Regularized State estimation and Parameter learning via augmented Lagrangian Kalman smoother method

Rui Gao, Filip Tronarp, Zheng Zhao, Simo Särkkä
Dept. Electrical Engineering and Automation, Aalto University, Finland

Motivation

State estimation and parameter learning in linear dynamic systems is an important problem that arises in many applications.

(i) Memory requirements and computation of learning algorithms are generally prohibitive when the dimensionality of state or parameters are large;
(ii) It is hard to promote sparsity using probabilistic approaches.

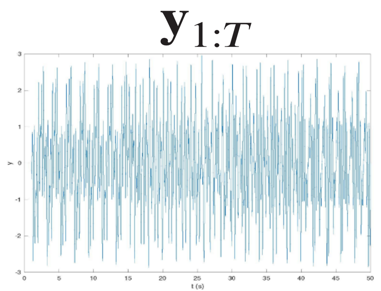
Contribution

We present an efficient method for regularized state estimation and parameter learning from noisy measurements. The main contributions are to:

(i) Build the generalized L_1-regularized model for jointly estimating of the state and learning of the parameters;
(ii) Propose a new augmented Lagrangian Kalman smoother method.

The overall architecture

$\mathbf{y}_{1:T}$



↓

Problem Formulation

$$\mathbf{x}_t = \mathbf{A}_t \mathbf{x}_{t-1} + \mathbf{q}_t$$
$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{r}_t$$

Linear system

+

$$||\Omega_t \mathbf{x}_t||_1$$

Sparse constraint

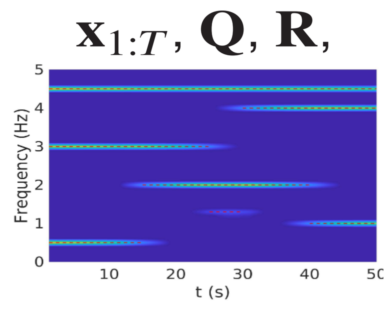
→

Objective Function

$$\min_{\mathbf{x}_{1:T},\mathbf{Q},\mathbf{R}} \sum_{t=1}^{T} \frac{1}{2}||\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t||^2_{\mathbf{R}^{-1}} + \frac{1}{2}\sum_{t=2}^{T}||\mathbf{x}_t - \mathbf{A}_t\mathbf{x}_{t-1}||^2_{\mathbf{Q}^{-1}} + \frac{1}{2}||\mathbf{x}_1 - \mathbf{m}_1||^2_{\mathbf{Q}^{-1}} + \lambda||\Omega_t\mathbf{x}_t||_1 + \frac{T}{2}\log|\mathbf{R}| + \frac{T}{2}\log|\mathbf{Q}|$$

↑

$\mathbf{x}_{1:T}$, $\mathbf{Q}$, $\mathbf{R}$,



↓

Proposed Method

Kalman smoother ↘
Augmented Lagrangian Kalman Smoother (ALKS)
Variable splitting ↗

→

New Linear System

$$\mathbf{x}_t = \mathbf{A}_t \mathbf{x}_{t-1} + \mathbf{q}_t$$
$$\mathbf{y}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{r}_t$$
$$\mathbf{z}_t = \Omega_t \mathbf{x}_t + \sigma_t$$

Augmented Lagrangian Kalman smoother (ALKS)

The main idea here is to build the augmented Lagrangian function and to decompose the objective into an iterative sequence of much easier subproblems. Based on our article [1], the subproblems are solved by applying Kalman smoother.

(i) **Build the augmented Lagrangian function**

$$\mathcal{L}(\mathbf{x}_{1:T}, \mathbf{Q}, \mathbf{R}, \mathbf{w}_{1:T}; \boldsymbol{\eta}_{1:T}) = \frac{1}{2}\sum_{t=1}^{T}\|\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t\|^2_{\mathbf{R}^{-1}} + \frac{1}{2}\sum_{t=2}^{T}\|\mathbf{x}_t - \mathbf{A}_t\mathbf{x}_{t-1}\|^2_{\mathbf{Q}^{-1}} + \lambda\sum_{t=1}^{T}\|\mathbf{w}_t\|_1 + \frac{1}{2}\|\mathbf{x}_1 - \mathbf{m}_1\|^2_{\mathbf{P}_1^{-1}} + \frac{T}{2}\log|\mathbf{Q}| + \frac{T}{2}\log|\mathbf{R}| + \frac{\rho}{2}\sum_{t=1}^{T}\|\boldsymbol{\Omega}_t\mathbf{x}_t - \mathbf{w}_t + \boldsymbol{\eta}_t\|^2 \quad (1)$$

(ii) **Solving the $\mathbf{x}_{1:T}$-subproblem**

$$\min_{\mathbf{x}_{1:T}} \frac{1}{2}\sum_{t=1}^{T}\|\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t\|^2_{\mathbf{R}^{-1}} + \frac{1}{2}\sum_{t=2}^{T}\|\mathbf{x}_t - \mathbf{A}_t\mathbf{x}_{t-1}\|^2_{\mathbf{Q}^{-1}} + \frac{1}{2}\sum_{t=1}^{T}\|\mathbf{z}_t - \boldsymbol{\Omega}_t\mathbf{x}_t\|^2_{\boldsymbol{\Sigma}^{-1}} + \frac{1}{2}\|\mathbf{x}_1 - \mathbf{m}_1\|^2_{\mathbf{P}_1^{-1}} \quad (2)$$

which is solved by Kalman smoother (see [2] for details).

(iii) **Solving the $\mathbf{Q}$, $\mathbf{R}$-subproblems**

$$\mathbf{Q}^{(k+1)} = \frac{1}{T-1}\sum_{t=2}^{T}(\mathbf{x}_t - \mathbf{A}_t\mathbf{x}_{t-1})(\mathbf{x}_t - \mathbf{A}_t\mathbf{x}_{t-1})^\top \quad (3a)$$

$$\mathbf{R}^{(k+1)} = \frac{1}{T}\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t)(\mathbf{y}_t - \mathbf{H}_t\mathbf{x}_t)^\top \quad (3b)$$

(iv) **Solving the $\mathbf{w}_{1:T}$-subproblem**

$$\mathbf{w}_t^{(k+1)} = \mathrm{sgn}(\boldsymbol{\Omega}_t\mathbf{x}_t^{(k+1)} + \boldsymbol{\eta}_t^{(k)}) \circ \max(|\boldsymbol{\Omega}_t\mathbf{x}_t^{(k+1)} + \boldsymbol{\eta}_t^{(k)}| - \lambda/\rho, 0) \quad (4)$$

Experimental Results

(i) **Simulated estimation scenario**

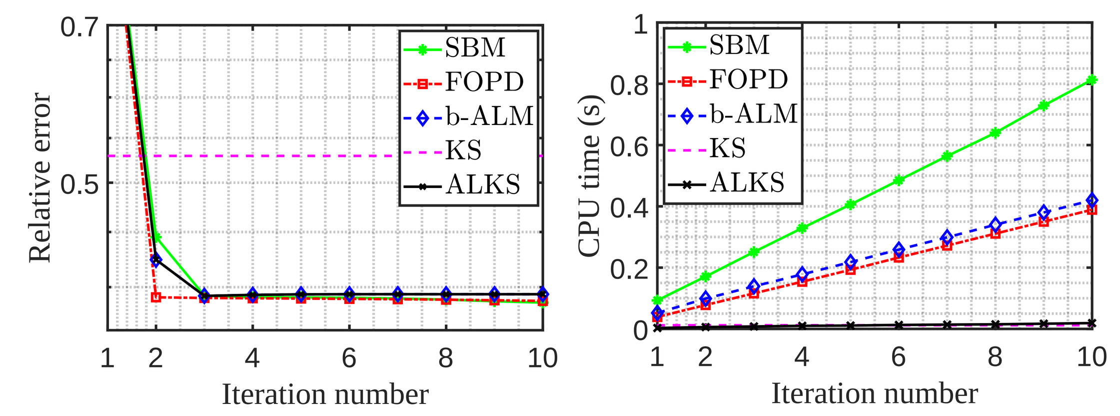


Table: Average CPU time for large-scale datasets

T	10^3	10^4	10^6	10^6	10^7
SBM	421	86286	-	-	-
FOPD	199	42643	-	-	-
b-ALM	302	74286	-	-	-
KS	1.1	11.6	100	535	6520
ALKS	1.56	13.3	134	767	12714

(ii) **Sparse spectro-temporal estimation**

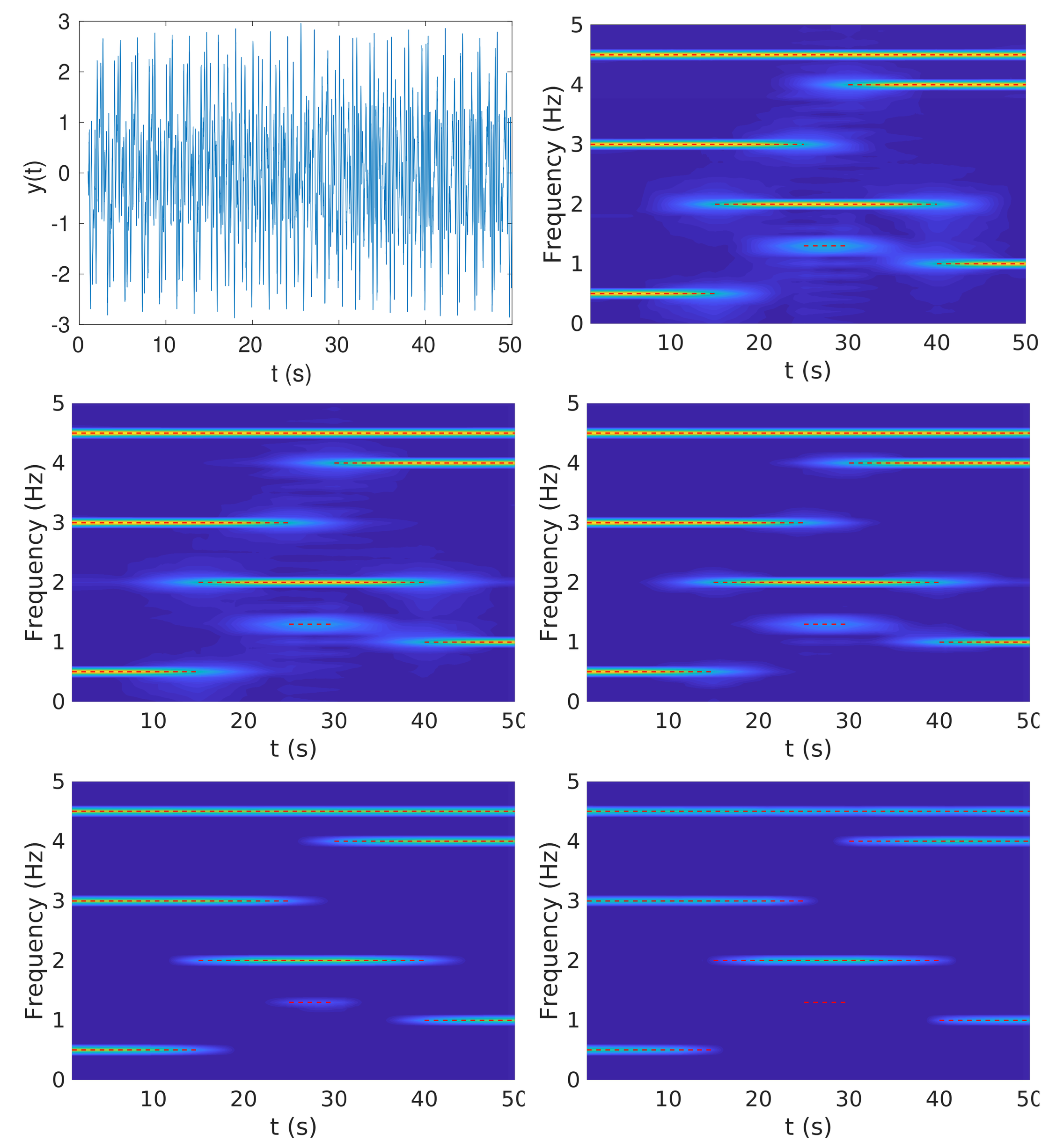


Figure: From left-top to right-bottom, comparison of the spectro-temporal estimation of noisy sinusoidal data, estimation results by using KS, ALKS with $\lambda = 0.01$, $\lambda = 0.2$, $\lambda = 1$, and $\lambda = 0.01$.

(iii) **Automatic music transcription**

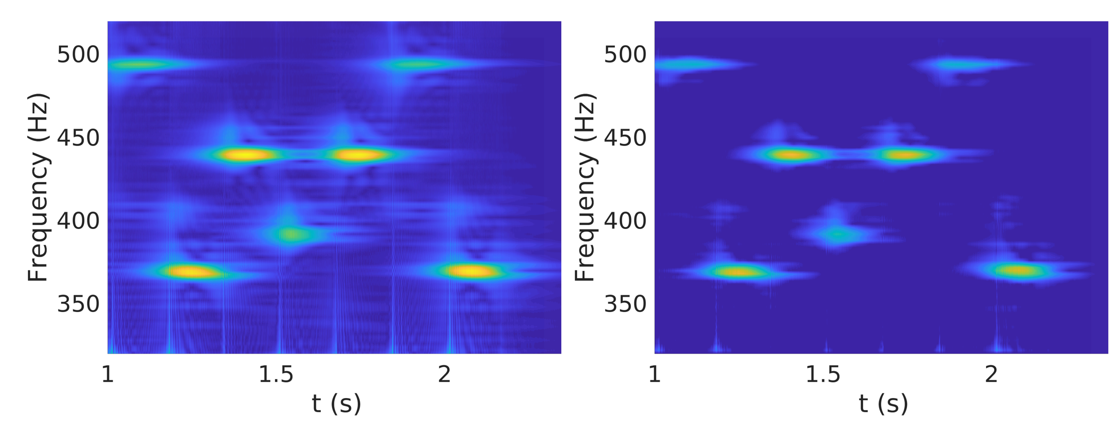


Figure: The musical note recognition using spectro-temporal estimation. The seven notes are B4, F4♯, A4, G4, A4, B4, and F4♯ with parameters $\lambda = 0.01$, $\rho = 100$, and 10 epochs by using Kalman smoother and the ALKS solver.

References

[1] R. Gao, F. Tronarp, and S. Särkkä, "Iterated extended Kalman smoother-based variable splitting for L1-regularized state estimation," *IEEE Trans. Signal Process.*, vol. 97, no. 19, pp. 5078–5092, Oct. 2019.

[2] S. Särkkä, *Bayesian Filtering and Smoothing*, Cambridge, U.K.: Cambridge Univ. Press, Aug. 2013.