Gao, Rui; Särkkä, Simo

Augmented sigma-point lagrangian splitting method for sparse nonlinear state estimation

Published in:
28th European Signal Processing Conference, EUSIPCO 2020 - Proceedings

DOI:
10.23919/Eusipco47968.2020.9287731

Published: 01/01/2020

Document Version
Publisher's PDF, also known as Version of record

Please cite the original version:
Gao, R., & Särkkä, S. (2020). Augmented sigma-point lagrangian splitting method for sparse nonlinear state estimation. In *28th European Signal Processing Conference, EUSIPCO 2020 - Proceedings* (pp. 2090-2094). Article 9287731 (European Signal Processing Conference). European Association For Signal and Image Processing. https://doi.org/10.23919/Eusipco47968.2020.9287731

Augmented Sigma-Point Lagrangian Splitting Method for Sparse Nonlinear State Estimation

Rui Gao and Simo Särkkä
Department of Electrical Engineering and Automation, Aalto University
{rui.gao, simo.sarkka}@aalto.fi

***Abstract*—Nonlinear state estimation using Bayesian filtering and smoothing is still an active area of research, especially when sparsity-inducing regularization is used. However, even the latest filtering and smoothing methods, such as unscented Kalman filters and smoothers and other sigma-point methods, lack a mechanism to promote sparsity in estimation process. Here, we formulate a sparse nonlinear state estimation problem as a generalized L_1-regularized minimization problem. Then, we develop an augmented sigma-point Lagrangian splitting method, which leads to iterated unscented, cubature, and Gauss–Hermite Kalman smoothers for computation in the primal space. The resulting method is demonstrated to outperform conventional methods in numerical experimentals.**

***Index Terms*—Nonlinear state estimation, sparsity, sigma-point, Kalman filter, variable splitting.**

I. Introduction

In many applications, the fundamental problem is to estimate the original states from degraded measurements, based upon prior knowledge on the nonlinear stochastic dynamics [1], [2]. This problem is of central importance, for example, in target tracking, biological processes, tomographic imaging reconstruction, and automatic music transcription [2], [3]. Generally, nonlinear Gaussian filtering and smoothing methods, for instance, Taylor series expansion based methods [4] or sigma-point based methods [2] can be used to estimate the state by analytical or statistical linearization of the nonlinear functions. Although such filtering and smoothing methods have been studied extensively, there is comparatively little work using sparsity-inducing regularization.

Promoting sparsity in state estimation is difficult but very much desired [3], [5]. Typically, the state is estimated by performing penalized least squares along with a sparsity-inducing regularized function [5]. It has already been demonstrated [6]–[10] that sparsity assumption can lead to significant gains in state estimation. For instance, the method [8] combined dynamic penalization with L_1-norm to obtain a cost function for sequentially estimating the state. A pseudo-measurement technique has been proposed for the Kalman innovations or filtering errors to promote sparsity [9]. The joint state estimation and parameter learning problem was solved by jointly using Kalman filter and L_1-regularization in [11].

Variable splitting optimization methods such as augmented Lagrangian splitting (ALS) [12], [13], Peaceman-Rachford splitting (PRS) [14], and first-order primal-dual (FOPD) method [15], have a significant impact in many engineering fields due to splitting of the problem into smaller parts. However, when the datasets are extremely large, consisting of hundreds of millions or billions of states in a dynamic system, solving such problems by conventional splitting methods becomes challenging. In our recent paper, we have proposed a general variable splitting framework for L_1-regularized state estimation, which is based on the combination of iterated extended Kalman smoother (IEKS) and alternating direction method of multipliers (ADMM) [3]. However, IEKS requires the Jacobians of the model functions, and may lead to an imprecise estimation result when the system is strongly nonlinear. For more accurate, but still computationally efficient estimation, our paper develops an augmented sigma-point Lagrangian splitting (ASPLS) method for the large-scale nonlinear state estimation problem.

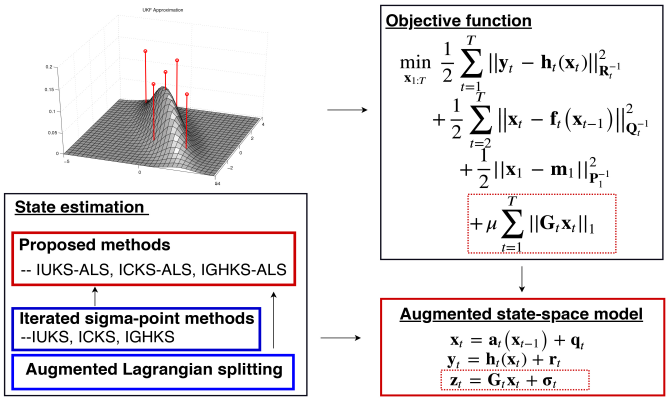

Figure 1. Our framework. With introducing the general L_1-norm, the sparse nonlinear state estimation problem is formulated as a generalized L_1-regularized minimization problem. The problem is solved by using iterated sigma-point smoothing methods combined with augmented Lagrangian splitting.

In this paper, we propose a novel approach based on a variable splitting strategy which aims at overcoming the limitations of the existing methods when introducing sparsity to nonlinear state estimation problems. First, we define a generalized L_1-regularized nonlinear least-square cost function for state estimation. We then develop a method which combines a sigma-point Kalman smoother (SPKS) with an augmented Lagrangian splitting (ALS) approach. The overall architecture is shown in Figure 1. Compared to our previous work [3], the main difference here is better capturing the higher order

The authors would like to thank Academy of Finland for funding.

moments caused by the nonlinear transforms via the use of sigma-point methods. To summarize, our main contribution is to propose a new sigma-point-based augmented Lagrangian splitting method to solve the sparse nonlinear state estimation problem. Experimental results demonstrate that the proposed method has low computational complexity and high estimation accuracy.

II. Problem Formulation

Let $\mathbf{x}_t$ be an N_x-dimensional state of a dynamic system and $\mathbf{y}_t$ be an N_y-dimensional noisy measurement. Then, the relations between the states and the measurements are governed by the nonlinear dynamic state-space model

$$\begin{aligned} \mathbf{x}_t &= \mathbf{f}_t(\mathbf{x}_{t-1}) + \mathbf{q}_t, \\ \mathbf{y}_t &= \mathbf{h}_t(\mathbf{x}_t) + \mathbf{r}_t, \quad t = 1, \ldots, T, \end{aligned} \tag{1}$$

where $\mathbf{f}_t \colon \mathbb{R}^{N_x} \to \mathbb{R}^{N_x}$ and $\mathbf{h}_t \colon \mathbb{R}^{N_x} \to \mathbb{R}^{N_y}$ are nonlinear functions, and t is the time step. The process and measurement noises $\mathbf{q}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_t)$ and $\mathbf{r}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_t)$ are assumed to be zero-mean Gaussian with covariances $\mathbf{Q}_t$ and $\mathbf{R}_t$, respectively. The initial condition is given by $\mathbf{x}_1 \sim \mathcal{N}(\mathbf{m}_1, \mathbf{P}_1)$ at $t = 1$. This paper is concerned with the estimation of the states $\mathbf{x}_{1:T}$ from the measurements $\mathbf{y}_{1:T}$.

When the system is underdetermined, that is, $N_y \ll N_x$, we can use analysis sparsity prior to stabilize the estimation of $\mathbf{x}_{1:T}$. This can be done by incorporating a generalized L_1-regularization on the solution. For this purpose, we define a pseudo-measurement variable $\mathbf{s}_t \in \mathbb{R}^P$, such that elements of $\mathbf{s}_t$ are conditionally independently Laplacian for $p = 1, \ldots, P$. The actual realizations of $\mathbf{s}_t$ are set to be zeros. Herein, the unnormalized posterior is then given by

$$\begin{aligned} &p(\mathbf{x}_{1:T} \mid \mathbf{y}_{1:T}, \mathbf{s}_{1:T}) \\ &\propto p(\mathbf{x}_1) \prod_{t=1}^{T} p(\mathbf{y}_t \mid \mathbf{x}_t) \prod_{t=2}^{T} p(\mathbf{x}_t \mid \mathbf{x}_{t-1}) \prod_{t=1}^{T} p(\mathbf{s}_t \mid \mathbf{x}_t), \end{aligned} \tag{2}$$

where

$$\begin{aligned} p(\mathbf{y}_t \mid \mathbf{x}_t) &\propto \exp\left(-\frac{1}{2}(\mathbf{y}_t - \mathbf{h}_t(\mathbf{x}_t))^{\mathsf{T}} \mathbf{R}_t^{-1} (\mathbf{y}_t - \mathbf{h}_t(\mathbf{x}_t))\right), \\ p(\mathbf{x}_t \mid \mathbf{x}_{t-1}) &\propto \exp\left(-\frac{1}{2}(\mathbf{x}_t - \mathbf{f}_t(\mathbf{x}_{t-1}))^{\mathsf{T}} \mathbf{Q}_t^{-1} (\mathbf{x}_t - \mathbf{f}_t(\mathbf{x}_{t-1}))\right), \\ p(\mathbf{s}_t \mid \mathbf{x}_t) &\propto \exp\left(-\frac{\mu}{P} \|\mathbf{s}_t - \mathbf{G}_t \mathbf{x}_t\|_1\right). \end{aligned} \tag{3}$$

Here, $\propto$ denotes proportionality and $\mathbf{G}_t \in \mathbb{R}^{P \times N_x}$. Then the maximum a posteriori (MAP) solution $\hat{\mathbf{x}}_{1:T}$ to the posterior (2) when $\mathbf{s}_t = \mathbf{0}$ is obtained by minimizing the negative logarithm of the posterior,

$$\begin{aligned} \hat{\mathbf{x}}_{1:T} = \underset{\mathbf{x}_{1:T}}{\arg\min}\, &\frac{1}{2} \sum_{t=1}^{T} \|\mathbf{y}_t - \mathbf{h}_t(\mathbf{x}_t)\|^2_{\mathbf{R}_t^{-1}} + \mu \sum_{t=1}^{T} \|\mathbf{G}_t \mathbf{x}_t\|_1 \\ &+ \frac{1}{2} \sum_{t=2}^{T} \|\mathbf{x}_t - \mathbf{f}_t(\mathbf{x}_{t-1})\|^2_{\mathbf{Q}_t^{-1}} + \frac{1}{2} \|\mathbf{x}_1 - \mathbf{m}_1\|^2_{\mathbf{P}_1^{-1}}, \end{aligned} \tag{4}$$

where $\|\cdot\|_{\mathbf{R}}$ denotes a $\mathbf{R}$-weighted Euclidean norm. Note that $\mathbf{G}_t$ can be any linear transform, for example, a frame analysis operator [6], learned analysis operator [7], or discrete gradient operator [10].

In the particular case corresponding to $\mu = 0$, the posterior distribution (or actually the smoothing distribution) can be approximated by a nonlinear Gaussian smoother (e.g., SPKS [2], [16]). Also, when the functions $\mathbf{h}_t(\mathbf{x}_t)$ and $\mathbf{f}_t(\mathbf{x}_{t-1})$ are affine, the minimization (4) can be done exactly by the Rauch–Tung–Striebel (RTS) smoother [2], [17]. However, when $\mu > 0$, these classical solutions are not applicable.

III. The Proposed Method

In this section, we present the method which is based on the combination of the ALS framework and sigma-point smoothing. We present the ALS framework in Section III-A, and introduce a sigma-point smoothing approximation as the primal variable update in Section III-B.

A. The augmented Lagrangian splitting (ALS) framework

ALS [13] is a splitting-type method that is particularly suitable for dealing with nondifferentiable term in (4). First, we introduce a sequence of auxiliary variables $\mathbf{v}_{1:T}$ and equality constraints $\mathbf{v}_t = \mathbf{G}_t \mathbf{x}_t$. Then, we reformulate the problem in (4) as an equivalent constrained problem, which is the idea of splitting methods in general. Our solution is computed by iteratively optimizing for a subset of variables at a time, while keeping all other variables fixed. We start with defining an augmented Lagrangian function

$$\begin{aligned} \mathcal{L}_\rho(\mathbf{x}_{1:T}, \mathbf{v}_{1:T}, \boldsymbol{\zeta}_{1:T}) &= \frac{1}{2} \sum_{t=1}^{T} \|\mathbf{y}_t - \mathbf{h}_t(\mathbf{x}_t)\|^2_{\mathbf{R}_t^{-1}} \\ &+ \frac{1}{2} \sum_{t=2}^{T} \|\mathbf{x}_t - \mathbf{f}_t(\mathbf{x}_{t-1})\|^2_{\mathbf{Q}_t^{-1}} + \frac{1}{2} \|\mathbf{x}_1 - \mathbf{m}_1\|^2_{\mathbf{P}_1^{-1}} \\ &+ \mu \sum_{t=1}^{T} \|\mathbf{v}_t\|_1 + \boldsymbol{\zeta}_t^{\mathsf{T}} (\mathbf{v}_t - \mathbf{G}_t \mathbf{x}_t) + \frac{\rho}{2} \sum_{t=1}^{T} \|\mathbf{v}_t - \mathbf{G}_t \mathbf{x}_t\|^2, \end{aligned} \tag{5}$$

where $\rho > 0$ is a penalty parameter. Then, we iteratively update the primal sequence $\mathbf{x}_{1:T}$, the dual sequence $\mathbf{v}_{1:T}$, and the Lagrange multiplier sequence $\boldsymbol{\zeta}_{1:T}$, until the convergence or the maximum iteration count $I_{\max}$. At each iteration i, we have the steps

$$\mathbf{x}_{1:T}^{(i)} = \underset{\mathbf{x}_{1:T}}{\arg\min}\, \mathcal{L}_\rho(\mathbf{x}_{1:T}, \mathbf{v}_{1:T}^{(i-1)}, \boldsymbol{\zeta}_{1:T}^{(i-1)}), \tag{6a}$$

$$\mathbf{v}_t^{(i)} = \max\left(|\mathbf{e}_t^{(i-1)}| - \mu/\rho, 0\right) \operatorname{sgn}\left(\mathbf{e}_t^{(i-1)}\right), \tag{6b}$$

$$\boldsymbol{\zeta}_t^{(i)} = \boldsymbol{\zeta}_t^{(i-1)} + \rho \left(\mathbf{v}_t^{(i)} - \mathbf{G}_t \mathbf{x}_t^{(i)}\right), \tag{6c}$$

for $t = 1, \ldots, T$, $\mathbf{e}_t^{(i-1)} = \mathbf{G}_t \mathbf{x}_t^{(i)} + \boldsymbol{\zeta}_t^{(i-1)}/\rho$, and sgn represents the signum function. The advantage of this iteration, using ALS, is to avoid the computational difficulties that arise when working only on the primal or the dual side.

Notice that the computational burden is mainly on the primal variable update. Ideally, we can compute the minimization

problem (6a) by batch solution, see [3] for details, but it involves the high-dimensional matrix inversion due to stacking the states of all T time steps into a batch form. Thus, we use filtering and smoothing here. Now let us describe our solution to the $\mathbf{x}_{1:T}$-subproblem (6a).

B. Sigma-point method in the primal space

Similarly to [3] we can rewrite the $\mathbf{x}_{1:T}$-subproblem (6a) as

$$\hat{\mathbf{x}}_{1:T} = \operatorname*{arg\,min}_{\mathbf{x}_{1:T}} \frac{1}{2}\sum_{t=1}^{T} \|\mathbf{y}_t - \mathbf{h}_t(\mathbf{x}_t)\|^2_{\mathbf{R}_t^{-1}} + \frac{1}{2}\|\mathbf{x}_1 - \mathbf{m}_1\|^2_{\mathbf{P}_1^{-1}} + \frac{1}{2}\sum_{t=2}^{T} \|\mathbf{x}_t - \mathbf{f}_t(\mathbf{x}_{t-1})\|^2_{\mathbf{Q}_t^{-1}} + \frac{1}{2}\sum_{t=1}^{T} \|\mathbf{z}_t - \mathbf{G}_t\mathbf{x}_t\|^2_{\mathbf{\Sigma}_t^{-1}}, \tag{7}$$

where $\boldsymbol{\sigma}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_t)$ is an artificial measurement noise with covariance $\mathbf{\Sigma}_t = \mathbf{I}/\rho$ and $\mathbf{z}_t = \mathbf{v}_t + \boldsymbol{\zeta}_t/\rho$ is a pseudo-measurement. Then, we can recognize (7) to be a special case of optimization problems, which can be solved by running a recursive smoother on the following augmented state-space model,

$$p(\mathbf{x}_t \mid \mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t \mid \mathbf{f}_t(\mathbf{x}_{t-1}), \mathbf{Q}_t), \tag{8a}$$
$$p(\mathbf{y}_t \mid \mathbf{x}_t) = \mathcal{N}(\mathbf{y}_t \mid \mathbf{h}_t(\mathbf{x}_t), \mathbf{R}_t), \tag{8b}$$
$$p(\mathbf{z}_t \mid \mathbf{x}_t) = \mathcal{N}(\mathbf{z}_t \mid \mathbf{G}_t\mathbf{x}_t, \mathbf{\Sigma}_t), \tag{8c}$$

that is also depicted in Figure 2. Strictly speaking, only a recursive smoother which computes the MAP estimate of the state is equivalent in the above sense (cf. [3]). However, it can be argued that the state estimation problem above is more fundamental than the corresponding optimization problem and hence we can replace a MAP-estimator seeking smoother with another one.

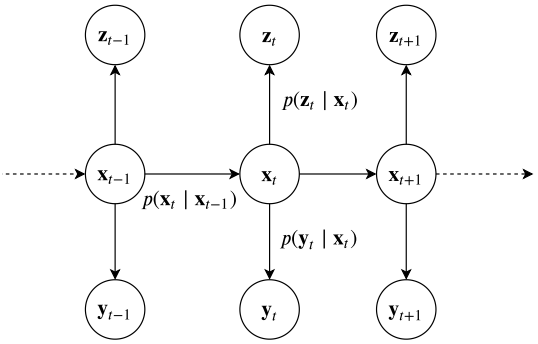

Figure 2. Augmented dynamic state space model with measurement $\mathbf{y}_t$ and argumented measurement $\mathbf{z}_t$.

As the recursive smoother we use the iterated posterior linearisation smoother [16]. In the approach, we linearize the nonlinear functions in (8) using statistical linear regression as

$$\begin{aligned} \mathbf{f}_t(\mathbf{x}_{t-1}) &\approx \mathbf{F}_t\mathbf{x}_{t-1} + \mathbf{a}_t + \mathbf{b}_t, \\ \mathbf{h}_t(\mathbf{x}_t) &\approx \mathbf{H}_t\mathbf{x}_t + \mathbf{e}_t + \mathbf{g}_t. \end{aligned} \tag{9}$$

Here $\mathbf{F}_t \in \mathbb{R}^{N_x \times N_x}$, $\mathbf{H}_t \in \mathbb{R}^{N_y \times N_x}$, $\mathbf{a}_t \in \mathbb{R}^{N_x}$, $\mathbf{e}_t \in \mathbb{R}^{N_y}$, $\mathbf{b}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Delta}_t)$, $\mathbf{g}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Xi}_t)$. In sigma-point-based statistical linear regression (see, e.g., [16]), for linearizing $\mathbf{f}_t$, we fix a base distribution $\mathcal{N}(\bar{\mathbf{m}}, \bar{\mathbf{P}})$ for $\mathbf{x}_{t-1}$, select K weight values $\omega_1, \ldots, \omega_K$, compute K sigma-points $\mathcal{X}^1, \ldots, \mathcal{X}^K$ using the base distribution, and compute transformed sigma points as $\mathcal{Z}^k = \mathbf{f}_t(\mathcal{X}^k)$. The linearization is then given by

$$\mathbf{F}_t = \mathbf{\Psi}^\mathsf{T}\bar{\mathbf{P}}^{-1}, \tag{10a}$$
$$\mathbf{a}_t = \bar{\mathbf{z}} - \mathbf{F}_t\,\bar{\mathbf{m}}, \tag{10b}$$
$$\mathbf{\Delta}_t = \mathbf{\Phi} - \mathbf{F}_t\,\bar{\mathbf{P}}\,\mathbf{F}_t^\mathsf{T}, \tag{10c}$$

where

$$\bar{\mathbf{z}} = \sum_k \omega_k \mathcal{Z}^k, \tag{11a}$$
$$\mathbf{\Psi} = \sum_k \omega_k (\mathcal{X}^k - \bar{\mathbf{m}})(\mathcal{Z}^k - \bar{\mathbf{z}})^\mathsf{T}, \tag{11b}$$
$$\mathbf{\Phi} = \sum_k \omega_k (\mathcal{Z}^k - \bar{\mathbf{z}})(\mathcal{Z}^k - \bar{\mathbf{z}})^\mathsf{T}. \tag{11c}$$

The linearization of $\mathbf{h}_t$ is similarly computed by

$$\mathbf{H}_t = \mathbf{\Psi}^T\bar{\mathbf{P}}^{-1}, \tag{12a}$$
$$\mathbf{e}_t = \bar{\mathbf{z}} - \mathbf{H}_t\,\bar{\mathbf{m}}, \tag{12b}$$
$$\mathbf{\Xi}_t = \mathbf{\Phi} - \mathbf{H}_t\,\bar{\mathbf{P}}\,\mathbf{H}_t^\mathsf{T}, \tag{12c}$$

where the sigma points for computing the quantities in (11) are now defined as $\mathcal{Z}^k = \mathbf{h}_t(\mathcal{X}^k)$ and the base distribution $\mathcal{N}(\bar{\mathbf{m}}, \bar{\mathbf{P}})$ is formed for $\mathbf{x}_t$.

The primal variable estimate is then computed by first running the Kalman filter (cf. [3]), from $t = 1$ to T,

$$\mathbf{m}_t^- = \mathbf{F}_t\,\mathbf{m}_{t-1} + \mathbf{a}_t, \tag{13a}$$
$$\mathbf{P}_t^- = \mathbf{F}_t\,\mathbf{P}_{t-1}\,\mathbf{F}_t^\mathsf{T} + \mathbf{\Delta}_t + \mathbf{Q}_t, \tag{13b}$$
$$\hat{\mathbf{y}}_t = \mathbf{H}_t\,\mathbf{m}_t^- + \mathbf{e}_t, \tag{13c}$$
$$\mathbf{S}_t^y = \mathbf{H}_t\,\mathbf{P}_t^-\,\mathbf{H}_t^\mathsf{T} + \mathbf{\Xi}_t + \mathbf{R}_t, \tag{13d}$$
$$\mathbf{K}_t^y = \mathbf{P}_t^-\,\mathbf{H}_t^\mathsf{T}\,[\mathbf{S}_t^y]^{-1}, \tag{13e}$$
$$\mathbf{m}_t^y = \mathbf{m}_t^- + \mathbf{K}_t^y\,[\mathbf{y}_t - \hat{\mathbf{y}}_t], \tag{13f}$$
$$\mathbf{P}_t^y = \mathbf{P}_t^- - \mathbf{K}_t^y\,\mathbf{S}_t^y\,[\mathbf{K}_t^y]^\mathsf{T}. \tag{13g}$$
$$\mathbf{S}_t^z = \mathbf{G}_t\,\mathbf{P}_t^y\,\mathbf{G}_t^\mathsf{T} + \mathbf{\Sigma}_t, \tag{13h}$$
$$\mathbf{K}_t^z = \mathbf{P}_t^y\,\mathbf{G}_t^\mathsf{T}\,[\mathbf{S}_t^z]^{-1}, \tag{13i}$$
$$\mathbf{m}_t = \mathbf{m}_t^- + \mathbf{K}_t^z\,[\mathbf{z}_t - \mathbf{G}_t\mathbf{m}_t^y], \tag{13j}$$
$$\mathbf{P}_t = \mathbf{P}_t^- - \mathbf{K}_t^z\,\mathbf{S}_t^z\,[\mathbf{K}_t^z]^\mathsf{T}. \tag{13k}$$

Then, we run a RTS smoother [2], [17] by the steps

$$\mathbf{G}_t = \mathbf{P}_t\,\mathbf{F}_{t+1}^\mathsf{T}\,[\mathbf{P}_{t+1}^-]^{-1}, \tag{14a}$$
$$\mathbf{m}_t^s = \mathbf{m}_t + \mathbf{G}_t\,[\mathbf{m}_{t+1}^s - \mathbf{m}_{t+1}^-], \tag{14b}$$
$$\mathbf{P}_t^s = \mathbf{P}_t + \mathbf{G}_t\,[\mathbf{P}_{t+1}^s - \mathbf{P}_{t+1}^-]\,\mathbf{G}_t^\mathsf{T}, \tag{14c}$$

for $t = T-1, \ldots, 1$.

In this article we use so-called sigma-point posterior linearization smoothers [16], where the linearization in (9) is iteratively computed by sigma points. The resulting steps of the augmented sigma-point Lagrangian splitting (ASPLS) method are summarized in Algorithm 1.

Algorithm 1: Augmented sigma-point Lagrangian splitting (ASPLS) method.

Input: $\mathbf{y}_{1:T}$, $\mathbf{f}_t$, $\mathbf{h}_t$, $\mathbf{Q}_t$, $\mathbf{R}_t$, and matrix $\mathbf{G}_t$; $\mathbf{m}_1$ and $\mathbf{P}_1$; and parameters μ and ρ.

Output: $\mathbf{x}_{1:T}$.

Initialize: $\mathbf{x}_{1:T}^{(0)}$, $\mathbf{v}_{1:T}^{(0)}$, $\boldsymbol{\zeta}_{1:T}^{(0)}$.

for $i = 1, \ldots, I_{\max}$ **do**

  Start from suitable $\mathbf{m}_{1:T}^{s,(0)}$ and $\mathbf{P}_{1:T}^{s,(0)}$;

  for $j = 1, \ldots, J_{\max}$ **do**

    for $t = 1, \ldots, T$ **do**

      Linearize function $\mathbf{f}_t$ using (10), where the nominal distribution is set to $\mathcal{N}(\mathbf{m}_{t-1}^{s,(j-1)}, \mathbf{P}_{t-1}^{s,(j-1)})$;

      Linearize function $\mathbf{h}_t$ using (12), where the nominal distribution is set to $\mathcal{N}(\mathbf{m}_{t}^{s,(j-1)}, \mathbf{P}_{t}^{s,(j-1)})$;

    end

    Run the Kalman filter for $t = 1, \ldots, T$ and the smoother for $t = T-1, \ldots, 1$ defined by (13) and (14), respectively, to obtain $\mathbf{m}_{1:T}^{s,(j)}$;

  end

  update the sequence $\mathbf{x}_{1:T}^{(i)} = \mathbf{m}_{1:T}^{s,(J_{\max})}$;

  compute the sequence $\mathbf{v}_{1:T}^{(i)}$ by (6b);

  compute the sequence $\boldsymbol{\zeta}_{1:T}^{(i)}$ by (6c);

end

IV. Experimental Results

In this numerical experiment, we consider a classical multi-sensor range measurement problem, in which the state $\mathbf{x}_t = [x_{t,1}; x_{t,2}; x_{t,3}; x_{t,4}]$ contains the location $(x_{t,1}, x_{t,2})$ and the corresponding velocities $(x_{t,3}, x_{t,4})$. The measurement dynamic model for sensor $k \in \{1, 2, 3\}$ is the following

$$\mathbf{h}_t^k(\mathbf{x}) = \sqrt{(x_{t,2} - s_y^k)^2 + (x_{t,1} - s_x^k)^2}, \tag{15}$$

and the dynamic model function is

$$\mathbf{f}_t(\mathbf{x}_{t-1}) = \begin{bmatrix} 1 & 0 & \triangle t & 0 \\ 0 & 1 & 0 & \triangle t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \mathbf{x}_{t-1}, \tag{16}$$

where (s_x^k, s_y^k) is the position of the sensor k, the measurement noise covariance is $\mathbf{R}_t = \sigma^2 \mathbf{I}$ with $\sigma = 0.2$, $\mathbf{Q}_t = \mathrm{diag}(0.01, 0.01, 0.1, 0.1)$, $\Delta t = 0.1$, and $T = 100$. We assume the target has many stops (the velocities $x_{t,3}$, $x_{t,4}$ are sparse), which corresponds to $\mathbf{G}_t = [0\,0\,1\,0; 0\,0\,0\,1]$. The performance is measured in terms of relative error $\mathbf{x}_{\text{err}} = \frac{\sum_{t=1}^{T} \|\mathbf{x}_t^{(i)} - \mathbf{x}_t^{\text{true}}\|_2}{\sum_{t=1}^{T} \|\mathbf{x}_t^{\text{true}}\|_2}$, where $\mathbf{x}_t^{\text{true}}$ is the ground truth.

Choice of the parameter μ. Figure 3 shows the relative error $\mathbf{x}_{\text{err}}$ at the iteration number i obtained by iterated unscented Kalman smoother (IUKS), iterated cubature Kalman smoother (ICKS), and iterated Gauss–Hermite Kalman smoother (IGHKS) based ALS methods for different values of the parameter $\mu \in \{0.01, 0.1, 1\}$. With the values of $\mu \in [0.1, 1]$, the relative errors are large. On the other hand, the value $\mu = 0.01$, appears to be a trade-off between iteration count and precision of the estimation. Figure 3 also shows IUKS-ALS (pink), ICKS-ALS (blue), and IGHKS-ALS (green) have similar convergence curves.

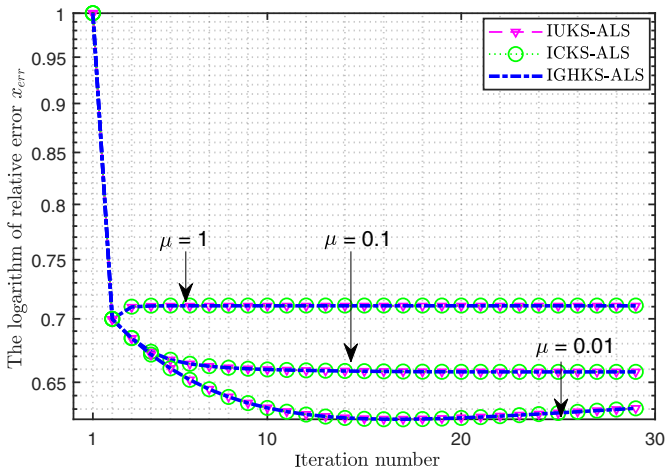

Figure 3. Convergence results of different parameter μ obtained by ASPLS. Please note that the y-axis is in log-scale.

Performance results. In Figure 4, the proposed methods (IUKS-ALS, ICKS-ALS, and IGHKS-ALS) are compared with IEKS-ADMM [3]. We also evaluate the performance without adding generalized L_1-penalized terms (i.e., $\mu = 0$) in which case we can use the classical sigma-point smoothers. We have set $I_{\max} = 20$, $J_{\max} = 20$, and $\mu = 0.01$. Figure 4 shows the relative error as function of the iteration number i. As can be seen, although the classical smoothers are computationally lighter as they only need a single outer iteration, their error is larger than the error of the proposed ASPLS methods. The error of IEKS-ADMM is slightly lower than of the classical methods, but the proposed methods outperform it significantly.

Table I
Average CPU time (sec) of the methods as the number of time steps T increases.

T	$I_{\max} = 20$			$I_{\max} = 20$, $J_{\max} = 2$		
	FOPD	PRS	ALS	IUKS-ALS	ICKS-ALS	IGHKS-ALS
10^2	0.28	0.21	0.21	0.26	0.26	1.36
10^3	12.3	6.9	6.8	1.3	1.3	11.4
10^4	2394	829	812	3.3	3.2	115.9
10^5	–	–	–	8.1	8.1	461
10^6	–	–	–	77	77	4669
10^7	–	–	–	750	749	42188
10^8	–	–	–	7267	7269	–

Large scale data computation. Table I reports the CPU time when the number of time steps T is varying from 10^2 to 10^8. We compare the performance with batch methods

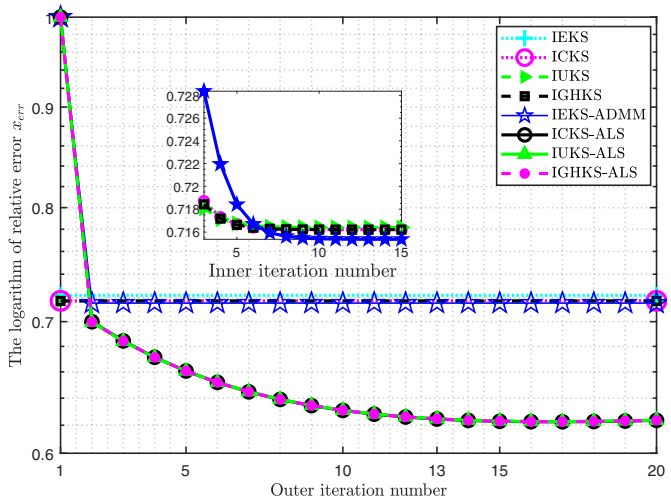

Figure 4. Relative error versus iteration number in the proposed methods and other state-of-the-art methods.

using first-order primal-dual (FOPD) method [15], Peaceman-Rachford splitting (PRS) [14], and ALS. We observed that $I_{\max} = 20$, $J_{\max} = 2$ were sufficient for convergence for out proposed methods. When $T \geq 10^5$, the batch optimization methods run out of memory ('–'). Figure 5 plots the average CPU time obtained by the mentioned methods, showing that the proposed methods successfully decrease the computation time as compared to batch methods.

It can be seen that ASPLS methods significantly outperform the batch optimization solvers, for example ALS, with respect to the CPU time. This is mainly because our methods deal with the objective using the smoother computations, while the batch methods use direct matrix or proximal gradient computations. Additionally, among the proposed methods IGHKS-ALS is the most time-consuming due to the large number of sigma points.

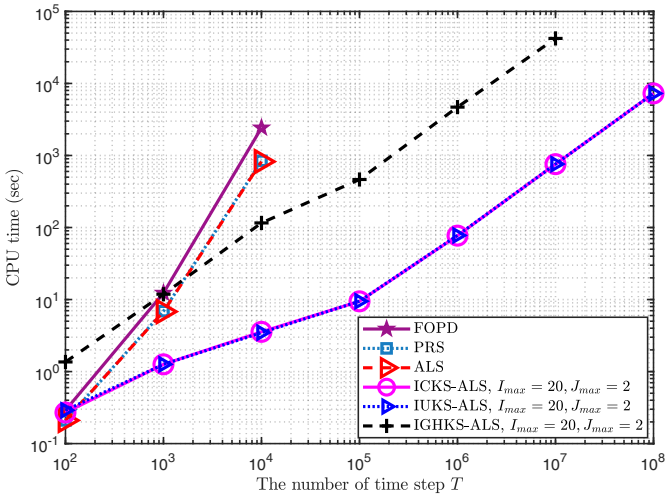

Figure 5. Average CPU time (sec) versus number of time steps in batch variable splitting methods and in the proposed methods.

V. Conclusion

In this paper, we have proposed a new method for solving nonlinear L_1-regularized state estimation problems. The methods uses a combination of sigma-point Kalman smoother (SPKS) and an augmented Lagrangian splitting (ALS) approach, where SPKF is used to efficiently update the variable in the primal space. Experiments demonstrate that when compared with other state-of-the-art methods for the same problem, the proposed augmented sigma-point Lagrangian splitting (ASPLS) method has a good accuracy while having a low computational complexity.

References

[1] Y. B. Shalom, X. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation*, Wiley, 2001.

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

[3] 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.

[4] B. Bell, "The iterated Kalman smoother as a Gauss-Newton method," *SIAM J. Optim.*, vol. 4, no. 3, pp. 626–636, Aug. 1994.

[5] A. Aravkin, J. V. Burke, L. Ljung, A. Lozano, and G. Pillonetto, "Generalized Kalman smoothing: Modeling and algorithms," *Automatica*, vol. 86, pp. 63–86, Dec. 2017.

[6] M. Elad, P. Milanfar, and R. Rubinstein, "Analysis versus synthesis in signal priors," *Inv. Probl.*, vol. 23, no. 3, pp. 947–968, Sept. 2007.

[7] R. Gao, S. A. Vorobyov, and H. Zhao, "Image fusion with cosparse analysis operator," *IEEE Signal Process. Lett.*, vol. 24, no. 7, pp. 943–947, July 2017.

[8] A. S. Charles, A. Balavoine, and C. J. Rozell, "Dynamic filtering of time-varying sparse signals via L1 minimization," *IEEE Trans. Signal Process.*, vol. 64, no. 21, pp. 5644–5656, Nov. 2016.

[9] A. Carmi, P. Gurfil, and D. Kanevsky, "Methods for sparse signal recovery using kalman filtering pseudo-measuremennt norms and quasi-norms," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2405–2409, Apr. 2010.

[10] R. Gao, F. Tronarp, and S. Särkkä, "Combined analysis-L1 and total variation ADMM with applications to MEG brain imaging and signal reconstruction," in *26th European Signal Proces. Conf. (EUSIPCO)*, Roma, Italy, Sept. 2018, pp. 1930–1934, IEEE.

[11] R. Gao, F. Tronarp, and S. Särkkä, "Regularized state estimation and parameter learning via augmented Lagrangian Kalman smoother method," in *29th International Workshop on Machine Learning for Signal Process.*, Pittsburgh, PA, USA, Oct. 2019, pp. 1–6, IEEE.

[12] S. J. Wright and J. Nocedal, *Numerical Optimization*, Springer Verlag,, 2006.

[13] R. Glowinski, S. J. Osher, and W. Yin, *Splitting Methods in Communication, Imaging, Science, and Engineerin*, Springer, 2017.

[14] B. S. He, H. Liu, Z. R. Wang, and X. M. Yuan, "A strictly contractive Peaceman–Rachford splitting method for convex programming," *SIAM J. Optim.*, vol. 24, no. 3, pp. 1011–1040, 2014.

[15] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *J. Math. Imaging. Vis.*, vol. 40, no. 1, pp. 120–145, May 2011.

[16] Á. F. Garcia-Fernández, L. Svensson, and S. Särkkä, "Iterated posterior linearisation smoother," *IEEE Trans. Automat. Control*, vol. 62, no. 4, pp. 2056–2063, 2017.

[17] H. E. Rauch, F. Tung, and C. T. Striebel, "Maximum likelihood estimates of linear dynamic system," *AIAA J.*, vol. 3, no. 8, pp. 1445–1450, Aug. 1965.