Blind Deconvolution of Sparse But Filtered Pulses With Linear State Space Models

Nour Zalmaï, Hampus Malmberg, and Hans-Andrea Loeliger

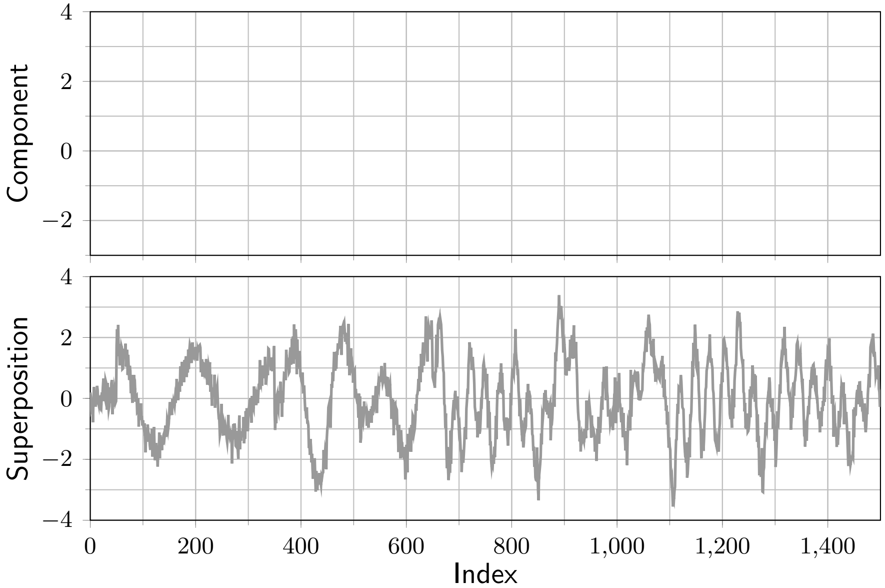
ETH Zurich
{zalmai, malmberg, loeliger}@isi.ee.ethz.ch

ICASSP Shanghai, 25 March 2016

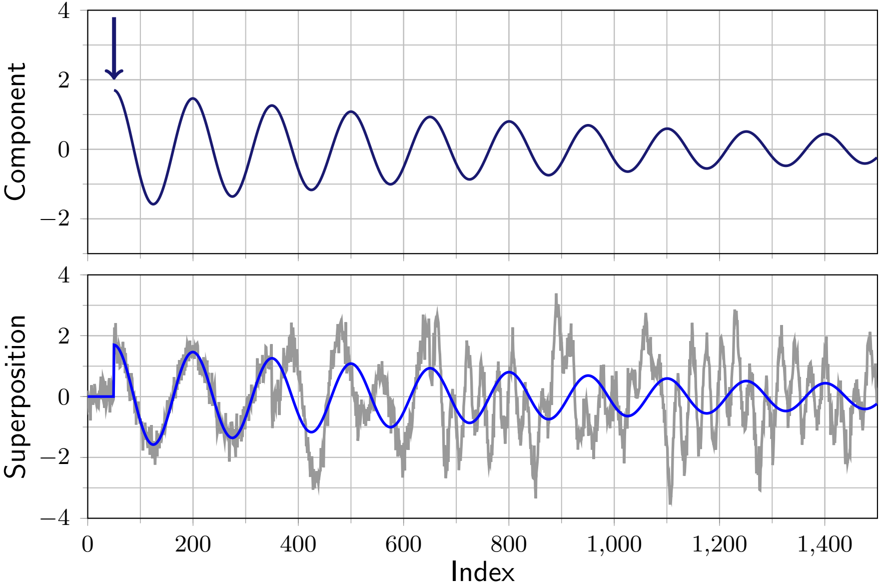
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Signal and Information Processing Laboratory
Institut für Signal- und Informationsverarbeitung

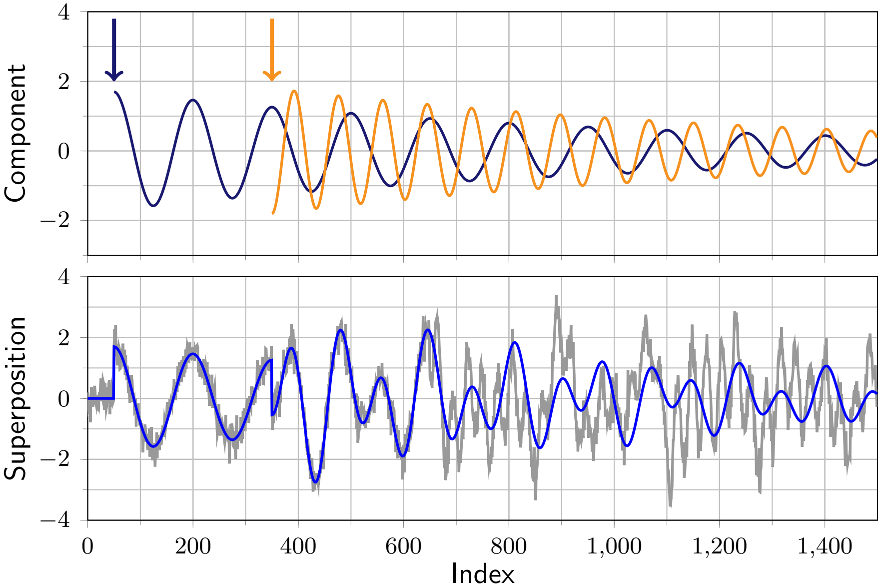
Introduction



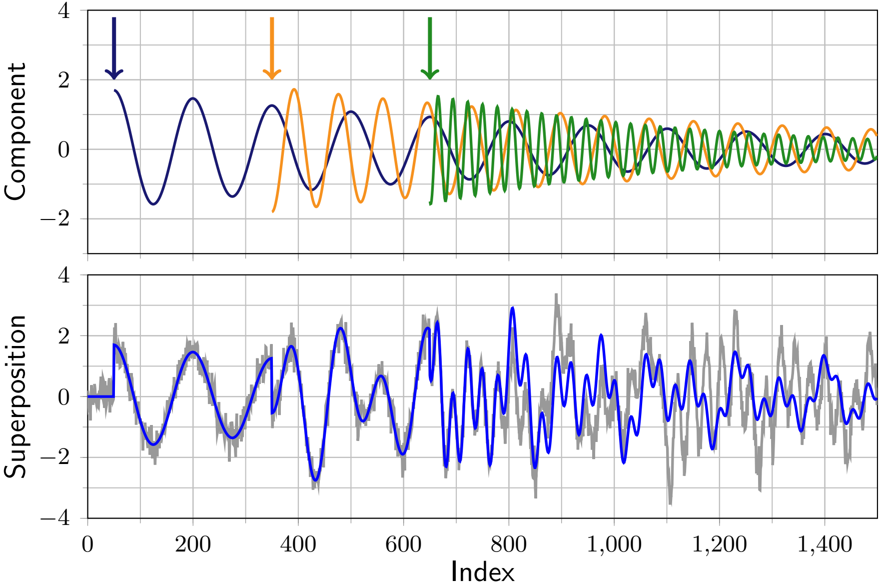
Introduction



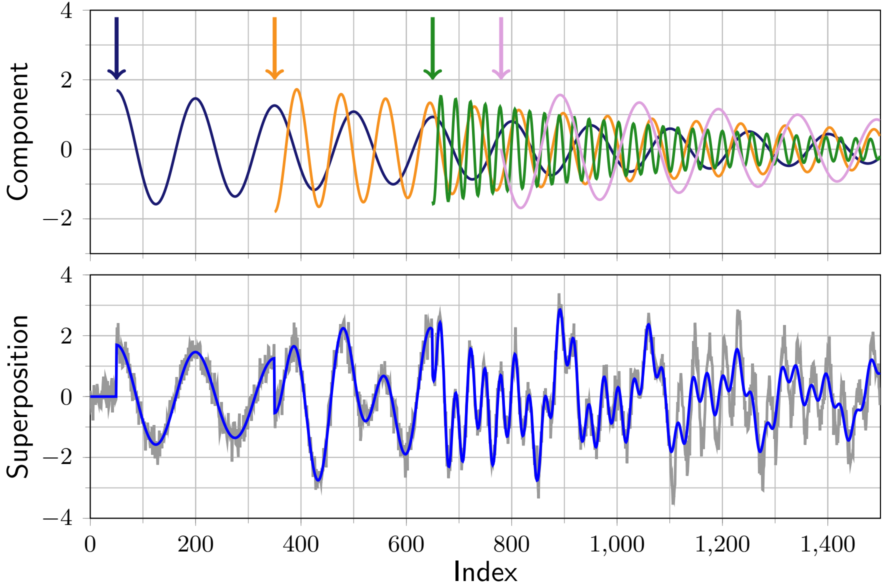
Introduction



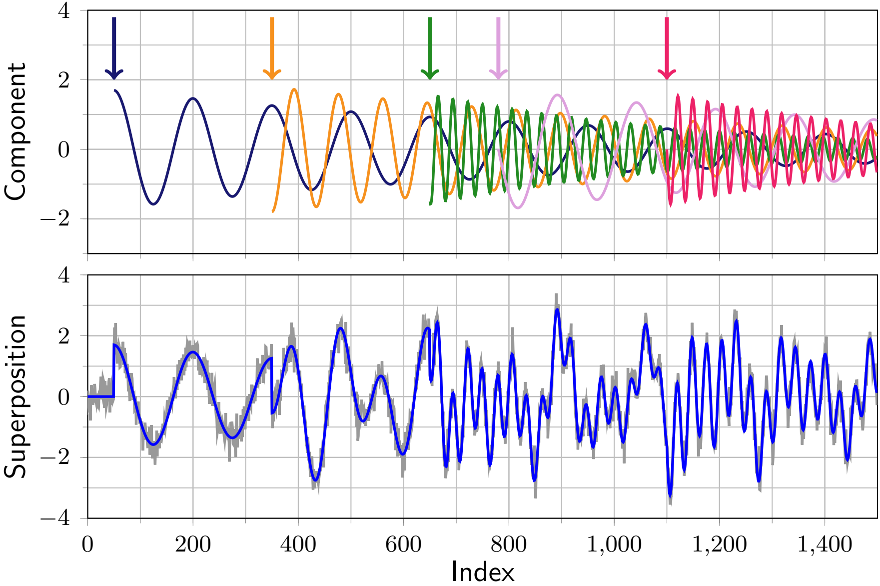
Introduction



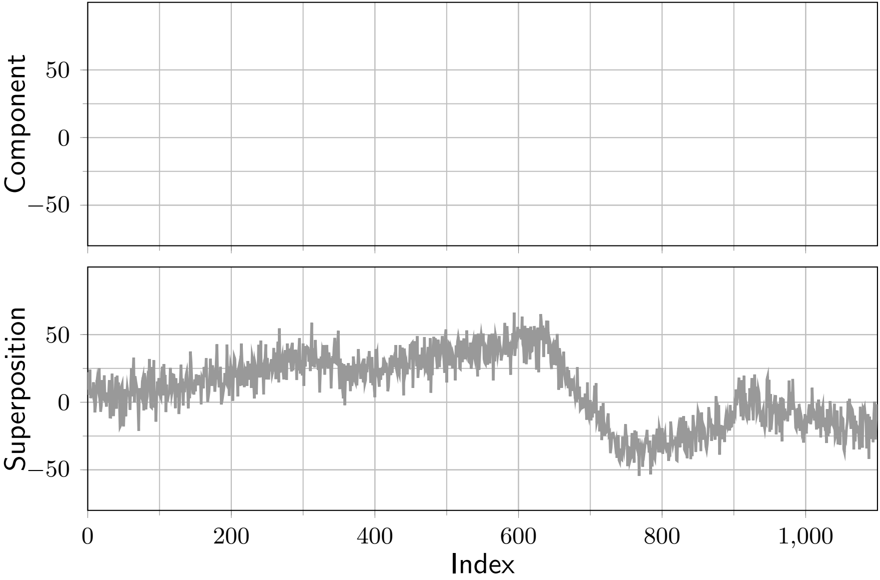
Introduction



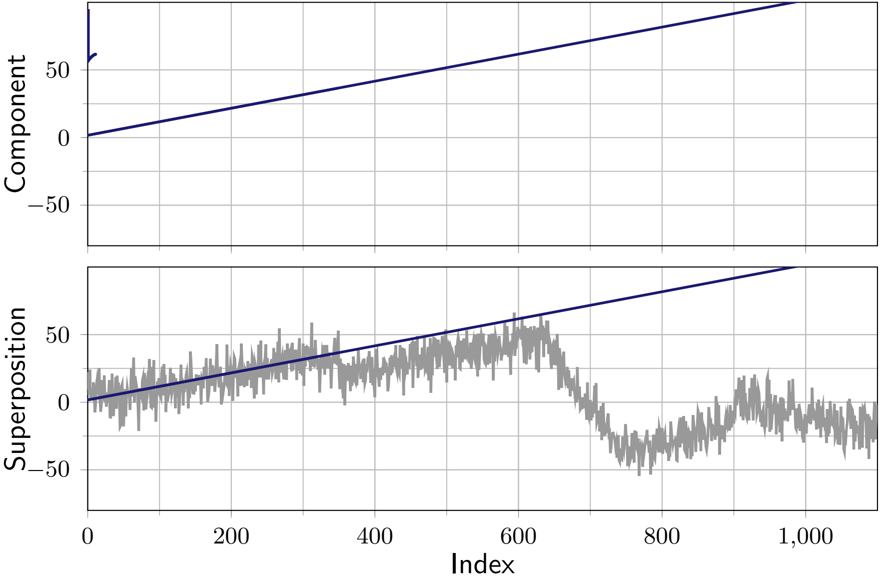
Introduction



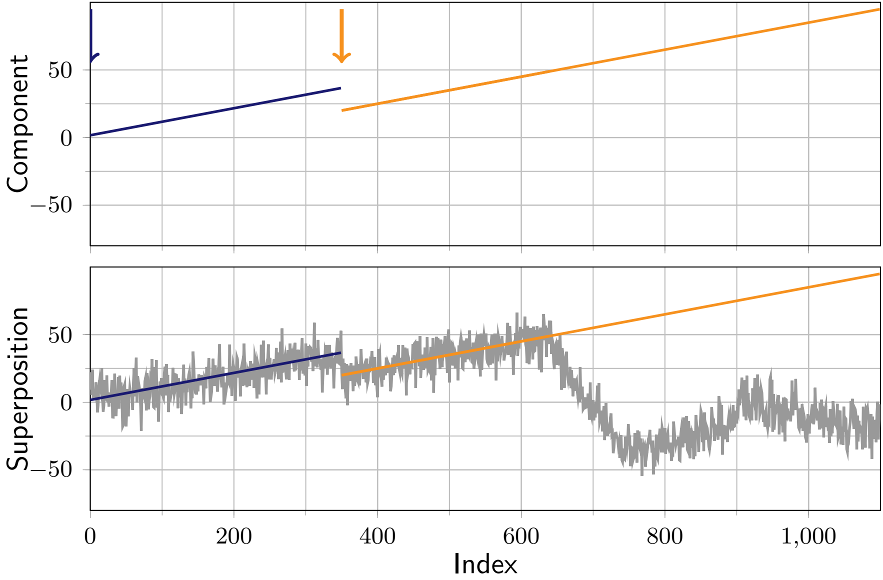
Introduction



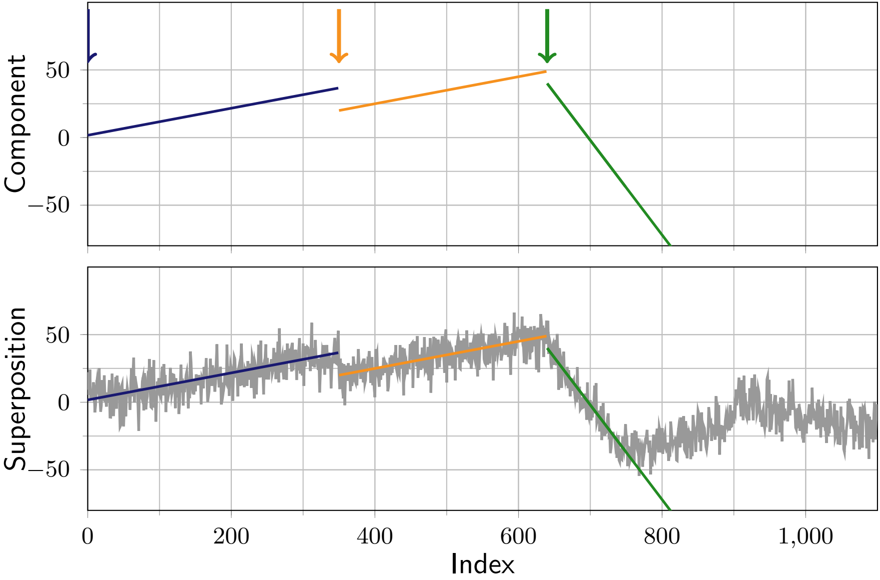
Introduction



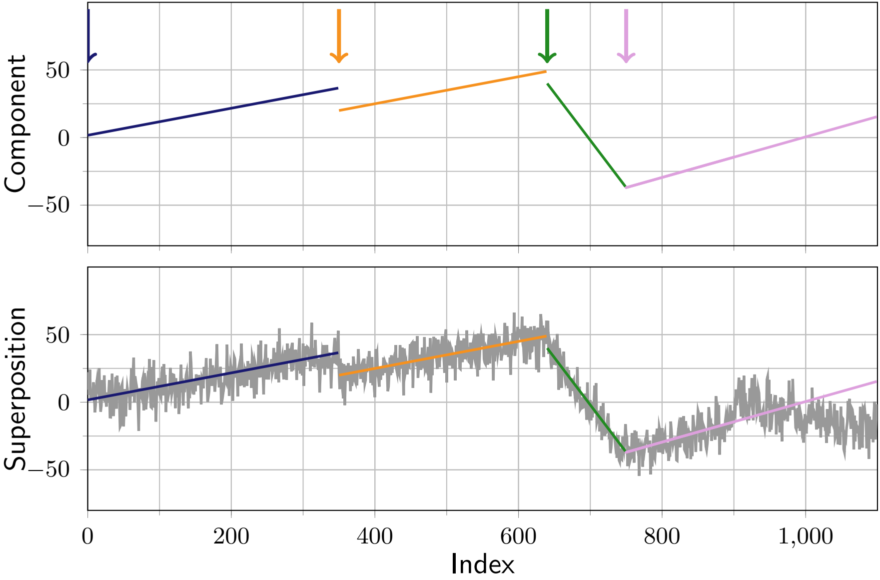
Introduction



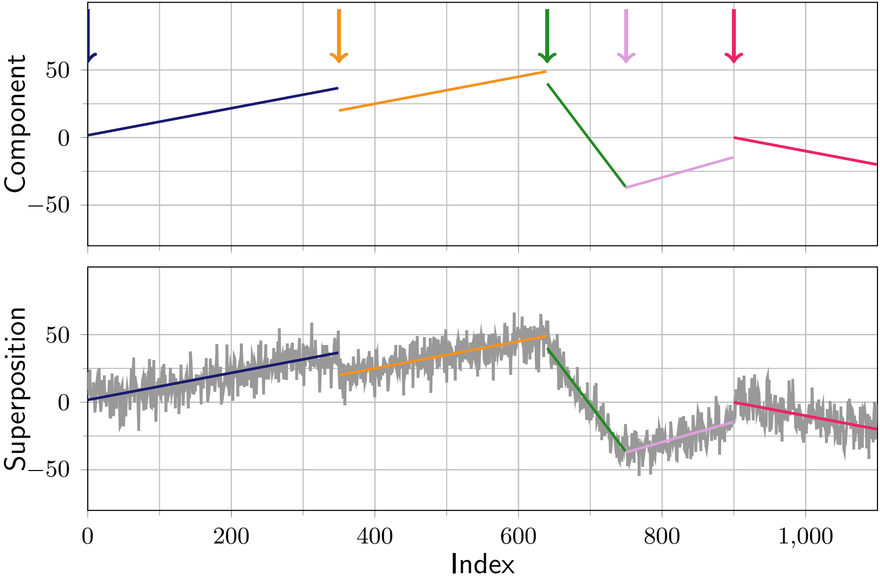
Introduction



Introduction



Introduction



Linear State Space Model

Goal: "Explain" observed signal $(y_1, \ldots, y_K) \in \mathbb{R}^K$ as output of a linear state space model (LSSM) of order n

$$\begin{cases} X_k & = AX_{k-1} + B_k u_k + \epsilon_k \\ y_k & = CX_k + Z_k \end{cases}$$

with (unknown)

- input signal $u = (u_1, \ldots u_K) \in \mathbb{R}^K$
- input vectors $B_1, \ldots, B_K \in \mathbb{R}^{n \times 1}$
- observation noise $Z_k \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_Z^2)$
- state noise $\epsilon_k \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\epsilon^2 I)$
- LSSM parameters: $A \in \mathbb{R}^{n \times n}$ and $C \in \mathbb{R}^{1 \times n}$

Linear State Space Model

Goal: "Explain" observed signal $(y_1, \dots, y_K) \in \mathbb{R}^K$ as output of a linear state space model (LSSM) of order n

$$\begin{cases} X_k &= AX_{k-1} + B_k u_k + \epsilon_k \\ y_k &= CX_k + Z_k \end{cases}$$

$\Rightarrow$ require $u = (u_1, \dots u_K)$ to be sparse

Tipping[1], automatic relevance determination (ARD)

$U_k \sim \mathcal{N}(0, \sigma^2_{U_k})$, with $\sigma^2_{U_k}$ estimated by maximum likelihood

$$p(y|\theta) = \int \underbrace{p(y|\theta, u)}_{\text{LSSM}} \ \underbrace{p(u|\theta)}_{\text{ARD prior}} \ \mathrm{d}u$$

[1] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *Journal of Machine Learning Research*, vol. 1, p. 211–244, 2001

Linear State Space Model

Goal: "Explain" observed signal $(y_1, \ldots, y_K) \in \mathbb{R}^K$ as output of a linear state space model (LSSM) of order n

$$\begin{cases} X_k &= AX_{k-1} + B_k u_k + \epsilon_k \\ y_k &= CX_k + Z_k \end{cases}$$

Why is it sparse?

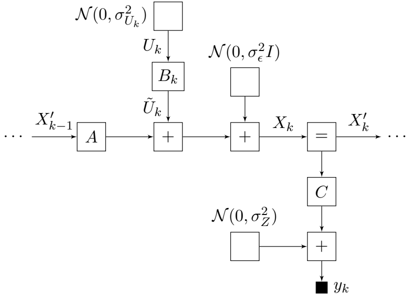
- look at local maxima

Linear State Space Model

Goal: "Explain" observed signal $(y_1, \ldots, y_K) \in \mathbb{R}^K$ as output of a linear state space model (LSSM) of order n

$$\begin{cases} X_k &= AX_{k-1} + B_k u_k + \epsilon_k \\ y_k &= CX_k + Z_k \end{cases}$$

Factor graph representation



Why is it sparse?

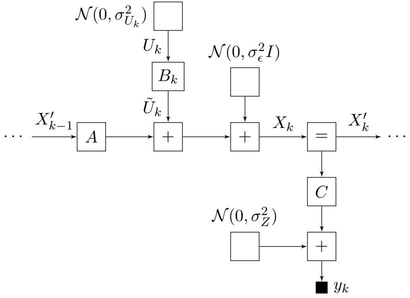
- look at local maxima

Linear State Space Model

Goal: "Explain" observed signal $(y_1, \ldots, y_K) \in \mathbb{R}^K$ as output of a linear state space model (LSSM) of order n

$$\begin{cases} X_k &= AX_{k-1} + B_k u_k + \epsilon_k \\ y_k &= CX_k + Z_k \end{cases}$$

Factor graph representation



Why is it sparse?

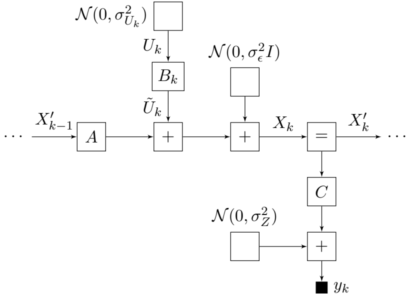
- look at local maxima
- local optimality of $(\sigma^2_{U_k}, B_k)$

Linear State Space Model

Goal: "Explain" observed signal $(y_1, \ldots, y_K) \in \mathbb{R}^K$ as output of a linear state space model (LSSM) of order n

$$\begin{cases} X_k &= AX_{k-1} + B_k u_k + \epsilon_k \\ y_k &= CX_k + Z_k \end{cases}$$

Factor graph representation



Why is it sparse?

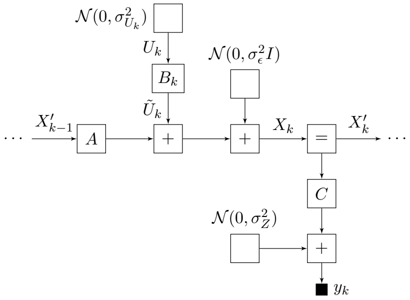
- look at local maxima
- local optimality of $(\sigma^2_{U_k}, B_k)$
- Gaussian message passing

Linear State Space Model

Goal: "Explain" observed signal $(y_1, \ldots, y_K) \in \mathbb{R}^K$ as output of a linear state space model (LSSM) of order n

$$\begin{cases} X_k &= AX_{k-1} + B_k u_k + \epsilon_k \\ y_k &= CX_k + Z_k \end{cases}$$

Factor graph representation



Why is it sparse?

- look at local maxima
- local optimality of $(\sigma^2_{U_k}, B_k)$
- Gaussian message passing

  In particular, on edge $\tilde{U}_k$:

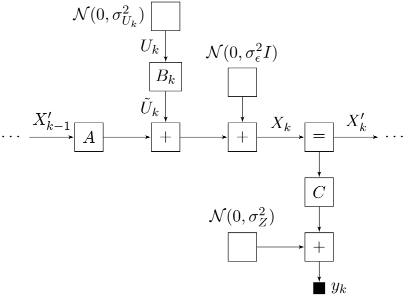
$$\overleftarrow{m}_{\tilde{U}_k} = \overleftarrow{m}_{X_k} - \overrightarrow{m}_{X_k}$$
$$\overleftarrow{V}_{\tilde{U}_k} = A\overrightarrow{V}_{X'_k}A^\mathsf{T} + \overleftarrow{V}_{X_k} + \sigma^2_\epsilon I$$

Sparsity Behavior

Marginal log-likelihood with respect to $(\sigma^2_{U_k}, B_k)$

$$2 \ln p(y|\hat{\theta}_k, B_k, \sigma^2_{U_k}) \propto \frac{\sigma^2_{U_k} (B_k^\mathsf{T} \overleftarrow{V}^{-1}_{\tilde{U}_k} \overleftarrow{m}_{\tilde{U}_k})^2}{1 + \sigma^2_{U_k} B_k^\mathsf{T} \overleftarrow{V}^{-1}_{\tilde{U}_k} B_k} - \ln(1 + \sigma^2_{U_k} B_k^\mathsf{T} \overleftarrow{V}^{-1}_{\tilde{U}_k} B_k)$$

Factor graph representation



Why is it sparse?

- look at local maxima
- local optimality of $(\sigma^2_{U_k}, B_k)$
- Gaussian message passing

  In particular, on edge $\tilde{U}_k$:

$$\overleftarrow{m}_{\tilde{U}_k} = \overleftarrow{m}_{X_k} - \overrightarrow{m}_{X_k}$$

$$\overleftarrow{V}_{\tilde{U}_k} = A \overrightarrow{V}_{X'_k} A^\mathsf{T} + \overleftarrow{V}_{X_k} + \sigma^2_\epsilon I$$

Sparsity Behavior

Marginal log-likelihood with respect to $(\sigma^2_{U_k}, B_k)$

$$2 \ln p(y|\hat{\theta}_k, B_k, \sigma^2_{U_k}) \propto \frac{\sigma^2_{U_k} (B_k^\mathsf{T} \overleftarrow{V}_{\tilde{U}_k}^{-1} \overleftarrow{m}_{\tilde{U}_k})^2}{1 + \sigma^2_{U_k} B_k^\mathsf{T} \overleftarrow{V}_{\tilde{U}_k}^{-1} B_k} - \ln(1 + \sigma^2_{U_k} B_k^\mathsf{T} \overleftarrow{V}_{\tilde{U}_k}^{-1} B_k)$$

Lemma : Local Optimality Condition

If $\overleftarrow{m}_{\tilde{U}_k}^\mathsf{T} \overleftarrow{V}_{\tilde{U}_k}^{-1} \overleftarrow{m}_{\tilde{U}_k} > 1$, *then*

$$\hat{\sigma}^2_{U_k} = \left(1 - \frac{1}{\overleftarrow{m}_{\tilde{U}_k}^\mathsf{T} \overleftarrow{V}_{\tilde{U}_k}^{-1} \overleftarrow{m}_{\tilde{U}_k}}\right) \|\overleftarrow{m}_{\tilde{U}_k}\|^2 > 0 \text{ and } \hat{B}_k = \frac{\overleftarrow{m}_{\tilde{U}_k}}{\|\overleftarrow{m}_{\tilde{U}_k}\|} .$$

Else, $\hat{\sigma}^2_{U_k} = 0$ *and* $\hat{B}_k$ *is any vector such that* $\|\hat{B}_k\| = 1$.

Sparsity Behavior

An input (i.e., $\hat{\sigma}^2_{U_k} > 0$) is introduced only to compensate a discrepancy $\overleftarrow{m}_{\tilde{U}_k} = \overrightarrow{m}_{X_k} - \overleftarrow{m}_{X_k}$ between $\overleftarrow{m}_{X_k}$ and $\overrightarrow{m}_{X_k}$ $\Rightarrow$ Awareness of input at k

Lemma : Local Optimality Condition

If $\overleftarrow{m}^{\mathsf{T}}_{\tilde{U}_k} \overleftarrow{V}^{-1}_{\tilde{U}_k} \overleftarrow{m}_{\tilde{U}_k} > 1$, *then*

$$\hat{\sigma}^2_{U_k} = \left(1 - \frac{1}{\overleftarrow{m}^{\mathsf{T}}_{\tilde{U}_k} \overleftarrow{V}^{-1}_{\tilde{U}_k} \overleftarrow{m}_{\tilde{U}_k}}\right) \|\overleftarrow{m}_{\tilde{U}_k}\|^2 > 0 \text{ and } \hat{B}_k = \frac{\overleftarrow{m}_{\tilde{U}_k}}{\|\overleftarrow{m}_{\tilde{U}_k}\|} .$$

Else, $\hat{\sigma}^2_{U_k} = 0$ *and* $\hat{B}_k$ *is any vector such that* $\|\hat{B}_k\| = 1$.

Likelihood Function

Goal: "Explain" observed signal $(y_1, \ldots, y_K) \in \mathbb{R}^K$ as output of a linear state space model (LSSM) of order n

$$\begin{cases} X_k &= AX_{k-1} + B_k u_k + \epsilon_k \\ y_k &= CX_k + Z_k \end{cases}$$

Likelihood Function: Integral form, X and U as hidden variables

$$p(y|\theta) = \int\int \underbrace{\prod_{k=1}^{K} p(y_k|x_k,\theta)p(x_k|x_{k-1},u_k,\theta)p(u_k|\theta)}_{p(y,x,u|\theta)} \, \mathrm{d}x \, \mathrm{d}u$$

with $\theta = (C, A, \sigma_Z^2, \sigma_\epsilon^2, B_1, \ldots, B_K, \sigma_{U_1}^2, \ldots, \sigma_{U_K}^2)$.

Likelihood Function

Likelihood Function: Integral form, X and U as hidden variables

$$p(y|\theta) = \int \int \underbrace{\prod_{k=1}^{K} p(y_k|x_k,\theta)p(x_k|x_{k-1},u_k,\theta)p(u_k|\theta)}_{p(y,x,u|\theta)} \, \mathrm{d}x \, \mathrm{d}u$$

with $\theta = (C, A, \sigma_Z^2, \sigma_\epsilon^2, B_1, \ldots, B_K, \sigma_{U_1}^2, \ldots, \sigma_{U_K}^2)$.

- maximize $p(y|\theta)$ for both input estimation and system identification
- use expectation maximization (EM) for a joint estimation

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \, \mathbb{E}\left[\ln p(y, U, X|\theta)\right] \ .$$

- for fixed $\hat{\theta}$, $p(y, x, u|\hat{\theta})$ is Gaussian
- Gaussian message passing for computing required expectations

EM Update (X and U as hidden variables)

Maximization step: for each parameter,
simple closed-form expression or minimization of a quadratic form

- $\hat{\sigma}^2_{U_k} = \mathbb{E}\left[U_k^2\right]$, for $k \in \{1, \ldots, K\}$
- $\hat{B}_k = \dfrac{\mathbb{E}[U_k X_k] - \hat{A}\,\mathbb{E}[U_k X_{k-1}]}{\mathbb{E}\left[U_k^2\right]}$, for $k \in \{1, \ldots, K\}$
- $\hat{A} = \underset{A}{\operatorname{argmin}} \operatorname{Tr}\left(A V_A A^{\mathsf{T}} - 2A\xi_A\right)$
- $\hat{\sigma}^2_{\epsilon}$, $\hat{C}$, $\hat{\sigma}^2_Z$

Required computations:
$\mathbb{E}\left[U_k^2\right]$, $\mathbb{E}[U_k X_{k-1}]$, $\mathbb{E}[U_k X_k]$, $\mathbb{E}\left[X_k X_k^{\mathsf{T}}\right]$, and $\mathbb{E}\left[X_{k-1} X_k^{\mathsf{T}}\right]$
$\Rightarrow$ efficiently computed using Gaussian message passing

Remarks

- the noise variance σ_Z^2 controls the sparsity level (should be fixed a priori)
- the state noise variance σ_ϵ^2 is the LSSM mismatch and should converge to a low value
- many elements of $(\sigma_{U_k} \cdot \|B_k\|)_{k\in\{1,\ldots,K\}}$ converge to zero but will typically not become exact zeros.
  To obtain exact zeros, use marginal likelihood update:
  $\overleftarrow{m}_{\tilde{U}_k}^{\mathsf{T}} \overleftarrow{V}_{\tilde{U}_k}^{-1} \overleftarrow{m}_{\tilde{U}_k} \leq 1 \Rightarrow \hat{\sigma}_{U_k}^2 = 0$

Likelihood Function

Goal: "Explain" observed signal $(y_1, \ldots, y_K) \in \mathbb{R}^K$ as output of a linear state space model (LSSM) of order n

$$\begin{cases} X_k &= AX_{k-1} + B_k u_k + \epsilon_k \\ y_k &= CX_k + Z_k \end{cases}$$

Special case: matrix A known

Likelihood Function: Integral form, X as hidden variable

$$p(y|\theta) = \int \underbrace{\prod_{k=1}^{K} p(y_k|x_k, \theta)p(x_k|x_{k-1}, \theta)}_{p(y,x|\theta)} \, \mathrm{d}x$$

with $\theta = (C, \sigma_Z^2, \sigma_\epsilon^2, B_1, \ldots, B_K, \sigma_{U_1}^2, \ldots, \sigma_{U_K}^2)$.

Likelihood Function

Special case: matrix A known

Likelihood Function: Integral form, X as hidden variable

$$p(y|\theta) = \int \underbrace{\prod_{k=1}^{K} p(y_k|x_k, \theta) p(x_k|x_{k-1}, \theta)}_{p(y,x|\theta)} \, \mathrm{d}x$$

with $\theta = (C, \sigma_Z^2, \sigma_\epsilon^2, B_1, \ldots, B_K, \sigma_{U_1}^2, \ldots, \sigma_{U_K}^2)$.

Use expectation maximization (EM) algorithm for a joint estimation of θ

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \, \mathbb{E} \left[\ln p(y, X|\theta) \right] \ .$$

EM Update (X as hidden variable)

- $\hat{B}_k$: eigenvector of $\mathbb{E}\left[(X_k - AX_{k-1})(X_k - AX_{k-1})^\mathsf{T}\right]$ corresponding to the maximum eigenvalue λ_k

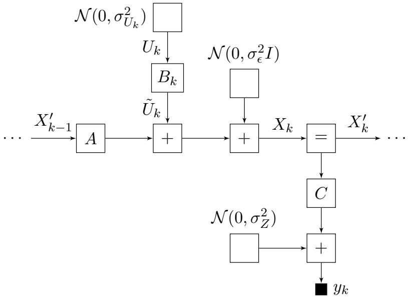
- $\hat{\sigma}^2_{U_k} = \max\left(0, \lambda_k - \hat{\sigma}^2_\epsilon\right)$, for $k \in \{1, \ldots, K\}$
  $\Rightarrow$ can create exact zeros

- $\hat{\sigma}^2_\epsilon = \underset{\sigma_\epsilon}{\operatorname{argmin}} \dfrac{M_A}{\sigma^2_\epsilon} + nK \ln(\sigma^2_\epsilon) + \displaystyle\sum_{\lambda_k > \sigma^2_\epsilon} -\frac{\lambda_k - \sigma^2_\epsilon}{\sigma^2_\epsilon} + \ln\left(\frac{\lambda_k}{\sigma^2_\epsilon}\right)$

Gaussian Message Passing

Quantities to be computed:

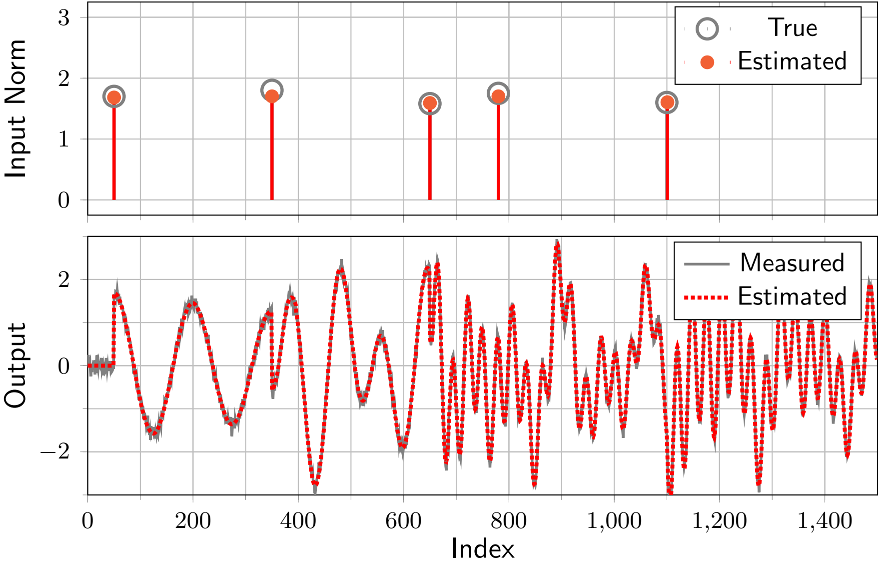
$$\mathbb{E}\left[U_k^2\right],\ \mathbb{E}[U_k X_{k-1}],\ \mathbb{E}[U_k X_k],\ \mathbb{E}\left[X_k X_k^\mathsf{T}\right],\ \mathbb{E}\left[X_{k-1} X_k^\mathsf{T}\right]$$



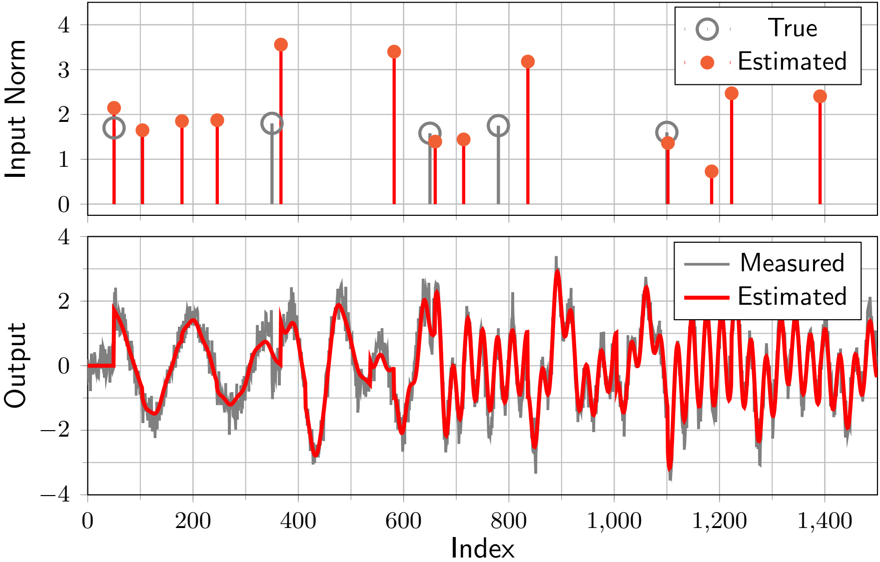
- efficient (matrix multiplications)
- stable while σ_ϵ^2 and any $\sigma_{U_k}^2$ tend to zero
- suitable choice: modified Bryson-Frazier smoother[1]
- avoid matrix inversions

[1] L. Bruderer, H. Malmberg, and H.-A. Loeliger, "Deconvolution of weakly-sparse signals and dynamical-system identification by Gaussian message passing," *Proc. 2015 IEEE Int. Symp. on Information Theory,* Hong Kong, June 14–19 2015, pp. 326-330.

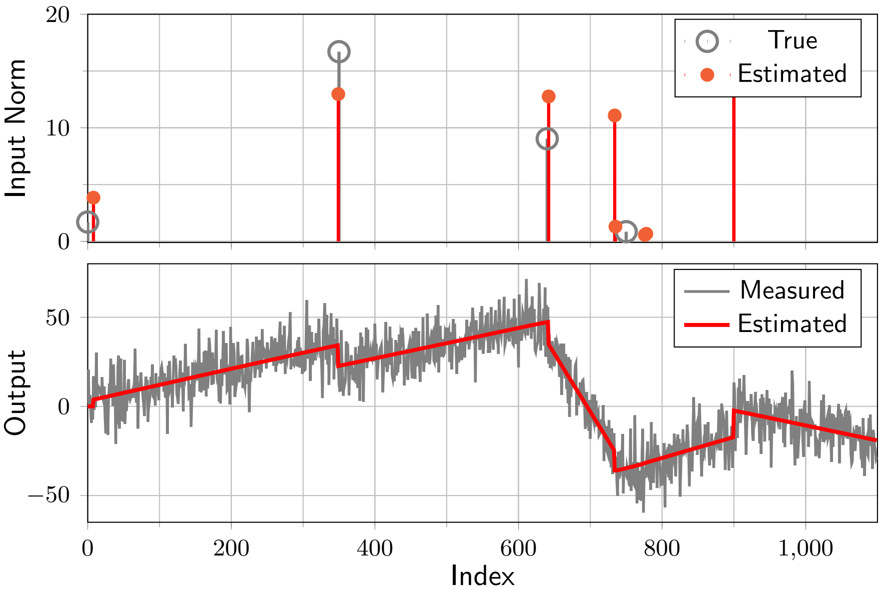
Overlapping Decaying Sines (high SNR)



Overlapping Decaying Sines (low SNR)



Lines with Jumps (A fixed)



Conclusion

- efficient algorithm for sparse input estimation and system identification
- ARD technique copes well with linear state space models
- robust joint estimation
- flexible framework