DATA SKETCHING FOR LARGE-SCALE KALMAN FILTERING

Dimitris Berberidis[1] *and Georgios B. Giannakis*[1]

[1] Dept. of ECE and Digital Tech. Center, University of Minnesota
Minneapolis, MN 55455, USA
Emails: {bermp001,georgios}@umn.edu

ABSTRACT

In an age of exponentially increasing data generation, performing inference tasks by utilizing the available information in its entirety is not always an affordable option. The present paper puts forth approaches to render tracking of large-scale dynamic processes affordable, by processing a reduced number of data. Two distinct methods are introduced for reducing the number of data involved per time step. The first method builds on reduction using low-complexity random projections, while the second performs censoring for data-adaptive measurement selection. Simulations on synthetic data, compare the proposed methods with competing alternatives, and corroborate their efficacy in terms of estimation accuracy over complexity reduction.

Index Terms– tracking, dimensionality reduction, censoring, random projections, Kalman filter.

1. INTRODUCTION

Tracking dynamically evolving processes is of paramount importance in a wide range of applications. In the context of large scale problems, being able to perform accurate and economical state estimation may render problems of prohibitive scale feasible. Weather prediction is an example of tracking a slowly-varying dynamic process, from a massive volume of observations acquired from fast-sampling sensors per time interval; see e.g., [1]. Monitoring large and dynamically evolving networks, where nodes may join or leave and connections may be established or lost as time progresses, provides an exciting domain in which the acquisition and processing of network-wide performance metrics becomes challenging as the network-size increases [2, Ch. 8]. For instance, monitoring path metrics such as delays or loss rates is challenging primarily because the number of paths generally grows as the square of the number of nodes in the network. Therefore, measuring and storing the delays of all possible source-destination pairs is hard in practice even for moderate-size networks[2].

Work in this paper was supported by the MURI AFOSR FA9550-10-1-0567, NIH Grant No. 1R01GM104975-01, and NSF grants 1343860, 1442686, 1500713 and 1514056.

In this context, channel-aware dimensionality reduction of observations was proposed in [3] and [4] using distributed wireless sensor networks (WSNs). A posterior-CRLB-based method to select sensors for tracking tracking was introduced in [5], and a greedy algorithm leveraging submodularity was developed in [6] for measurement selection in sequential estimation.

The present paper draws on interval censoring to discard "less informative" observations online. Censoring has recently been employed to select data for distributed parameter estimation using resource-constrained WSNs, thus trading off performance for tractability [7, 8]. Furthermore, censoring has been proposed for signal estimation using WSNs, for tracking, and control of dynamical processes [9, 10, 11, 12]. However, existing works on censoring mostly focus on reducing the rate at which sensors communicate their observations, and pertinent methods exhibit large computational complexity and storage requirements.

The goal of this paper is to perform reliable tracking using the Kalman filter (KF), while reducing the amount of data and the computational complexity involved. Towards this goal, two algorithms are proposed for dimensionality reduction and tracking. The first is based on random projections (RPs) and thus is data-agnostic, while the second adopts censoring for joint tracking and rejection of "uninformative" data (see also our work in [13]). Corroborating simulations compare with the state-of-the-art greedy measurement selection algorithm, and illustrate the efficiency of the proposed methods.

2. PROBLEM FORMULATION

Consider the following linear dynamical system model

$$\boldsymbol{\theta}_n = \mathbf{F}_n\boldsymbol{\theta}_{n-1} + \mathbf{G}_n\mathbf{u}_n + \mathbf{w}_n \tag{1}$$

$$\mathbf{y}_n = \mathbf{X}_n\boldsymbol{\theta}_n + \mathbf{v}_n \tag{2}$$

where $\boldsymbol{\theta}_n \in \mathbb{R}^p$ denotes the state vector at time n; $\mathbf{F}_n$ is the known state-transition matrix;$\mathbf{G}_n$ and $\mathbf{u}_n$ are known and deterministic control-input model and control-input vector respectively; $\mathbf{y}_n \in \mathbb{R}^D$ the measurement vector and $\mathbf{X}_n$ is the known $D \times p$ measurement matrix; while $\mathbf{w}_n$ and $\mathbf{v}_n$ are zero-mean, mutually uncorrelated and individually uncorrelated across time random noise vectors, with respective covariance matrices $\mathbf{Q}_n$ and $\mathbf{R}_n$. The initial state $\boldsymbol{\theta}_0$ has mean $\mathbf{m}_0$, and covariance $\mathbf{P}_0$.

Given the information-bearing data $\mathcal{I}_n := \{\mathbf{y}_n, \mathbf{X}_n, \mathbf{R}_n\}$ of the measurement equation (2) at time n, the most recent estimate $\hat{\boldsymbol{\theta}}_{n-1|n-1}$ and its covariance matrix $\mathbf{P}_{n-1|n-1}$, the celebrated KF yields the minimum mean square error (MMSE) optimal estimate $\hat{\boldsymbol{\theta}}_{n|n}$ in two steps. First, the state prediction $\hat{\boldsymbol{\theta}}_{n|n-1}$ and its covariance matrix $\mathbf{P}_{n|n-1}$ are obtained using the model dynamics $\{\mathbf{F}_n, \mathbf{Q}_n\}$ as [cf. (1)]

$$\begin{aligned}\hat{\boldsymbol{\theta}}_{n|n-1} &= \mathbf{F}_n\hat{\boldsymbol{\theta}}_{n-1|n-1} + \mathbf{G}_n\mathbf{u}_n\\ \mathbf{P}_{n|n-1} &= \mathbf{F}_n\mathbf{P}_{n-1|n-1}\mathbf{F}_n^T + \mathbf{Q}_n.\end{aligned}$$

Subsequently, as $\mathcal{I}_n$ becomes available, $\hat{\boldsymbol{\theta}}_{n|n}$ is obtained as

$$\hat{\boldsymbol{\theta}}_{n|n} = \arg\min_{\boldsymbol{\theta}} \|\mathbf{y}_n - \mathbf{X}_n\boldsymbol{\theta}\|^2_{\mathbf{R}_n^{-1}} + \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{n|n-1}\|^2_{\mathbf{P}^{-1}_{n|n-1}}. \tag{4}$$

The first term of the cost in (4) is a weighted least-squares term fitting the state $\boldsymbol{\theta}$ with $\mathcal{I}_n$ that arises from the linear observation model in (2); while the second regularization term corresponds to treating $\hat{\boldsymbol{\theta}}_{n|n-1}$ as a prior of $\boldsymbol{\theta}_n$. Solving (4) and applying the matrix inversion lemma (MIL) yields the well-known KF correction step, e.g., [14, p. 205]

$$\hat{\boldsymbol{\theta}}_{n|n} = \hat{\boldsymbol{\theta}}_{n|n-1} + \mathbf{K}_n(\mathbf{y}_n - \mathbf{X}_n\hat{\boldsymbol{\theta}}_{n|n-1})$$

where the so-termed KF gain $\mathbf{K}_n$ and the state covariance update are given by

$$\begin{aligned}\mathbf{K}_n &= \mathbf{P}_{n|n-1}\mathbf{X}_n^T\left(\mathbf{X}_n\mathbf{P}_{n|n-1}\mathbf{X}_n^T + \mathbf{R}_n\right)^{-1}\\ \mathbf{P}_{n|n} &= (\mathbf{I}_p - \mathbf{K}_n\mathbf{X}_n)\,\mathbf{P}_{n|n-1}.\end{aligned}$$

A dual form of the KF known as the information filter (IF) relies on the MIL to offer a more efficient solver of (4) as D grows large [14, Ch. 7]. Nevertheless, even the low-complexity IF requires $\mathcal{O}(Dp^2)$ multiplications to solve (4) in the case of uncorrelated observations ($\mathbf{R}_n$ diagonal), and $\mathcal{O}(D^2p)$ in general. Therefore, for large-scale KF problems where $D \gg p$, dimensionality reduction of the datasets $\mathcal{I}_n$ is an attractive tool for rendering the solution of (4) computationally tractable, while also reducing other data-related costs, such as storage and transmission.

Towards this goal, we introduce a reduced-complexity Kalman-like filter (see Algorithm 1) that extracts a reduced (size $d < D$), yet informative dataset $\mathcal{I}_n^d := \{\check{\mathbf{y}}_n, \check{\mathbf{X}}_n, \check{\mathbf{R}}_n\}$ from the original $\mathcal{I}_n$, where $\check{\mathbf{y}}_n \in \mathbb{R}^d, \check{\mathbf{X}}_n \in \mathbb{R}^{d\times p}$ and $\check{\mathbf{R}}_n \in \mathbb{R}^{d\times d}$ are the corresponding reduced-dimension observation vector, measurement matrix, and covariance matrix. Consequently, the problem reduces to the design of low-complexity sketching modules for informative dimensionality reduction. In the ensuing two sections, a *data-agnostic* method based on RPs followed by a *data-adaptive* method based on censoring are developed.

Algorithm 1 Reduced-complexity KF

<u>**Initialization:**</u> $\hat{\boldsymbol{\theta}}_{0|0} = \mathbf{m}_0,\ \mathbf{P}_{0|0} = \mathbf{P}_0$
for $n = 1 : N$ **do**
  <u>**Prediction Step:**</u>
  $\hat{\boldsymbol{\theta}}_{n|n-1} = \mathbf{F}_n\hat{\boldsymbol{\theta}}_{n-1|n-1} + \mathbf{G}_n\mathbf{u}_n$
  $\mathbf{P}_{n|n-1} = \mathbf{F}_n\mathbf{P}_{n-1|n-1}\mathbf{F}_n^T + \mathbf{Q}_n$
  <u>**Data Reduction :**</u>
  $\{\check{\mathbf{y}}_n, \check{\mathbf{X}}_n, \check{\mathbf{R}}_n\} = \text{Sketching}\big(\{\mathbf{y}_n, \mathbf{X}_n, \mathbf{R}_n\},\ \hat{\boldsymbol{\theta}}_{n|n-1}\big)$
  <u>**Correction Step:**</u>
  $\hat{\boldsymbol{\theta}}_{n|n} = \hat{\boldsymbol{\theta}}_{n|n-1} + \mathbf{K}_n(\check{\mathbf{y}}_n - \check{\mathbf{X}}_n\hat{\boldsymbol{\theta}}_{n|n-1})$
  $\mathbf{K}_n = \mathbf{P}_{n|n-1}\check{\mathbf{X}}_n^T\left(\check{\mathbf{X}}_n\mathbf{P}_{n|n-1}\check{\mathbf{X}}_n^T + \check{\mathbf{R}}_n\right)^{-1}$
  $\mathbf{P}_{n|n} = \left(\mathbf{I}_p - \mathbf{K}_n\check{\mathbf{X}}_n\right)\mathbf{P}_{n|n-1}$
end for

Algorithm 2 RP scetching module

Dimensionality reduction with RPs
$\check{\mathbf{y}}_n = \mathbf{S}_d\mathbf{H}\boldsymbol{\Lambda}\mathbf{y}_n$
$\check{\mathbf{X}}_n = \mathbf{S}_d\mathbf{H}\boldsymbol{\Lambda}\mathbf{X}_n$
$\check{\mathbf{R}}_n = \mathbf{S}_d\mathbf{H}\boldsymbol{\Lambda}\mathbf{R}_n(\mathbf{S}_d\mathbf{H}\boldsymbol{\Lambda})^T$

3. RP-BASED KF

RP-based dimensionality reduction amounts to premultiplying measurements and regressors $\{\mathbf{y}_n, \mathbf{X}_n\}$ with a random matrix $\mathbf{H}$, and a diagonal matrix $\boldsymbol{\Lambda}$, whose entries take the values $\{+1/\sqrt{D}, -1/\sqrt{D}\}$ equiprobably. The net result is a linear transformation of the system of equations in which all rows are of approximately equal importance. A subset of d rows of the transformed system is then extracted by simple random sampling, implemented by left multiplication with a random $d \times D$ selection matrix $\mathbf{S}_d$.

Originally developed in the context of linear regression [15, 16], RPs can be readily adapted to reduce dimensionality in tracking dynamical processes too. Applying the Hadamard preconditioning and random sampling matrices on (2) yields the reduced-dimension observation model

$$\check{\mathbf{y}}_n = \mathbf{S}_d\mathbf{H}\boldsymbol{\Lambda}\mathbf{y}_n = \mathbf{S}_d\mathbf{H}\boldsymbol{\Lambda}(\mathbf{X}_n\boldsymbol{\theta}_n + \mathbf{v}_n) = \check{\mathbf{X}}_n\boldsymbol{\theta}_n + \check{\mathbf{v}}_n$$

where $\check{\mathbf{v}}_n := \mathbf{S}_d\mathbf{H}\boldsymbol{\Lambda}\mathbf{v}_n$ is zero-mean with covariance $\check{\mathbf{R}}_n = \mathbf{S}_d\mathbf{H}\boldsymbol{\Lambda}\mathbf{R}_n(\mathbf{S}_d\mathbf{H}\boldsymbol{\Lambda})^T$. Given $\hat{\boldsymbol{\theta}}_{n|n-1}$ and the reduced data $\mathcal{I}_n^d$, state estimate $\hat{\boldsymbol{\theta}}_{n|n}$ can be obtained similar to (4) as

$$\hat{\boldsymbol{\theta}}_{n|n} = \arg\min_{\boldsymbol{\theta}} \|\check{\mathbf{y}}_n - \check{\mathbf{X}}_n\boldsymbol{\theta}\|^2_{\check{\mathbf{R}}_n^{-1}} + \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_{n|n-1}\|^2_{\mathbf{P}^{-1}_{n|n-1}}. \tag{5}$$

Solving (5) and applying the MIL gives rise to the novel RP-based KF, which is equivalent to Algorithm 1 using Algorithm 2 as sketching module.

Implementing RPs can have affordable complexity if $\mathbf{H}$ is chosen to be a pseudo-random Hadamard matrix of size D. Different from the more elaborate approaches in [3] and [4], the proposed RP-KF is an easy-to-implement, "one-size-fits-all" reduced-complexity tracker, using data-agnostic dimensionality reduction. Furthermore, RP-KF's estimation performance can be guaranteed as asserted in the ensuing proposition, which provides a benchmark for the data-driven methods introduced in the following section.[1]

Proposition 1. *With* $\mathbf{R}_n = \sigma_n^2 \mathbf{I}_D$*, let* $\mathbf{A}_n := [\mathbf{P}_{n|n-1}^{-1/2}, \sigma_n^{-1}\mathbf{X}_n^T]^T$*,* $\mathbf{b}_n := [\mathbf{P}_{n|n-1}^{-1/2}\hat{\boldsymbol{\theta}}_{n|n-1}, \ \sigma_n^{-1}\mathbf{y}_n^T]^T$*, and* $\mathbf{A}_n = \mathbf{U}_n\mathbf{S}_n\mathbf{V}_n^T$*. If* $\|\mathbf{U}_n\mathbf{U}_n^T\mathbf{b}_n\|_2 \geq \gamma\|\mathbf{b}_n\|_2$ *for some* $\gamma \in (0,1]$*, then by choosing* $d = \mathcal{O}(p\ln(pD)/\epsilon)$ *the following bound for the RP-KF estimates holds w.h.p.*

$$\|\hat{\boldsymbol{\theta}}_{n|n} - \hat{\boldsymbol{\theta}}^{\star}_{n|n}\|_2 \leq \sqrt{\epsilon}\left(\kappa(\mathbf{A}_n)\sqrt{\gamma^{-2}-1}\right)\|\hat{\boldsymbol{\theta}}^{\star}_{n|n}\|_2$$

where $\kappa(\mathbf{A}_n)$ *denotes the condition number of* $\mathbf{A}_n$*, and* $\hat{\boldsymbol{\theta}}^{\star}_{n|n}$ *is the full-data KF estimate.*

4. CENSORING-BASED KF

Measurement censoring for estimating dynamical processes has recently been advocated as a means of reducing the inter-sensor transmission overhead when WSNs are employed for distributed tracking [10, 11]; see also [9, 17], where censoring is employed for event-based estimation. Since the goal in the aforementioned applications is saving communication resources, censoring is performed solely on measurements $\mathbf{y}_n$, with $\mathbf{X}_n$ and $\mathbf{R}_n$ assumed to be known; thus, in our notation [10, 9, 17, 8, 11] rely on $\mathcal{I}_n := \{\mathbf{y}_n\}$. A set of d observations $\mathcal{I}_n^d := \{[\mathbf{y}_n]_{\mathcal{S}_n}\}$ is obtained, where $[\mathbf{y}_n]_i$ denotes the $i-$th element of $\mathbf{y}_n$, and $\mathcal{S}_n \subseteq \{1,\ldots,D\}$ denotes a set collecting the indices of uncensored observations. Given $[\mathbf{y}_n]_{\mathcal{S}_n}, \mathbf{X}_n$ and $\mathbf{R}_n$, [10, 9, 17, 8, 11] develop sequential estimators to optimally estimate $\boldsymbol{\theta}_n$. In the context of reducing communicational load, optimal (in the ML or MMSE sense) estimation from censored observations comes with complexity comparable to that of using the full number of measurements.

Since our goal is dimensionality and complexity reduction, the starting point is on censoring entire rows of the full dataset $\mathcal{I}_n^d := \{\mathbf{y}_n, \mathbf{X}_n, \mathbf{R}_n\}$, in order to obtain a reduced set $\mathcal{I}_n^d := \{[\mathbf{y}_n]_{\mathcal{S}_n}, [\mathbf{X}_n]_{\mathcal{S}_n}, [\mathbf{R}_n]_{\mathcal{S}_n}\}$, where $[\mathbf{X}_n]_i$ denotes the $i-$th row of $\mathbf{X}_n$ and $[\mathbf{R}_n]_{\mathcal{S}_n} := \mathrm{cov}([\mathbf{v}_n]_{\mathcal{S}_n})$. In this context, the goal is to develop censoring rules in order to obtain $\mathcal{S}_n$, so that $\mathcal{I}_n^d$ is an "informative" subset of $\mathcal{I}_n$. Most existing censoring schemes consider the innovation $\tilde{\mathbf{y}}_n := \mathbf{y}_n - \mathbf{X}_n\hat{\boldsymbol{\theta}}_{n|n-1}$ as a measure of information contained in $\mathbf{y}_n$. One approach –henceforth referred to as *block censoring* (BC)– is to censor the entire vector $\mathbf{y}_n$. From an information-theoretic viewpoint [11], the optimal BC rule is based on the magnitude of the prewhitened innovation $\boldsymbol{\Sigma}_n^{-1/2}\tilde{\mathbf{y}}_n$, where $\boldsymbol{\Sigma}_n := \mathrm{cov}(\tilde{\mathbf{y}}_n) = \mathbf{X}_n\mathbf{P}_{n|n-1}\mathbf{X}_n^T + \mathbf{R}_n$; thus, $\mathcal{S}_n$ is obtained as

$$\mathcal{S}_n := \begin{cases} \{1,\ldots,D\}, & \|\boldsymbol{\Sigma}_n^{-1/2}\tilde{\mathbf{y}}_n\|_2 > \tau_n \\ \emptyset, & \text{otherwise} \end{cases}. \tag{6}$$

Clearly, having $\mathcal{S}_n = \emptyset$ corresponds to skipping the correction step of the KF. A major shortcoming of (6) is the cubic complexity $\mathcal{O}(D^3)$ associated with calculating $\boldsymbol{\Sigma}_n$. Furthermore, BC-KF may only reduce the data cost *on average* across iterations by entirely skipping correction steps.

Our idea of a more attractive alternative is to censor separately each entry of $\mathcal{I}_n^D$. In our context, entry-wise censoring yields $\mathcal{S}_n$ as

$$\mathcal{S}_n := \{1 \leq i \leq D \mid |[\tilde{\mathbf{y}}_n]_i| > \tau_n\} \tag{7}$$

where τ_n can be designed so that the set cardinality $|\mathcal{S}_n| \approx d$. Compared to BC-KF, the innovation-based entry-wise rule of (7) is more flexible in reducing the available data. To accurately perform measurement selection with (7), $|[\tilde{\mathbf{y}}_n]_i|$ must reflect how informative $[\mathbf{y}_n]_i$ is for the purpose of tracking $\boldsymbol{\theta}_n$. Using the predictor-based innovations for this purpose $[\tilde{\mathbf{y}}_n]_i := \mathbf{y}_n - [\mathbf{X}_n]_{i,:}\hat{\boldsymbol{\theta}}_{n|n-1}$ is intuitive, but unsuitable for the proposed reduced-complexity KF, since uncensored observations adhere to a highly nonlinear measurement model.

Towards a more suitable censoring rule, the adaptive censoring least-mean-square (AC-LMS) algorithm we introduced in [18] can be employed to discard uninformative rows of $\mathcal{I}_n^D$. Within time slot n, rows of $(\mathbf{y}_n, \mathbf{X}_n)$ are processed sequentially; given a temporary estimate $\hat{\boldsymbol{\theta}}^{(i-1)}_{n|n-1}$, the $i-$th row is discarded when indicated so by the censoring variable ($\mathbb{1}_.$ denotes the indicator function)

$$c_i = \mathbb{1}\{\left|[\mathbf{y}_n]_i - [\mathbf{X}_n]_{i,:}\hat{\boldsymbol{\theta}}^{(i-1)}_{n|n-1}\right| \leq \tau_n\} \tag{8}$$

where $[\mathbf{X}_n]_{i,:}$ is the $i-$th row of $\mathbf{X}_n$. If deemed informative enough, the $i-$th row is added to the collection $\mathcal{S}_n$, and subsequently involved in updating $\hat{\boldsymbol{\theta}}^{(i-1)}_{n|n-1}$ as

$$\hat{\boldsymbol{\theta}}^{(i)}_{n|n-1} = \hat{\boldsymbol{\theta}}^{(i-1)}_{n|n-1} + \mu[\mathbf{X}_n]_{i,:}\left([\mathbf{y}_n]_i - [\mathbf{X}_n]_{i,:}\hat{\boldsymbol{\theta}}^{(i-1)}_{n|n-1}\right). \tag{9}$$

The role of the update in (9) is to perturb the censoring region (slab) towards the direction of $([y_n]_i, [\mathbf{X}_n]_{i,:})$, thus making it less likely for similar points to be obtained; intuitively speaking, such updates eliminate redundancies in the selected data and reduce estimation error. The AC sketching module is summarized as Algorithm 3, and when plugged into Algorithm 1, it yields the proposed adaptive censoring (AC)-KF scheme.

Proposition 2. *If* $\mathbf{w}_n$ *and* $\mathbf{v}_n$ *are Gaussian, then the AC-KF yields unbiased estimates* $\forall\tau,\mu$*.*

[1] Detailed proofs are ommited due to space limitation but they will be included in the journal version of this paper.

Algorithm 3 AC sketching module

Measurement selection with AC-LMS

Input: $\hat{\boldsymbol{\theta}}_{n|n-1}, \ \{\mathbf{y}_n, \mathbf{X}_n, \mathbf{R}_n\}$

Initialization: $\hat{\boldsymbol{\theta}}^{(0)}_{n|n-1} = \hat{\boldsymbol{\theta}}_{n|n-1}, \ \mathcal{S}_n^{(0)} = \emptyset$

for $i = 1 : D$ **do**

  Obtain c_i as in (8)

  if $c_i = 0$, **then**

    $\mathcal{S}_n^{(i)} = \mathcal{S}_n^{(i-1)} \cup \{i\}$

    Update $\hat{\boldsymbol{\theta}}^{(i-1)}_{n|n-1}$ as in (9)

  end if

end for

Return :$\{\check{\mathbf{y}}_n, \check{\mathbf{X}}_n, \check{\mathbf{R}}_n\} = \{[\mathbf{y}_n]_{\mathcal{S}_n^{(D)}}, [\mathbf{X}_n]_{\mathcal{S}_n^{(D)},:}, [\mathbf{R}_n]_{\mathcal{S}_n^{(D)}}\}$

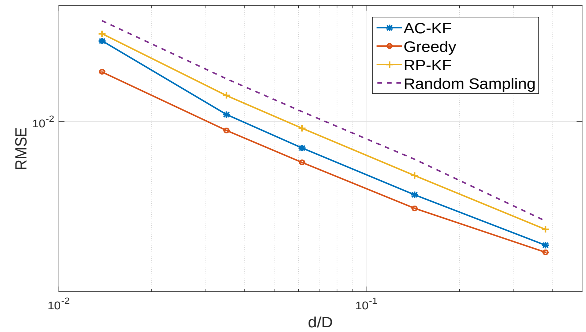
Proposition 2, whose proof is omitted due to space limitations, asserts the unbiasdness of AC-KF. Nevertheless, the variance of AC-KF largely depends on the choice of μ. Tuning μ to optimize the MSE performance of AC-KF is a challenging task; the development of accurate rules for the selection of μ is part of ongoing our research. Nevertheless, even for possibly suboptimal values of μ, the proposed scheme yields promising results. Simulations in Section 5 will demonstrate that the proposed AC-KF attains estimation accuracy close to that of the KF using the greedy measurement selection method in [6]. As importantly, the proposed AC performs a single pass over the data, and requires $\mathcal{O}(Dp)$ computations, which is markedly lower than the $\mathcal{O}(Ddp^2)$ required to perform greedy selection. Furthermore, AC-KF is suitable for online implementation by processing rows of $\mathcal{I}_n^D$ sequentially.

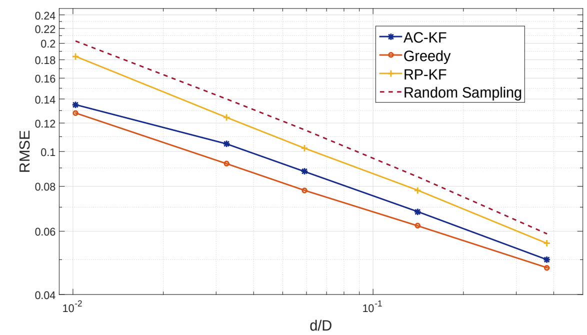
5. NUMERICAL TESTS

The novel AC-KF and RP-KF algorithms are tested here on a simulated linear dynamical system modeling a random spiral trajectory, which consists of a rotation on the $\mathrm{x}-\mathrm{y}$ plane, and a downward movement along the z axis with

$$\mathbf{F}_n = \begin{bmatrix} \cos(\phi) & \sin(\phi) & 0 \\ -\sin(\phi) & \cos(\phi) & 0 \\ 0 & 0 & a_z \end{bmatrix}, \ \forall n$$

where ϕ determines the angular speed of rotation set to $\pi/60$, and a_z the rate of descent set to 0.997. The state noise $\{\mathbf{w}_n\}_{n=1}^N$ was generated i.i.d. with $\mathbf{w}_n \sim \mathcal{N}(\mathbf{0}, \sigma_w^2 \mathbf{Q}_n)$, where $[\mathbf{Q}_n]_{i,j} = 0.5^{|i-j|}$, and $\sigma_w = 0.02$. Finally, the initial state is $\boldsymbol{\theta}_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{P}_0)$, with $\mathbf{m}_0 = [1, 1, 10]^T$ and $\mathbf{P}_0 = 0.09\mathbf{I}$. Per time instant $n \in \{1, \ldots, N\}$ with $N = 100$, $D = 1{,}000$ measurements are obtained and concatenated in vector $\mathbf{y}_n = \mathbf{X}_n^T \boldsymbol{\theta}_n + \mathbf{v}_n$, where rows of $\mathbf{X}_n$ are generated as i.i.d. standardized Gaussian vectors. For this experiment, observations are assumed correlated; thus, $\mathbf{v}_n \sim \mathcal{N}(0, \sigma_v^2 \mathbf{R}_n)$, where $[\mathbf{R}_n]_{i,j} = 0.5^{|i-j|}$.





(b)

Figure 1: Average RMSE for AC-KF, Greedy algorithm, RP-KF and random sampling as a function of data reduction ratio d/D. (Upper) High SNR case with $\sigma_v^2 = 25 \times 10^{-4}$ and (lower) Low SNR case with $\sigma_v^2 = 1$.

To determine the average performance in terms of estimation error and computational complexity of AC-KF and RP-KF for different values of d/D, 20 Monte Carlo realizations were run on the same simulated linear dynamical system. The experiment was repeated for different levels of signal-to-noise-ratio (SNR) at the observation model. One high ($\sigma_v^2 = 25 \times 10^{-4}$) and one low ($\sigma_v^2 = 1$) SNR scenario are presented here. The estimation performance was measured in terms of averaged across realizations, root-mean-square error (RMSE) of the estimates across iterations; that is, $\mathrm{RMSE} = N^{-1}\sqrt{\sum_{n=1}^N \|\hat{\boldsymbol{\theta}}_{n|n} - \boldsymbol{\theta}_n\|}$. AC-KF was run first, with $\tau_n = \tau$ tuned so that a constant number of approximately d observations were selected per time slot.

The average RMSE of the four methods as a function of d/D is plotted in Figs. 1(a) and 1(b), for high and low SNR, respectively. These plots confirm that the proposed data-agnostic RP-KF is useful for increasing the accuracy (compared to plain random sampling) when estimating dynamic processes. With regards to the more elaborate algorithms, the proposed AC-KF has comparable performance with the KF using greedy measurement selection, while being orders of magnitude faster in terms of runtime. Furthermore, the gap between estimation accuracy of the two methods closes as SNR decreases, indicating that the AC-KF is more robust to noisy observations.

6. REFERENCES

[1] P. Louka, G. Galanis, N. Siebert, G. Kariniotakis, P. Katsafados, I. Pytharoulis, and G. Kallos, "Improvements in wind speed forecasts for wind power prediction purposes using Kalman filtering," *J. of Wind Engin. and Indust. Aero.*, vol. 96, no. 12, pp. 2348–2362, Nov. 2008.

[2] E. D. Kolaczyk, *Statistical Analysis of Network data*. Springer, 2009.

[3] H. Zhu, I. D. Schizas, and G. B. Giannakis, "Power-efficient dimensionality reduction for distributed channel-aware Kalman tracking using WSNs," *IEEE Trans. Sig. Proc.*, vol. 57, pp. 3193–3207, Aug. 2009.

[4] H. Ma, Y.-H. Yang, Y. Chen, K. R. Liu, and Q. Wang, "Distributed state estimation with dimension reduction preprocessing," *IEEE Trans. Sig. Proc.*, vol. 62, no. 12, pp. 3098–3110, Dec. 2014.

[5] L. Zuo, R. Niu, and P. K. Varshney, "Posterior CRLB based sensor selection for target tracking in sensor networks," in *Proc. of Inte. Conf. on Acous., Speech and Signal Proc.*, vol. 2, Honolulu, Hawaii, 2007, pp. II–1041.

[6] M. Shamaiah, S. Banerjee, and H. Vikalo, "Greedy sensor selection: Leveraging submodularity," in *Proc. of 49th Conf. on Decision and Control*, Atlanta, Dec. 2010, pp. 2572–2577.

[7] E. Msechu and G. B. Giannakis, "Sensor–centric data reduction for estimation with WSNs via censoring and quantization," *IEEE Trans. Sig. Proc.*, vol. 60, no. 1, pp. 400–414, Jan. 2012.

[8] K. You, L. Xie, and S. Song, "Asymptotically optimal parameter estimation with scheduled measurements," *IEEE Trans. Sig. Proc.*, vol. 61, no. 14, pp. 3521–3531, July 2013.

[9] Q. Liu, Z. Wang, X. He, and D. Zhou, "A survey of event-based strategies on control and estimation," *Systems Science & Control Engineering Open Access Journal*, vol. 2, no. 1, pp. 90–97, 2014.

[10] G. Battistelli, A. Benavoli, and L. Chisci, "Data-driven strategies for selective data transmission in sensor networks," in *Proc. of 51st Conf. on Decision and Control*, Grand Wailea, Maui, 2012, pp. 800–805.

[11] Y. Zheng, R. Niu, and P. K. Varshney, "Sequential Bayesian estimation with censored data for multi-sensor systems," *IEEE Trans. Sig. Proc.*, vol. 62, no. 10, pp. 2626–2641, Oct. 2014.

[12] G. Wang, J. Chen, J. Sun, and Y. Cai, "Power scheduling of Kalman filtering in wireless sensor networks with data packet drops," *arXiv preprint arXiv:1312.3269v2*, 2013.

[13] D. Berberidis and G. B. Giannakis, "Budgeted kalman filtering and smoothing for economical tracking with big distributed data," in *Proc. of 49th Asilomar Conf. on Signals, Systems and Computers (to appear)*, Pacific Grove, CA, Nov. 2015.

[14] Y. Bar-Shalom, X. R. Li, and T. Kirubarajan, *Estimation with Applications to Tracking and Navigation: Theory, Algorithms and Software*. John Wiley & Sons, 2004.

[15] M. Mahoney, "Randomized algorithms for matrices and data," *Found. Trends. in Mach. Learn.*, vol. 3, no. 2, pp. 123–224, 2011.

[16] P. Drineas, M. W. Mahoney, and S. Muthukrishnan, "Sampling algorithms for l 2 regression and applications," in *Proc. of the 17-th SIAM-ACM Symp. on Discrete Alg.*, Miami, Florida, 2006, pp. 1127–1136.

[17] Y. Wang, V. Krishnaswami, and G. Rizzoni, "Event-based estimation of indicated torque for ic engines using sliding-mode observers," *Control Engineering Practice*, vol. 5, no. 8, pp. 1123–1129, 1997.

[18] D. Berberidis, V. Kekatos, G. Wang, and G. B. Giannakis, "Adaptive censoring for large-scale regressions," in *Proc. of 40th Conf. on Acoust., Speech and Signal Proc.*, Brisbane, Australia, April 2015, pp. 5475–5479.