
In *IEEE Statistical Signal Processing Workshop (SSP)*, pp. 35–39, Freiburg, June 2018.

Algorithm 1: $[\boldsymbol{\lambda}_{\mathrm{o,B}}, \Lambda_{\mathrm{o,B}}] = \texttt{VAMP}(\boldsymbol{y},\, \boldsymbol{m}_{\mathsf{x}},\, \sigma_{\mathsf{x}}^2,\, \boldsymbol{A},\, \sigma_n^2)$

$\boldsymbol{r} = \frac{1}{\sigma_n^2}\boldsymbol{A}^\top \boldsymbol{y},\ \Lambda_{\mathrm{A}} = 1/\sigma_{\mathsf{x}}^2,$ // $\hat{\sigma}_{\mathrm{A}}^2 = \sigma_{\mathsf{x}}^2$

$\boldsymbol{\lambda}_{\mathrm{A}} = \boldsymbol{m}_{\mathsf{x}}/\sigma_{\mathsf{x}}^2$ // $\boldsymbol{m}_{\mathrm{A}} = \boldsymbol{m}_{\mathsf{x}}$

while *stopping criterion not met* **do**

- $\boldsymbol{m}_{\mathrm{o,A}} = \sigma_n^2(\boldsymbol{A}^\top\boldsymbol{A} + \sigma_n^2\Lambda_{\mathrm{A}}\boldsymbol{I}_N)^{-1}(\boldsymbol{r} + \boldsymbol{\lambda}_{\mathrm{A}})$ // Linear Estimator
- $\hat{\sigma}_{\mathrm{o,A}}^2 = \mathrm{trace}(\sigma_n^2(\boldsymbol{A}^\top\boldsymbol{A} + \sigma_n^2\Lambda_{\mathrm{A}}\boldsymbol{I}_N)^{-1})/N$
- $\Lambda_{\mathrm{o,A}} = 1/\hat{\sigma}_{\mathrm{o,A}}^2 = \mathrm{fct}(\Lambda_{\mathrm{A}})$
- $\boldsymbol{\lambda}_{\mathrm{o,A}} = \boldsymbol{m}_{\mathrm{o,A}}/\hat{\sigma}_{\mathrm{o,A}}^2 = \mathrm{fct}(\Lambda_{\mathrm{A}}, \boldsymbol{r} + \boldsymbol{\lambda}_{\mathrm{A}})$
- $\Lambda_{\mathrm{B}} = \Lambda_{\mathrm{o,A}} - \Lambda_{\mathrm{A}}$ // $\hat{\sigma}_{\mathrm{B}}^2 = (\frac{1}{\hat{\sigma}_{\mathrm{o,A}}^2} - \frac{1}{\hat{\sigma}_{\mathrm{A}}^2})^{-1}$
- $\boldsymbol{\lambda}_{\mathrm{B}} = \boldsymbol{\lambda}_{\mathrm{o,A}} - \boldsymbol{\lambda}_{\mathrm{A}}$ // $\boldsymbol{m}_{\mathrm{B}} = \hat{\sigma}_{\mathrm{B}}^2(\frac{\boldsymbol{m}_{\mathrm{o,A}}}{\hat{\sigma}_{\mathrm{o,A}}^2} - \frac{\boldsymbol{m}_{\mathrm{A}}}{\hat{\sigma}_{\mathrm{A}}^2})$
- $\boldsymbol{m}_{\mathrm{o,B}} = \mathrm{E}_{\mathbf{x},\mathrm{B}}\{\boldsymbol{g}_\lambda(\mathbf{x})\}$ // Non-Linear Estimator
- $\hat{\sigma}_{\mathrm{o,B}}^2 = (-2\mathrm{E}_{\mathbf{x},\mathrm{B}}\{g_\Lambda(\mathbf{x})\} - \boldsymbol{m}_{\mathrm{o,B}}^\top\boldsymbol{m}_{\mathrm{o,B}})/N$
- $\Lambda_{\mathrm{o,B}} = 1/\hat{\sigma}_{\mathrm{o,B}}^2 = \mathrm{fct}(\Lambda_{\mathrm{B}}, \boldsymbol{\lambda}_{\mathrm{B}})$
- $\boldsymbol{\lambda}_{\mathrm{o,B}} = \boldsymbol{m}_{\mathrm{o,B}}/\hat{\sigma}_{\mathrm{o,B}}^2 = \mathrm{fct}(\Lambda_{\mathrm{B}}, \boldsymbol{\lambda}_{\mathrm{B}})$
- $\Lambda_{\mathrm{A}} = \Lambda_{\mathrm{o,B}} - \Lambda_{\mathrm{B}}$ // $\hat{\sigma}_{\mathrm{A}}^2 = (\frac{1}{\hat{\sigma}_{\mathrm{o,B}}^2} - \frac{1}{\hat{\sigma}_{\mathrm{B}}^2})^{-1}$
- $\boldsymbol{\lambda}_{\mathrm{A}} = \boldsymbol{\lambda}_{\mathrm{o,B}} - \boldsymbol{\lambda}_{\mathrm{B}}$ // $\boldsymbol{m}_{\mathrm{A}} = \hat{\sigma}_{\mathrm{A}}^2(\frac{\boldsymbol{m}_{\mathrm{o,B}}}{\hat{\sigma}_{\mathrm{o,B}}^2} - \frac{\boldsymbol{m}_{\mathrm{B}}}{\hat{\sigma}_{\mathrm{B}}^2})$

Algorithm for EC with Variable Separation

Algorithm 2: $[\boldsymbol{\lambda}_{\mathrm{o,B}}, \Lambda_{\mathrm{o,B}}] = \texttt{VAMPvar}(\boldsymbol{y},\ \boldsymbol{m}_{\mathsf{x}},\ \sigma_{\mathsf{x}}^2,\ \boldsymbol{A},\ \sigma_n^2, \gamma)$

$\boldsymbol{r} = \frac{1}{\sigma_n^2}\boldsymbol{A}^\top\boldsymbol{y}$, $\Lambda_{\mathrm{A}} = 1/\sigma_{\mathsf{x}}^2$, // $\hat{\sigma}_{\mathrm{A}}^2 = \sigma_{\mathsf{x}}^2$

$\boldsymbol{\lambda}_{\mathrm{A}} = \boldsymbol{m}_{\mathsf{x}}/\sigma_{\mathsf{x}}^2 + (1-\gamma)\boldsymbol{r}$ // $\boldsymbol{m}_{\mathrm{A}} = \boldsymbol{m}_{\mathsf{x}} + \sigma_{\mathsf{x}}^2(1-\gamma)\boldsymbol{r}$

while *stopping criterion not met* **do**

- $\boldsymbol{m}_{\mathrm{o,A}} = \sigma_n^2(\boldsymbol{A}^\top\boldsymbol{A} + \sigma_n^2\Lambda_{\mathrm{A}}\boldsymbol{I}_N)^{-1}(\gamma\boldsymbol{r} + \boldsymbol{\lambda}_{\mathrm{A}})$ // Linear Estimator
- $\hat{\sigma}_{\mathrm{o,A}}^2 = \mathrm{trace}(\sigma_n^2(\boldsymbol{A}^\top\boldsymbol{A} + \sigma_n^2\Lambda_{\mathrm{A}}\boldsymbol{I}_N)^{-1})/N$
- $\Lambda_{\mathrm{o,A}} = 1/\hat{\sigma}_{\mathrm{o,A}}^2 = \mathrm{fct}(\Lambda_{\mathrm{A}})$
- $\boldsymbol{\lambda}_{\mathrm{o,A}} = \boldsymbol{m}_{\mathrm{o,A}}/\hat{\sigma}_{\mathrm{o,A}}^2 = \mathrm{fct}(\Lambda_{\mathrm{A}}, \gamma\boldsymbol{r} + \boldsymbol{\lambda}_{\mathrm{A}})$
- $\Lambda_{\mathrm{B}} = \Lambda_{\mathrm{o,A}} - \Lambda_{\mathrm{A}}$ // $\hat{\sigma}_{\mathrm{B}}^2 = (\frac{1}{\hat{\sigma}_{\mathrm{o,A}}^2} - \frac{1}{\hat{\sigma}_{\mathrm{A}}^2})^{-1}$
- $\boldsymbol{\lambda}_{\mathrm{B}} = \boldsymbol{\lambda}_{\mathrm{o,A}} - \boldsymbol{\lambda}_{\mathrm{A}}$ // $\boldsymbol{m}_{\mathrm{B}} = \hat{\sigma}_{\mathrm{B}}^2(\frac{\boldsymbol{m}_{\mathrm{o,A}}}{\hat{\sigma}_{\mathrm{o,A}}^2} - \frac{\boldsymbol{m}_{\mathrm{A}}}{\hat{\sigma}_{\mathrm{A}}^2})$
- $\boldsymbol{m}_{\mathrm{o,B}} = \mathrm{E}_{\mathsf{x},\mathrm{B}}\{\boldsymbol{g}_\lambda(\mathsf{x})\}$ // Non-Linear Estimator
- $\hat{\sigma}_{\mathrm{o,B}}^2 = (-2\mathrm{E}_{\mathsf{x},\mathrm{B}}\{g_\Lambda(\mathsf{x})\} - \boldsymbol{m}_{\mathrm{o,B}}^\top\boldsymbol{m}_{\mathrm{o,B}})/N$
- $\Lambda_{\mathrm{o,B}} = 1/\hat{\sigma}_{\mathrm{o,B}}^2 = \mathrm{fct}(\Lambda_{\mathrm{B}}, (1-\gamma)\boldsymbol{r} + \boldsymbol{\lambda}_{\mathrm{B}})$
- $\boldsymbol{\lambda}_{\mathrm{o,B}} = \boldsymbol{m}_{\mathrm{o,B}}/\hat{\sigma}_{\mathrm{o,B}}^2 = \mathrm{fct}(\Lambda_{\mathrm{B}}, (1-\gamma)\boldsymbol{r} + \boldsymbol{\lambda}_{\mathrm{B}})$
- $\Lambda_{\mathrm{A}} = \Lambda_{\mathrm{o,B}} - \Lambda_{\mathrm{B}}$ // $\hat{\sigma}_{\mathrm{A}}^2 = (\frac{1}{\hat{\sigma}_{\mathrm{o,B}}^2} - \frac{1}{\hat{\sigma}_{\mathrm{B}}^2})^{-1}$
- $\boldsymbol{\lambda}_{\mathrm{A}} = \boldsymbol{\lambda}_{\mathrm{o,B}} - \boldsymbol{\lambda}_{\mathrm{B}}$ // $\boldsymbol{m}_{\mathrm{A}} = \hat{\sigma}_{\mathrm{A}}^2(\frac{\boldsymbol{m}_{\mathrm{o,B}}}{\hat{\sigma}_{\mathrm{o,B}}^2} - \frac{\boldsymbol{m}_{\mathrm{B}}}{\hat{\sigma}_{\mathrm{B}}^2})$

Individual Variances and Variable Separation

Algorithm 3: $[\boldsymbol{\lambda}_{\text{o,B}}, \boldsymbol{\Lambda}_{\text{o,B}}] = \texttt{ECind\_var}(\boldsymbol{y},\ \boldsymbol{m}_{\mathsf{x}},\ \sigma_{\mathsf{x}}^2,\ \boldsymbol{A},\ \sigma_n^2, \gamma)$

$\boldsymbol{r} = \frac{1}{\sigma_n^2}\boldsymbol{A}^\top \boldsymbol{y} = [r_1,\ \ldots,\ r_N]^\top$, $\boldsymbol{\Lambda}_{\text{A}} = 1/\sigma_{\mathsf{x}}^2 \boldsymbol{I}$, // $\varsigma_{\text{A},i}^2 = \sigma_{\mathsf{x}}^2\ \forall i \in \{1,\ \ldots,\ N\}$

$\boldsymbol{\lambda}_{\text{A}} = \boldsymbol{\Lambda}_{\text{A}}\boldsymbol{m}_{\mathsf{x}} + (1-\gamma)\boldsymbol{r}$ // $\boldsymbol{m}_{\text{A}} = \boldsymbol{m}_{\mathsf{x}} + \sigma_{\mathsf{x}}^2(1-\gamma)\boldsymbol{r}$

while *stopping criterion not met* **do**

- $\boldsymbol{m}_{\text{o,A}} = \sigma_n^2(\boldsymbol{A}^\top\boldsymbol{A} + \sigma_n^2\boldsymbol{\Lambda}_{\text{A}})^{-1}(\gamma\boldsymbol{r} + \boldsymbol{\lambda}_{\text{A}})$ // Linear Estimator
- $\boldsymbol{\Psi}_{\text{o,A}} = \textbf{diag}(\sigma_n^2(\boldsymbol{A}^\top\boldsymbol{A} + \sigma_n^2\boldsymbol{\Lambda}_{\text{A}})^{-1})$ // $\sigma_{\text{o,A},i}^2 = \Psi_{\text{o,A},ii}\ \forall i \in \{1,\ \ldots,\ N\}$
- $\boldsymbol{\Lambda}_{\text{o,A}} = \boldsymbol{\Psi}_{\text{o,A}}^{-1} = \text{fct}(\boldsymbol{\Lambda}_{\text{A}})$
- $\boldsymbol{\lambda}_{\text{o,A}} = \boldsymbol{\Psi}_{\text{o,A}}^{-1}\boldsymbol{m}_{\text{o,A}} = \text{fct}(\boldsymbol{\Lambda}_{\text{A}}, \gamma\boldsymbol{r} + \boldsymbol{\lambda}_{\text{A}})$
- $\boldsymbol{\Lambda}_{\text{B}} = \boldsymbol{\Lambda}_{\text{o,A}} - \boldsymbol{\Lambda}_{\text{A}}$ // $\sigma_{\text{B},i}^2 = (\frac{1}{\sigma_{\text{o,A},i}^2} - \frac{1}{\varsigma_{\text{A},i}^2})^{-1}$
- $\boldsymbol{\lambda}_{\text{B}} = \boldsymbol{\lambda}_{\text{o,A}} - \boldsymbol{\lambda}_{\text{A}}$ // $m_{\text{B},i} = \sigma_{\text{B},i}^2(\frac{m_{\text{o,A},i}}{\sigma_{\text{o,A},i}^2} - \frac{m_{\text{A},i}}{\varsigma_{\text{A},i}^2})$
- $\boldsymbol{m}_{\text{o,B}} = \text{E}_{\mathbf{x},\text{B}}\{\boldsymbol{g}_\lambda(\mathbf{x})\} = [m_{\text{o,B},1}, \ldots, m_{\text{o,B},N}]^\top$ // Non-Linear Estimator
- **for** $i \in \{1,\ \ldots,\ N\}$ **do**
  - $\varsigma_{\text{o,B},i}^2 = -2\text{E}_{\mathsf{x}_i,\text{B}}\{g_{\Lambda,i}(\mathsf{x}_i)\} - m_{\text{o,B},i}^2$
  - $\Lambda_{\text{o,B},i} = 1/\varsigma_{\text{o,B},i}^2 = \text{fct}(\Lambda_{\text{B},ii}, (1-\gamma)r_i + \lambda_{\text{B},i})$
  - $\lambda_{\text{o,B},i} = m_{\text{o,B},i}/\varsigma_{\text{o,B},i}^2 = \text{fct}(\Lambda_{\text{B},ii}, (1-\gamma)r_i + \lambda_{\text{B},i})$
  - $\Lambda_{\text{A},ii} = \Lambda_{\text{o,B},i} - \Lambda_{\text{B},ii}$ // $\varsigma_{\text{A},i}^2 = (\frac{1}{\varsigma_{\text{o,B},i}^2} - \frac{1}{\sigma_{\text{B},i}^2})^{-1}$
- $\boldsymbol{\lambda}_{\text{A}} = \boldsymbol{\lambda}_{\text{o,A}} - \boldsymbol{\lambda}_{\text{B}}$ // $m_{\text{A},i} = \varsigma_{\text{A},i}^2(\frac{m_{\text{o,B},i}}{\varsigma_{\text{o,B},i}^2} - \frac{m_{\text{B},i}}{\sigma_{\text{B},i}^2})$

Individual Variances—MMSE Improved

Algorithm 4: $[\boldsymbol{\lambda}_{\mathrm{o,B}}, \boldsymbol{\Lambda}_{\mathrm{o,B}}] = \texttt{ECindMMSE}(\boldsymbol{y},\ \boldsymbol{m}_{\mathsf{x}},\ \sigma_{\mathsf{x}}^2,\ \boldsymbol{A},\ \sigma_n^2)$

[...] // Initialization

while *stopping criterion not met* **do**

$\boldsymbol{m}_{\mathrm{o,A}} = \sigma_n^2(\boldsymbol{A}^\top\boldsymbol{A} + \sigma_n^2\boldsymbol{\Lambda}_{\mathrm{A}})^{-1}(\boldsymbol{r} + \boldsymbol{\lambda}_{\mathrm{A}})$ // Linear Estimator

$\boldsymbol{\Psi}_{\mathrm{o,A}} = \mathbf{diag}(\sigma_n^2(\boldsymbol{A}^\top\boldsymbol{A} + \sigma_n^2\boldsymbol{\Lambda}_{\mathrm{A}})^{-1})$ // $\sigma^2_{\mathrm{o,A},i} = \Psi_{\mathrm{o,A},ii}\ \forall i \in \{1, \ldots, N\}$

$\boldsymbol{\Lambda}_{\mathrm{o,A}} = \boldsymbol{\Psi}_{\mathrm{o,A}}^{-1} = \mathrm{fct}(\boldsymbol{\Lambda}_{\mathrm{A}})$

$\boldsymbol{\lambda}_{\mathrm{o,A}} = \boldsymbol{\Psi}_{\mathrm{o,A}}^{-1}\boldsymbol{m}_{\mathrm{o,A}} = \mathrm{fct}(\boldsymbol{\Lambda}_{\mathrm{A}},\ \boldsymbol{r} + \boldsymbol{\lambda}_{\mathrm{A}})$

$\boldsymbol{\Lambda}_{\mathrm{B}} = \boldsymbol{\Lambda}_{\mathrm{o,A}} - \boldsymbol{\Lambda}_{\mathrm{A}}$ // $\sigma^2_{\mathrm{B},i} = \left(\frac{1}{\sigma^2_{\mathrm{o,A},i}} - \frac{1}{\varsigma^2_{\mathrm{A},i}}\right)^{-1}$

$\boldsymbol{\lambda}_{\mathrm{B}} = \boldsymbol{\lambda}_{\mathrm{o,A}} - \boldsymbol{\lambda}_{\mathrm{A}}$ // $m_{\mathrm{B},i} = \sigma^2_{\mathrm{B},i}\left(\frac{m_{\mathrm{o,A},i}}{\sigma^2_{\mathrm{o,A},i}} - \frac{m_{\mathrm{A},i}}{\varsigma^2_{\mathrm{A},i}}\right)$

$\boldsymbol{m}_{\mathrm{o,B}} = \mathrm{E}_{\mathbf{x},\mathrm{B}}\{\boldsymbol{g}_{\lambda}(\mathbf{x})\} = [m_{\mathrm{o,B},1}, \ldots, m_{\mathrm{o,B},N}]^\top$ // Non-Linear Estimator

for $i \in \{1, \ldots, N\}$ **do**

$\varsigma^2_{\mathrm{o,B},i} = -2\mathrm{E}_{\mathsf{x}_i,\mathrm{B}}\{g_{\Lambda,i}(\mathsf{x}_i)\} - m^2_{\mathrm{o,B},i}$

$\sigma^2_{\mathrm{b},i} = \mathrm{E}_{\mathsf{m}_{\mathrm{B,i}}}\{\varsigma^2_{\mathrm{o,B},i}(\mathsf{m}_{\mathrm{B,i}})\}$

$\Lambda_{\mathrm{o,B},i} = 1/\varsigma^2_{\mathrm{o,B},i} = \mathrm{fct}(\Lambda_{\mathrm{B},i},\ \lambda_{\mathrm{B},i})$

$\lambda_{\mathrm{o,B},i} = m_{\mathrm{o,B},i}/\varsigma^2_{\mathrm{o,B},i} = \mathrm{fct}(\Lambda_{\mathrm{B},i},\ \lambda_{\mathrm{B},i})$

$\lambda_{\mathrm{b},i} = m_{\mathrm{o,B},i}/\sigma^2_{\mathrm{b},i}$

$\Lambda_{\mathrm{A},ii} = \left(\frac{1}{\Lambda_{\mathrm{o,B},i}} + \left(\frac{\Lambda_{\mathrm{B},ii}}{\Lambda_{\mathrm{b},i} - \Lambda_{\mathrm{B},ii}}\left(\frac{\lambda_{\mathrm{b},i}}{\Lambda_{\mathrm{b},i}} - \frac{\lambda_{\mathrm{B},i}}{\Lambda_{\mathrm{B},ii}}\right)\right)^2\right)^{-1}$ // $\varsigma^2_{\mathrm{A},i} = \ldots$

$\lambda_{\mathrm{A},i} = \frac{\Lambda_{\mathrm{A},ii}}{\Lambda_{\mathrm{b},i} - \Lambda_{\mathrm{B},ii}}(\lambda_{\mathrm{b},i} - \lambda_{\mathrm{B},i})$ // $m_{\mathrm{A},i} = \mathrm{fct}(\varsigma^2_{\mathrm{o,B},i}, \sigma^2_{\mathrm{B},i}, m_{\mathrm{o,B},i}, m_{\mathrm{B},i})$

Algorithm 5: $[\boldsymbol{\lambda}_{\mathrm{o,B}}, \boldsymbol{\Lambda}_{\mathrm{o,B}}] = \texttt{ECindMMSE\_var}(\boldsymbol{y},\, \boldsymbol{m}_{\mathsf{x}},\, \sigma_{\mathsf{x}}^2,\, \boldsymbol{A},\, \sigma_n^2,\, \gamma)$

[...] `// Initialization`

while *stopping criterion not met* **do**

- $\boldsymbol{m}_{\mathrm{o,A}} = \sigma_n^2(\boldsymbol{A}^\top\boldsymbol{A} + \sigma_n^2\boldsymbol{\Lambda}_{\mathrm{A}})^{-1}(\gamma\boldsymbol{r} + \boldsymbol{\lambda}_{\mathrm{A}})$ `// Linear Estimator`
- $\boldsymbol{\Psi}_{\mathrm{o,A}} = \mathbf{diag}(\sigma_n^2(\boldsymbol{A}^\top\boldsymbol{A} + \sigma_n^2\boldsymbol{\Lambda}_{\mathrm{A}})^{-1})$ `//` $\sigma_{\mathrm{o,A},i}^2 = \Psi_{\mathrm{o,A},ii}\ \forall i \in \{1, \ldots, N\}$
- $\boldsymbol{\Lambda}_{\mathrm{o,A}} = \boldsymbol{\Psi}_{\mathrm{o,A}}^{-1} = \mathrm{fct}(\boldsymbol{\Lambda}_{\mathrm{A}})$
- $\boldsymbol{\lambda}_{\mathrm{o,A}} = \boldsymbol{\Psi}_{\mathrm{o,A}}^{-1}\boldsymbol{m}_{\mathrm{o,A}} = \mathrm{fct}(\boldsymbol{\Lambda}_{\mathrm{A}}, \gamma\boldsymbol{r} + \boldsymbol{\lambda}_{\mathrm{A}})$
- $\boldsymbol{\Lambda}_{\mathrm{B}} = \boldsymbol{\Lambda}_{\mathrm{o,A}} - \boldsymbol{\Lambda}_{\mathrm{A}}$ `//` $\sigma_{\mathrm{B},i}^2 = (1/\sigma_{\mathrm{o,A},i}^2 - 1/\varsigma_{\mathrm{A},i}^2)^{-1}$
- $\boldsymbol{\lambda}_{\mathrm{B}} = \boldsymbol{\lambda}_{\mathrm{o,A}} - \boldsymbol{\lambda}_{\mathrm{A}}$ `//` $m_{\mathrm{B},i} = \sigma_{\mathrm{B},i}^2(m_{\mathrm{o,A},i}/\sigma_{\mathrm{o,A},i}^2 - m_{\mathrm{A},i}/\varsigma_{\mathrm{A},i}^2)$
- $\boldsymbol{\lambda}_{\mathrm{in,B}} = (1-\gamma)\boldsymbol{r} + \boldsymbol{\lambda}_{\mathrm{B}}$ `//` $\boldsymbol{m}_{\mathrm{in,B}} = \boldsymbol{\Lambda}_{\mathrm{B}}^{-1}\boldsymbol{\lambda}_{\mathrm{in,B}}$
- $\boldsymbol{m}_{\mathrm{o,B}} = \mathrm{E}_{\mathbf{x},\mathrm{B}}\{\boldsymbol{g}_{\lambda}(\mathbf{x})\} = [m_{\mathrm{o,B},1}, \ldots, m_{\mathrm{o,B},N}]^\top$ `// Non-Linear Estimator`
- **for** $i \in \{1, \ldots, N\}$ **do**
  - $\varsigma_{\mathrm{o,B},i}^2 = -2\mathrm{E}_{\mathsf{x}_i,\mathrm{B}}\{g_{\Lambda,i}(\mathsf{x}_i)\} - m_{\mathrm{o,B},i}^2$
  - $\sigma_{\mathrm{b},i}^2 = \mathrm{E}_{\mathsf{m}_{\mathrm{in,B,i}}}\{\varsigma_{\mathrm{o,B},i}^2(\mathsf{m}_{\mathrm{in,B,i}})\}$
  - $\Lambda_{\mathrm{o,B},i} = 1/\varsigma_{\mathrm{o,B},i}^2 = \mathrm{fct}(\Lambda_{\mathrm{B},i},\, (1-\gamma)r_i + \lambda_{\mathrm{B},i})$
  - $\lambda_{\mathrm{o,B},i} = m_{\mathrm{o,B},i}/\varsigma_{\mathrm{o,B},i}^2 = \mathrm{fct}(\Lambda_{\mathrm{B},i},\, (1-\gamma)r_i + \lambda_{\mathrm{B},i})$
  - $\lambda_{\mathrm{b},i} = m_{\mathrm{o,B},i}/\sigma_{\mathrm{b},i}^2$
  - $\Lambda_{\mathrm{A},ii} = \left(\frac{1}{\Lambda_{\mathrm{o,B},i}} + \left(\frac{\Lambda_{\mathrm{B},ii}}{\Lambda_{\mathrm{b},i} - \Lambda_{\mathrm{B},ii}}\left(\frac{\lambda_{\mathrm{b},i}}{\Lambda_{\mathrm{b},i}} - \frac{\lambda_{\mathrm{in,B},i}}{\Lambda_{\mathrm{B},ii}}\right)\right)^2\right)^{-1}$ `//` $\varsigma_{\mathrm{A},i}^2 = \ldots$
  - $\lambda_{\mathrm{A},i} = \frac{\Lambda_{\mathrm{A},ii}}{\Lambda_{\mathrm{b},i} - \Lambda_{\mathrm{B},ii}}(\lambda_{\mathrm{b},i} - \lambda_{\mathrm{in,B},i}) + (1-\gamma)r_i$ `//` $m_{\mathrm{A},i} = \ldots$