Quantized Filtering Schemes for Multi-Sensor Linear State Estimation: Stability and Performance Under High Rate Quantization

Alex S. Leong*, Subhrakanti Dey, and Girish N. Nair

Abstract—In this paper we consider state estimation of a discrete time linear system using multiple sensors, where the sensors quantize their individual innovations, which are then combined at the fusion center to form a global state estimate. We prove the stability of the estimation scheme under sufficiently high bit rates. We obtain asymptotic approximations for the error covariance matrix that relates the system parameters and quantization levels used by the different sensors. Numerical results show close agreement with the true error covariance for quantization at high rates. An optimal rate allocation problem amongst the different sensors is also considered.

EDICS: SSP-FILT, SEN-FUSE

I. INTRODUCTION

Linear state estimation using multiple sensors is a commonly performed task in areas such as radar tracking and industrial monitoring. Nowadays, much of the communication systems used in practice are digital in nature. Therefore, analog measurements made by sensors will need to be quantized before transmission to a central processor or fusion center over a bandwidth limited wireless channel. Proposing a quantized estimation scheme that is stable, and characterizing the performance loss due to quantization, for a multi-sensor linear state estimation problem, is the primary focus of this paper.

We consider a discrete time linear system. A number of sensors take measurements, perform some local processing before transmitting a processed signal to a fusion center, which then combines these signals to form a global state estimate. At the sensor level, each sensor will quantize their innovations.1 This is motivated by the fact that for unstable systems,2 while the state will become unbounded, the (true) innovations process remains of bounded variance [2]. Thus, quantizing the innovations rather than the state estimates avoids possible saturation of the quantizer. These quantized innovations are then sent to a fusion center to form a global state estimate, using a modification of the decentralized scheme for unquantized Kalman filtering in [3].

The authors are with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Vic. 3010, Australia. E-mail {asleong,sdey,gnair}@unimelb.edu.au.

This work was supported by the Australian Research Council

1In this paper, we refer to innovations as \( y_k - C \dot{z}_k \), where \( y_k \) is a measurement and \( \dot{z}_k \) is a state estimate obtained using a quantized filter. This differs from the definition of innovations used in Kalman filtering (which we will refer to as the true innovations), which is \( y_k - C \hat{x}_k^{kf} \), where \( \hat{x}_k^{kf} \) is the state estimate obtained using the (unquantized) Kalman filter.

2For instance, the use of unstable systems in kinematic models for radar tracking is well established [1].

Related Work

The works of [4], [5] gave structural results on optimal coding for state estimation with measurements obtained over a finite rate digital link, though the focus is on determining minimum bit rates required for stability rather than performance analysis. For control problems with quantized state feedback, the performance with high rate quantization has been studied in e.g. [6] and [7]. The idea of quantizing innovations for estimation has been considered in [8], [9], [10] with different filtering equations from ours. However [9] and [10] only consider the case of a single sensor, while the multi-sensor setup in [8] does not involve a fusion center but instead requires sensors to broadcast their quantized innovations to all other sensors. Furthermore, as pointed out in [11], the schemes of [8], [9] are not guaranteed to be stable for unstable systems, while [10] proves stability of their scheme only for systems with bounded noise support. In [12] quantization of measurements is carried out after performing an optimization of the quantization levels, but their scheme requires feedback of the state estimates from the fusion center back to the sensors. In [11], a filter which involves quantizing the true innovations at the sensor is given, but it is shown that for unstable systems the mean squared error always becomes unbounded with this scheme.

Particle filtering schemes are also considered in [11], though the performance of such schemes are difficult to analyze theoretically.

Another related area is the CEO problem [13], [14], where a number of agents observe a memoryless source and then communicate these observations over rate limited channels (of rate \( R \)) to a central CEO, which then reconstructs the source with a certain distortion \( D(R) \). Here however we consider sources which are not memoryless but governed by a linear state space dynamical system.

Summary of Contributions

In this paper we consider and analyze a multi-sensor quantized filtering scheme. In particular, the main contributions of this paper are:

- We prove that even for unstable systems, this quantized filtering scheme is stable for sufficiently high bit rates.
- We derive asymptotic approximations relating the estimation error performance to the system parameters and the number of quantization levels used by the different sensors. This can be seen as a first step towards achieving a quantization rate versus state estimation error trade-off for multi-terminal estimation of linear dynamical systems, which is largely unavailable in the current literature.
For systems with scalar measurements/quantizers, the performance is analyzed for the uniform quantizer, while for systems with vector measurements, the asymptotic performance is derived for lattice vector quantizers.

- We will consider a rate allocation problem, for allocating a total rate among the sensors in order to optimize the asymptotic error performance.

The paper is organized as follows. Scalar quantizers are studied in Section II. We first review the unquantized decentralized Kalman filter equations in Section II-A, in order to motivate our choice of quantized filtering equations which are presented in Section II-B. We then prove stability of our filtering scheme in Section II-C. In Section II-D we obtain an asymptotic approximation for the error covariance in terms of the number of quantization levels used by the different sensors, as well as the system parameters. In Section II-E we study a rate allocation problem for minimizing the trace of the error covariance matrix at the fusion center when the total rate across the sensors is constrained. In Section III extensions of the results to lattice vector quantizers is considered. Numerical comparisons are made between the asymptotic expression and Monte Carlo simulations of the true error covariance matrix in Section IV. While our asymptotic expressions are derived assuming high rate quantization, numerical results suggest that they are quite accurate even for rates as low as 3-4 bits per sample (per sensor/dimension).

Notations: In this paper, $\mathbb{E}[\cdot]$ will denote the expected value, $\text{tr}(\cdot)$ the trace, $\Gamma(\cdot)$ the Gamma function, and a random vector is Gaussian with mean $\mu$ and covariance matrix $\Sigma$. A matrix $X \succ 0$ if it is positive definite. We use the big-O notation (see e.g. [15]), where for functions $f(\cdot)$ and $g(\cdot)$, we say that $f(x) = O(g(x))$ as $x \to x_0$, if there exists a constant $K$ such that $|f(x)| \leq K|g(x)|$ for all $x$ within some neighbourhood of $x_0$. We also say that $f(x) \sim g(x)$ as $x \to x_0$, if $\lim_{x \to x_0} f(x)/g(x) = 1$.

II. SYSTEMS WITH SCALAR MEASUREMENTS

Throughout this paper, we will use $k$ to denote the discrete time index, and $i$ the sensor index. We consider a discrete time vector linear system

$$x_{k+1} = Ax_k + w_k,$$

where $x_k \in \mathbb{R}^n$ and $w_k$ is i.i.d. zero mean Gaussian with covariance matrix $\Sigma_w \succeq 0$. There are $M$ different sensors each making scalar measurements:

$$y_{i,k} = C_ix_k + v_{i,k}, \quad i = 1, \ldots, M$$

where $y_{i,k} \in \mathbb{R}$, and $v_{i,k}$ is i.i.d. zero mean Gaussian with covariance $\Sigma_{v,i} > 0$. We assume that $\{w_k\}$ and $\{v_{i,k}\}$ are mutually independent, and that the pair $(A, \Sigma_w^{1/2})$ is stabilizable. The pair $(A, C)$ is detectable, where $C \triangleq \begin{bmatrix} C^T_1 & \cdots & C^T_M \end{bmatrix}^T$, however the individual sensor pairs $(A, C_i)$ are not necessarily detectable for all $i$. The case where the individual pairs $(A, C_i)$ are all detectable has been previously studied in [16].

It is assumed that the individual sensors can perform some local processing, with a fusion center then using an appropriate fusion rule to compute a global estimate of the state $x_k$. See Fig. 1 for a diagram of the system model.

![System model](image)

In this paper, we will use similar analytical methods to [5], [17] to prove the stability of our quantized filtering scheme, which will require the assumption that $w_k$ and $v_{i,k}$ for $i = 1, \ldots, M$ have uniformly bounded $(2 + \epsilon)-$th absolute moments for some $\epsilon > 0$. Since we are dealing with Gaussian noise here, which have moments of all orders, such an assumption is automatically satisfied.

A. Decentralized Kalman filter

In [3], it is shown that in the case where there is no quantization, each sensor can run its own individual Kalman filter to obtain local estimates of the full state $x_k$, which can then be combined at the fusion center to obtain a global state estimate, that is the same as if the fusion center had access to the individual measurements. However, if for some sensor $i$ the pair $(A, C_i)$ is not detectable and $A$ is an unstable matrix, then the local error covariance of this sensor becomes unbounded over time. Consequently the local true innovations (defined here as $y_{i,k} - C_i\hat{x}_{i,k|k-1}$, where $\hat{x}_{i,k|k-1}$ is sensor $i$'s estimate of $x_k$) will also have unbounded variance, making quantization of the local innovations in Section II-B infeasible.

The approach taken in this paper is for sensors to only estimate their observable parts of the state. Due to the overall system $(A, C)$ being detectable, the fusion center can use these local estimates (or true innovations) to form estimates of the full state $x_k$. Such an approach has also been used in e.g. [18] in the context of state estimation with data-driven communications. In this subsection we present the decentralized Kalman filter equations (without quantization). In Section II-B these equations will be modified in our quantized filtering scheme.

It is well-known (see e.g. [19]) that one can always find nonsingular matrices $T_i$ for $i = 1, \ldots, M$, such that

$$T_i^{-1}AT_i = \begin{bmatrix} A_{1,i} & 0 \\ A_{21,i} & A_{2,i} \end{bmatrix}, \quad C_iT_i = \begin{bmatrix} C_{1,i} \\ 0 \end{bmatrix},$$

with the pair $(A_{1,i}, C_{1,i})$ being observable. Note that the $T_i$'s are not unique and many different choices are possible.

Another possible approach is to allow some feedback from the fusion center back to the sensors (e.g. a broadcast to all sensors), which will enable sensors to have estimates of the full state. However, due to space restrictions this approach will not be considered here.
Partition $T_i^{-1}$ as

$$T_i^{-1} = \begin{bmatrix} D_{1,i} \\ D_{2,i} \end{bmatrix}$$

and call

$$\begin{bmatrix} x_{i,k}^o \\ x_{i,k}^f \end{bmatrix} \triangleq T_i^{-1} x_k = \begin{bmatrix} D_{1,i} x_k \\ D_{2,i} x_k \end{bmatrix}$$

where $x_{i,k}^o$ and $x_{i,k}^f$ are respectively the observable and unobservable components of the state $x_k$ at sensor $i$. Then the subsystem

$$x_{i,k+1}^o = A_{1,i} x_{i,k}^o + D_{1,i} w_k$$

$$y_{i,k} = C_{1,i} x_{i,k}^o + v_{i,k} \quad (= C_i x_k + v_{i,k})$$

is observable. Note that if the pair $(A, C_i)$ is observable, then one can obviously choose $T_i = I$, and then $A_{1,i} = A$, $C_{1,i} = C_i$, $D_{1,i} = I$, $x_{i,k}^o = x_k$. In the case where the pair $(A, C_i)$ is detectable but not observable, one can still set $A_{1,i} = A$, $C_{1,i} = C_i$, $D_{1,i} = I$ in the equations below.

Define the local estimates and error covariances:

$$\hat{x}_{i,k|1} = \mathbb{E}[x_{i,k}^o | y_{i,0}, \ldots, y_{i,k-1}]$$

$$\tilde{x}_{i,k|k} = \mathbb{E}[x_{i,k}^o - y_{i,k} | y_{i,0}, \ldots, y_{i,k}]$$

$$P_{i,k|k} = \mathbb{E}[(x_{i,k}^o - \tilde{x}_{i,k|k})(x_{i,k}^o - \tilde{x}_{i,k|k})^T | y_{i,0}, \ldots, y_{i,k}]$$

The local true innovations process at sensor $i$ is defined as:

$$\hat{y}_{i,k} = y_{i,k} - C_{1,i} \tilde{x}_{i,k|1|k}$$

It is well-known (see e.g. [2]) that $\hat{y}_{i,k}$ is Gaussian with zero mean and variance $C_{1,i} P_{i,k|k-1} C_{1,i}^T + \Sigma_{i,v}$. The local estimates can be computed at the sensors using local Kalman filters as follows:

$$\tilde{x}_{i,k|k} = A_{1,i} \tilde{x}_{i,k|k-1} + K_{i,k} (y_{i,k} - C_{1,i} \tilde{x}_{i,k|1|k})$$

$$K_{i,k} = P_{i,k|k-1} C_{1,i}^T (C_{1,i} P_{i,k|k-1} C_{1,i}^T + \Sigma_{i,v})^{-1}$$

$$P_{i,k|k} = A_{1,i} P_{i,k|k-1} A_{1,i}^T + D_{1,i} \Sigma_w D_{1,i}^T$$

The fusion center can then use these local estimates and error covariances to form a global estimate of the full state $x_k$ using various different fusion rules. In [18], the fusion center fuses the local estimates together using a BLUE (best linear unbiased estimate) criterion. However this approach requires cross covariances between different sensors to be computed, and furthermore doesn’t appear to be equivalent to the Kalman filter estimate where the fusion center has access to all the sensor measurements. Instead, here we will give the decentralized Kalman filter equations with which the fused global state estimate is the same as if the fusion center had access to the individual measurements. The equations can be derived using similar techniques to [3],[20] (which only considered the case where sensors had local estimates of the full state $x_k$) and thus the derivations are omitted for brevity.

Define the global quantities:

$$\tilde{x}_{k|k-1} = \mathbb{E}\{x_k | y_{0,0}, \ldots, y_{k-1}\}$$

$$\tilde{y}_{k} = \mathbb{E}\{x_k - \tilde{x}_{k|k-1} \}^T y_{0,0}, \ldots, y_{k-1}\}$$

$$P_{k|k} = \mathbb{E}\{(x_k - \tilde{x}_{k|k})(x_k - \tilde{x}_{k|k})^T | y_{0,0}, \ldots, y_{k}\}$$

where $y_{k} = (y_{k,0}^T, \ldots, y_{k,M}^T)^T$. The fusion center makes use of the local estimates $\tilde{x}_{i,k|k-1}$ and $\tilde{x}_{i,k|k}$, local error covariances $P_{i,k|k-1}$ and $P_{i,k|k}$, and knowledge of $D_{1,i}$, to compute global estimates as follows:

$$\tilde{x}_{k|k} = A \tilde{x}_{k|k-1}$$

$$\tilde{y}_{k|k} = P_{k|k}(1 - P_{k|k}^{-1} A P_{k|k}^{-1} + \Sigma_{w})^{-1} \{P_{k|k}^{-1} \tilde{y}_{k|k} - P_{k|k}^{-1} \tilde{x}_{k|k} - \Sigma_{w}\}$$

$$P_{k|k} = A P_{k|k-1} A^T + \Sigma_{w}$$

$$P_{k|k} = P_{k|k-1} - P_{k|k-1} C_{k|k}^T (C_{k|k} P_{k|k-1} C_{k|k}^T + \Sigma_{v})^{-1} C_{k|k} P_{k|k-1}$$

where $C = [C_{1}^T | \ldots | C_{M}^T]^T$ and $\Sigma_{v}$ is a diagonal matrix given by $\Sigma_{v} = \text{diag}(\Sigma_{1,v}, \ldots, \Sigma_{M,v})$. Note that instead of the sensors sending their local estimates and error covariances, the local true innovations $\hat{y}_{i,k} = y_{i,k} - C_{1,i} \tilde{x}_{i,k|1|k}$ can be sent to the fusion center first, since the fusion center can reconstruct $\tilde{x}_{i,k|k} = \tilde{x}_{i,k|k-1|k} + P_{i,k|k} \hat{y}_{i,k}$ from $\hat{y}_{i,k}$ provided it has knowledge of all the sensor parameters $C_i$ and $\Sigma_{i,v}$, $i = 1, \ldots, M$, and the sensor observability decompositions. Such parameters can for instance be transmitted by the sensors to the fusion center offline before the estimation begins. In order to perform the reconstruction, the fusion center will also need to run copies of the local Kalman filter equations (4) of each sensor. Note however that it does not require feedback from the fusion center back to the sensors.

As $k \to \infty$, the local error covariance matrices $P_{i,k|k-1}$ converge to steady state values $P_{i,\infty}$ satisfying the algebraic Riccati equations:

$$P_{i,\infty} = A_{1,i} P_{i,\infty} A_{1,i}^T + D_{1,i} \Sigma_{w} D_{1,i}^T - A_{1,i} P_{i,\infty} C_{i,\infty} (C_{i,\infty} P_{i,\infty} C_{i,\infty}^T + \Sigma_{v})^{-1} C_{i,\infty} P_{i,\infty} A_{1,i}^T$$

$$i = 1, \ldots, M,$$

and the global error covariance matrix $P_{k|k}$ has steady state value $P_{k,\infty}$ that satisfies the algebraic Riccati equation

$$P_{k,\infty} = A P_{k,\infty} A^T + \Sigma_{w} - A P_{k,\infty} C (C P_{k,\infty} C^T + \Sigma_{v})^{-1} C P_{k,\infty} A^T$$

$$i = 1, \ldots, M.$$
Remark 2.1: In [3], [20] the information form of the Kalman filter is used, so that the equation for the updating of $P_{k|k}$ in (5) can be given equivalently as

$$P_{k|k}^{f^{-1}} = P_{k|k-1}^{f^{-1}} + \sum_{i=1}^{M} D_{i}^{T} \left( P_{i,k|k}^{f^{-1}} - P_{i,k|k-1}^{f^{-1}} \right) D_{i}^{T}.$$  

If the sensors transmit their local state estimates and error covariances and $D_{i}$, then knowledge of the sensor parameters at the fusion center is not required when using the information form. However, since the focus of this paper is on quantization, where we want to quantize the innovations instead of the measurements/states, we try to minimize the amount of information that needs to be transmitted by assuming more knowledge at the fusion center, as will be also done in the next subsection.

B. Quantized filtering scheme

We will consider a suboptimal quantized filtering scheme which are a modified version of the unquantized decentralized Kalman filtering equations given in (4)-(5). In this scheme, the individual sensors run the following equations, for $i = 1, \ldots, M$:

$$\begin{align*}
\hat{x}_{i,k|k-1} &= A_{i} \hat{x}_{i,k|k-1} - 1 \\
\hat{x}_{i,k|k} &= \hat{x}_{i,k|k-1} + K_{i,k} q_{i,k} \left( y_{i,k} - C_{i} \hat{x}_{i,k|k-1} \right) \\
K_{i,k} &= P_{i,k|k-1} C_{i}^{T} \left( C_{i} P_{i,k|k-1} C_{i}^{T} + \Sigma_{i,v} \right)^{-1} \\
P_{i,k|k-1} &= A_{i} P_{i,k|k-1} A_{i}^{T} + D_{i} \Sigma_{i,w} D_{i}^{T} \\
P_{i,k|k} &= P_{i,k|k-1} - K_{i,k} C_{i} P_{i,k|k-1} \\
\end{align*}$$

while the fusion center runs the following equations:

$$\begin{align*}
\hat{x}_{k|k-1} &= A \hat{x}_{k|k-1} \\
\hat{x}_{k|k} &= P_{k|k} \left( P_{k|k}^{-1} \hat{x}_{k|k-1} + \sum_{i=1}^{M} D_{i}^{T} \left( P_{i,k|k}^{-1} \hat{x}_{i,k|k-1} - P_{i,k|k}^{-1} \hat{x}_{i,k|k-1} \right) \right) \\
P_{k|k} &= A P_{k|k} A^{T} + \Sigma_{w} \\
P_{k|k} &= P_{k|k} - K_{k} C_{k} \left( C_{k}^{T} + \Sigma_{n,k} \right)^{-1} C_{k} P_{k|k-1} \\
\end{align*}$$

In (8)-(9), $q_{i,k}$ is the quantization of $y_{i,k} - C_{i} \hat{x}_{i,k|k-1}$ that is sent by sensor $i$ to the remote fusion center. $\hat{x}_{i,k|k-1}$ and $\hat{x}_{i,k|k}$ are the local state estimates and $P_{i,k|k-1}$, $P_{i,k|k}$ are the local error covariances. Similarly $\hat{x}_{k|k-1}$, $\hat{x}_{k|k}$, $P_{k|k-1}$, $P_{k|k}$ are the corresponding global quantities. We will often use the shorthand $P_{i,k} \triangleq P_{i,k|k-1}$ and $P_{k} \triangleq P_{k|k-1}$. Note that due to quantization $\hat{x}_{i,k|k-1}$, $P_{i,k|k}$, and $y_{i,k} - C_{i} \hat{x}_{i,k|k-1}$ are not the true conditional mean, error covariance matrix and innovations respectively, but for high rate quantization the approximations will be quite accurate, see e.g. [10]. As stated in the introduction, the motivation for quantizing the innovations is due to the fact that for unstable systems, while the state will become unbounded, the (true) innovations process remains of bounded variance. 8

In (9), $\Sigma_{n,k} = \text{diag}(\Sigma_{1,n,k}, \ldots, \Sigma_{M,n,k})$ is a diagonal matrix with terms $\Sigma_{i,n,k}$ to account for the quantization noise variances of each sensor, similar to [22]. See Section II-B1 on how the quantizers $q_{i,k}(\cdot)$ are chosen and the corresponding expressions for $\Sigma_{i,n,k}$. The terms $l_{i,k}$ are the scaling factors of each sensor $i$, which allows one to adaptively change the quantizer range to account for possible quantizer overload, similar to e.g. [23], [5], and is needed in order to prove the stability of the quantized filtering scheme for noises with infinite support [5]. See Section II-B2 for details on how $l_{i,k}$ are chosen.

Similar to the previous subsection, the fusion center can reconstruct $\hat{x}_{i,k|k}$, $\hat{x}_{i,k|k+1}$, $P_{i,k|k}$ and $P_{i,k|k+1}$ from $l_{i,k} q_{i,k} \left( y_{i,k} - C_{i} \hat{x}_{i,k|k-1} \right)$, and knowledge of the sensor parameters, observability decompositions, quantizers $q_{i,k}(\cdot)$, and how the scaling factors $l_{i,k}$ are updated. As before, such knowledge can be provided by the sensors to the fusion center offline before the estimation begins, and no feedback is required from the fusion center back to the sensors.

1) Choice of quantizer: In this paper the quantizers will be assumed to be fixed rate quantizers, but with time-varying quantizer ranges. In particular, the performance of our quantized filtering scheme using (scalar) uniform quantizers will be analyzed. The scalar Lloyd-Max “optimal” quantizer can be analyzed using similar techniques but the results will only be mentioned briefly to avoid repetition.9 The non-uniform quantizer of [5] can also be used to give a stable quantized filtering scheme, however its performance seems to be more difficult to analyze.

Let $N_{i}$ denote the number of quantization levels for the quantizer used by sensor $i$. The rate of the quantizer used by sensor $i$ is denoted by

$$R_{i} = \log_{2}(N_{i}),$$

see also Section II-E. Thus the case of high rate quantization will refer to either large $R_{i}$ or large $N_{i}$ interchangeably.

Under high rate quantization, we assume that the quantity $y_{i,k} - C_{i} \hat{x}_{i,k|k-1}$ is approximately $N(0, C_{i} P_{i,k|k} C_{i}^{T} + \Sigma_{i,v})$, since the quantization noise is dominated by the Gaussian process and measurement noise. Some studies on the accuracy of the Gaussian approximation of the quantization error, for the case of logarithmic quantizers, can be found in [10].

Suppose now that $q_{i}(\cdot)$ is a quantizer of $N_{i}$ levels designed for quantization of $N(0,1)$ random variables. Then in (8) we can rewrite

$$l_{i,k} q_{i,k} \left( y_{i,k} - C_{i} \hat{x}_{i,k|k-1} \right) = \sigma_{i,k} l_{i,k} q_{i,k} \left( \frac{y_{i,k} - C_{i} \hat{x}_{i,k|k-1}}{\sigma_{i,k} l_{i,k}} \right),$$

where $\sigma_{i,k}^{2} \triangleq C_{i} P_{i,k|k} C_{i}^{T} + \Sigma_{i,v}$. 8

This is the case for open loop systems. For closed loop systems with feedback control to stabilize the system the states will remain of bounded variance and quantization of the measurements could also be feasible. 9

The uniform quantizer can be generalized to the case of lattice vector quantizers, see Section III, whereas optimal vector quantizers are not easy to find in general.
For uniform quantization of Gaussian random variables, the asymptotically optimal step sizes of the quantizer for large $N_i$ has been derived in [24]. Under high rate quantization, the step size $\Delta_{N_i}$ is asymptotically

$$\Delta_{N_i} \sim \frac{4\sqrt{\ln N_i}}{N_i} \sigma$$

where $\sigma^2$ is the variance of the Gaussian random variable that is to be quantized. Using similar notation to [5], the uniform quantizer of [24], for variance $\sigma^2 = 1$, can then be expressed as follows: Partition the real line into $N$ intervals $(-\infty, -\frac{2(N_i-1)\sqrt{\ln N_i}}{N_i}), [-\frac{2(N_i-2)\sqrt{\ln N_i}}{N_i}, -\frac{2(N_i-2)\sqrt{\ln N_i}}{N_i} + \Delta_{N_i}), \ldots, [\frac{2(N_i-2)\sqrt{\ln N_i}}{N_i}, \infty)$. Label these intervals $I(1)$, $I(2), \ldots, I(N_i)$ respectively. The quantized value of $x$ is then

$$\tilde{q}_i(x) = \begin{cases} \text{midpoint of } I(\omega), & x \in I(\omega), \omega \in \{2, \ldots, N_i - 1\} \\ -\frac{2(N_i-1)\sqrt{\ln N_i}}{N_i}, & x \in I(\omega), \omega = 1 \\ \frac{2(N_i-1)\sqrt{\ln N_i}}{N_i}, & x \in I(\omega), \omega = N_i \end{cases}$$

where $\omega$ represents the index of the quantizer region that $x$ lies in. The resulting squared error distortion is asymptotically (see [24])

$$D_{N_i} \sim \frac{4\ln N_i}{3N_i^2} \triangleq \delta_{N_i}$$

The term $\Sigma_{i,n,k}$ in (8) is then defined as

$$\Sigma_{i,n,k} \triangleq \frac{\delta_{N_i}(C_{1,i}P_{i,k}C_{1,i}^T + \Sigma_{i,v})}{1 + \delta_{N_i}}$$

where $\delta_{N_i} = \frac{4\ln N_i}{3N_i^2}$. For the case of Lloyd-Max quantization, one can similarly derive that $\delta_{N_i} = \frac{\pi N_i}{2\sqrt{\ln N_i}}$.

The recursion for $P_{i,k}$ in (8) can thus be written as

$$P_{i,k+1} = A_{1,i}P_{i,k}A_{1,i}^T + D_{1,i}\Sigma_{i,v}D_{1,i}^T = A_{1,i}P_{i,k}C_{1,i}^T(C_{1,i}P_{i,k}C_{1,i}^T + \Sigma_{i,v})^{-1}C_{1,i}P_{i,k}A_{1,i}^T + \delta_{N_i}$$

Such recursions have been previously studied in the literature in the context of the modified algebraic Riccati equation, see [25] and the references therein. In particular, since $C_{1,i}$ is a row vector which has rank 1, it is shown in [26] that $P_{i,k}$ converges to a steady state value $P_{i,\infty}$ satisfying the modified algebraic Riccati equation

$$P_{i,\infty} = A_{1,i}P_{i,\infty}A_{1,i}^T + D_{1,i}\Sigma_{i,v}D_{1,i}^T = A_{1,i}P_{i,\infty}C_{1,i}^T(C_{1,i}P_{i,\infty}C_{1,i}^T + \Sigma_{i,v})^{-1}C_{1,i}P_{i,\infty}A_{1,i}^T$$

if and only if

$$\frac{1}{1 + \delta_{N_i}} > \frac{1}{\prod_{j} |\lambda_{ij}(A_{1,i})|^2}$$

where $\lambda_{ij}(A_{1,i})$ are the unstable eigenvalues of $A_{1,i}$. Since $\delta_{N_i} \to 0$ as $N_i \to \infty$, the condition (10) will be met for $N_i$ sufficiently large. Note that $K_{i,k}$ and $\sigma_{i,k}$ will then also tend towards steady state values as $k \to \infty$, facts which will be used in the sequel, e.g. (13)-(14).

2) Choice of scaling factors: We now describe how the scaling factors $l_{i,k}$ are chosen. Following a similar approach to [17], the scaling factors $l_{i,k}$ are updated recursively as follows:

$$l_{i,k} = \frac{||C_{1,i}||l_{i,k} + d_{i,v}}{\sigma_{i,k}}$$

$$l_{i,k} = ||A_{1,i}(I - K_{i,k}C_{1,i})||l_{i,k-1} + d_{i,w} + ||A_{1,i}K_{i,k}||d_{i,v} + ||A_{1,i}K_{i,k}||(||C_{1,i}||l_{i,k-1} + d_{i,v})\kappa(\omega_{i,k-1})$$

where $d_{i,v} > 0, d_{i,w} > 0, \forall i$ are arbitrary constants, and $\kappa(\omega_{i,k})$ is defined as:

$$\kappa(\omega_{i,k}) = \begin{cases} \beta(N_i) & \omega_{i,k} \in \{2, \ldots, N_i - 1\} \\ \gamma(N_i) & \omega_{i,k} \in \{1, N_i\} \end{cases}$$

where $\beta(\cdot)$ and $\gamma(\cdot)$ are functions of $N_i$ that need to be chosen appropriately in order to prove stability in Section II-C, see e.g. Lemma 2.2. We will choose here

$$\beta(N_i) = \frac{2\sqrt{\ln N_i}}{N_i}$$

which corresponds to the half length of a quantizer interval, similar to [5]. The function $\gamma(\cdot)$ will be chosen to be

$$\gamma(N_i) = \sqrt{\ln N_i}$$

The choice of values of $d_{i,v}$ and $d_{i,w}$ in (11) will affect the performance of the quantizing scheme. The intuitive reason is that the locations of the quantizer points for the quantizer $\tilde{q}_i(\cdot)$ are designed assuming $l_{i,k} = 1$, so the performance is expected to perform well when $l_{i,k} \approx 1$ most of the time. Under high rate quantization, this can be achieved as follows. For a given $d_{i,w}$, let $l_{i,\text{min}}$ satisfy the equation

$$l_{i,\text{min}} = ||A_{1,i}(I - K_{i,k}C_{1,i})||l_{i,\text{min}} + d_{i,w} + ||A_{1,i}K_{i,k}||d_{i,v} + ||A_{1,i}K_{i,k}||(||C_{1,i}||l_{i,\text{min}} + d_{i,v})\beta(N_i)$$

with $d_{i,v}$ chosen such that

$$\frac{||C_{1,i}||l_{i,\text{min}} + d_{i,v}}{\sigma_{i}} = 1,$$
Proof See Appendix A.

Thus in the case of high rate quantization, where quantizer saturation is rare so that \( \omega_{i,k} \in \{2, \ldots, N_i - 1\} \) for much of the time, one can keep \( l_{i,k} \) close to 1 with this choice of \( d_{i,v} \) and \( d_{i,w} \).

C. Stability of quantized filtering scheme

Define the estimation error

\[
f_k = x_k - \hat{x}_{k|k-1}
\]

The objective in this subsection is to prove Theorem 2.4, which says that \( \mathbb{E}[|f_k|^2] \) is always bounded, when using our choice of \( l_{i,k} \) in (11) and sufficiently high bit rates for all sensors (or sufficiently large \( N_i \)). In order to do this, we will first need to prove some preliminary results.

Define also the estimation error at the local sensors as

\[
f_{i,k} = x^o_{i,k} - \hat{x}_{i,k|k-1}
\]

One of the results we will show is that \( \mathbb{E}[|f_{i,k}|^2] \) is bounded for sensor \( i \) when \( N_i \) is sufficiently large. Similar to [5], [17], the approach used to prove this is as follows: instead of showing directly that \( \mathbb{E}[|f_{i,k}|^2] \) is bounded, we show instead that an upper bound to \( \mathbb{E}[|f_{i,k}|^2] \), given by \( |f_{i,k}|^2 \), is bounded, where \( ||\cdot,\cdot||_s \) is defined as

\[
||X, L||_s = \sqrt{\mathbb{E}[L^2 + ||X||^{2+\epsilon}L^{-\epsilon}]}
\]

for some random vector \( X \) and random variable \( L > 0 \), and some \( \epsilon > 0 \). The fact that \( ||X, L||_s \) is bounded for \( ||X||^2 \) is bounded in [5], and further pseudo-norm properties of \( ||\cdot,\cdot||_s \), namely

\[
||dx, dl||_s = d||X,L||_s, \forall d > 0
\]

\[
||X_1 + X_2, L_1 + L_2||_s \leq ||X_1, L_1||_s + ||X_2, L_2||_s
\]

are proved in [17]. We first have the following result, which is similar to Lemma 5.2 of [5], but for the uniform quantizer of [24].

\[
\text{Lemma 2.2:} \text{ Let } X \in \mathbb{R}, L > 0 \text{ be random variables with } \mathbb{E}[X]^{2+\epsilon} < \infty \text{ for some } \epsilon > 0. \quad \text{Suppose } \gamma(N) = \sqrt{\ln N} \text{ in (12). Then for } N \geq 4 \text{ and the uniform quantizer of [24],}
\]

\[
||X - Lq \left( \frac{X}{L} \right), Lq(\Omega)||^2_s \leq \max \left( 2\beta(N)^2, \frac{3}{(\ln N)^{\epsilon/2}} \right)||X, L||^2
\]

\[
\text{Proof See Appendix B.}
\]

In the case where all moments of \( X \) exist (e.g. if \( X \) is a Gaussian random variable), Lemma 2.2 will hold for any \( \epsilon > 0 \). In particular, for a given \( N \), one can choose \( \epsilon(N) \) such that \( \frac{3}{(\ln N)^{\epsilon/2}} \leq 2\beta(N)^2 \), so one then has for suitably chosen \( \epsilon(N) \) that \( ||X - Lq \left( \frac{X}{L} \right), Lq(\Omega)||^2_s \leq 2\beta(N)^2||X, L||^2 \).

We now have the stability result for a single sensor.

\[
\text{Theorem 2.3:} \text{ Suppose that for the uniform quantizer } N_i \text{ is sufficiently large that (10) and}
\]

\[
||A_{i,i}(I - K_{i,C_{1,i}})|| + ||A_{i,i}K_{i,i}||C_{1,i}||\sqrt{2}\beta(N_i) < 1 \quad (18)
\]

are satisfied, where \( K_{i,i} \) is the steady state value of \( K_{i,k} \) (which exists if (10) is satisfied), and \( \beta(N_i) = \frac{2\sqrt{\ln N_i}}{N_i} \). Then \( \mathbb{E}[|f_{i,k}|^2] \) is bounded \( \forall k \).

\[
\text{Proof See Appendix C.}
\]

Comparing (18) and (16) we see that (18) is a more stringent condition than (16) that \( N_i \) needs to meet in order to guarantee stability for a single sensor.

Finally, for the case of multiple sensors we have the following stability result.

\[
\text{Theorem 2.4:} \text{ Let } N_i, i = 1, \ldots, M \text{ be such that for each sensor } i, \mathbb{E}[|f_{i,k}|^2] \text{ is bounded } \forall k. \text{ Then } \mathbb{E}[|f_k|^2] \text{ is bounded } \forall k.
\]

\[
\text{Proof See Appendix D.}
\]

D. Asymptotic analysis

The quantity \( P_k \) in (9) can be regarded as an approximation to the true error covariance. In this subsection we will determine the asymptotic behaviour of \( tr(P_{\infty}) \) for large \( N_i, i = 1, \ldots, M \) (corresponding to the situation of high rate quantization at all sensors), where \( P_{\infty} \) is the limit of \( P_k \) as \( k \to \infty \), that satisfies the equation

\[
P_{\infty} = AP_{\infty}AT + \Sigma_U - AP_{\infty}C^T(\Sigma_P \Sigma_U + \Sigma_v + \Sigma_n)^{-1}C^TP_{\infty}A^T.
\]

In (19), \( \Sigma_n = diag(\Sigma_1, \ldots, \Sigma_M.n) \) where \( \Sigma_i.n = \delta_{N_i}(C_{i,1}P_{i,\infty}C_{i,1}^T + \Sigma_{i,0} + \Sigma_{i,v}) \) and \( P_{i,\infty} \) is the steady state value of \( P_{i,k} \) that satisfies

\[
P_{i,\infty} = A_{i,i}P_{i,\infty}A_{i,i}^T + D_i, \Sigma_u D_i^T - A_{i,i}P_{i,\infty} \times C_{i,1}(C_{i,1}P_{i,\infty}C_{i,1}^T + \Sigma_{i,0} + \Sigma_{i,v})^{-1}C_{i,1}P_{i,\infty}A_{i,i}^T.
\]

1) Scalar systems: We first give the result for scalar systems (where the system state and all sensor measurements are scalar), which can be derived in a more direct manner than for vector systems. We will call \( A = a, C_i = c_i, \Sigma_u = \sigma_u^2, \Sigma_{i,v} = \sigma_{i,v}^2, \Sigma_i = \delta_{N_i}C_{i,1}\), \( i, \ldots, M \) here. For scalar systems, the equation (19) can be solved for \( P_{i,\infty} \), to obtain

\[
P_{\infty} = \frac{\left(1 - \sigma_u^2\sum_{i=1}^{M} \frac{c_i^2}{\sigma_{i,v}^2 + \sigma_i^2} - a^2\right)}{2\sum_{i=1}^{M} \frac{c_i^2}{\sigma_{i,v}^2 + \sigma_i^2}}
\]

\[
+ \frac{(1 - \sigma_u^2\sum_{i=1}^{M} \frac{c_i^2}{\sigma_{i,v}^2 + \sigma_i^2} - a^2)^2 + 4\sigma_u^2\sum_{i=1}^{M} \frac{c_i^2}{\sigma_{i,v}^2 + \sigma_i^2}}{2\sum_{i=1}^{M} \frac{c_i^2}{\sigma_{i,v}^2 + \sigma_i^2}}
\]

(21)

The asymptotic behaviour of (21) is given by the following:

\[
\text{Lemma 2.5:} \text{ For scalar systems, we have}
\]

\[
P_{\infty} = P^{SF} + \sum_{i=1}^{M} \frac{s_i(\xi_i)P^{SF}}{2}E_1^2 \delta_{N_i}
\]

\[
+ \sum_{i,j} O(\delta_{N_i} \delta_{N_j})
\]

(22)
where \( s_i = c_i^2 / \sigma_i^2, ~ E_i = (1 - a^2 - \sigma_i^2 \sum_{j=1}^{M} s_j)^2 + 4\sigma_i^2 \sum_{j=1}^{M} s_j \), and

\[
P_{\infty}^{bf} = \frac{-(1 - \sigma_i^2 \sum_{j=1}^{M} s_j - a^2)}{2 \sum_{j=1}^{M} s_j} \left[ \sqrt{(1 - \sigma_i^2 \sum_{j=1}^{M} s_j - a^2)^2 + 4\sigma_i^2 \sum_{j=1}^{M} s_j} \right]
\]

is the steady state error covariance when there is no quantization (which can be obtained by solving (7)).

**Proof**  The proof is omitted, but can be derived in a similar manner to [27].

2) **Vector systems:** In the scalar case the analytical expression (21) for \( P_{\infty} \) can be derived and analyzed to find asymptotic approximations. However, in the vector case we do not have a closed form expression for either \( P_{\infty} \) or \( \text{tr}(P_{\infty}) \). Instead we will use a different technique, based on the method used to find asymptotic solutions to algebraic equations in perturbation theory (see e.g. [28]), but extended to matrices. With this technique, we can in fact derive an asymptotic expression for the whole matrix \( P_{\infty} \), and not just its trace.

Notation: We will call a matrix \( O(\cdot) \) if all its entries are \( O(1) \), and call a matrix \( O(\varepsilon) \) if all its entries are \( O(\varepsilon) \).

Motivated by the asymptotic result (22) for scalar systems, we assume that \( P_{\infty} \) takes the form

\[
P_{\infty} = \Phi_0 + \sum_{i=1}^{M} \delta N_i \Phi_{1,i} + \sum_{i,j} O(\delta N_i \delta N_j)
\]  

(23)

where \( \Phi_0, \Phi_{1,i}, i = 1, \ldots, M \) are matrices not dependent on \( N_i \). Substituting (23) into (19) we obtain

\[
\Phi_0 + \sum_{i=1}^{M} \delta N_i \Phi_{1,i} + \ldots = A(\Phi_0 + \sum_{i=1}^{M} \delta N_i \Phi_{1,i} + \ldots) A^T
\]

\[
+ \Sigma_w - A(\Phi_0 + \sum_{i=1}^{M} \delta N_i \Phi_{1,i} + \ldots) C^T
\]

\[
\times (C(\Phi_0 + \sum_{i=1}^{M} \delta N_i \Phi_{1,i} + \ldots)C^T + \Sigma_v + \Sigma_n)^{-1}
\]

\[
\times C(\Phi_0 + \sum_{i=1}^{M} \delta N_i \Phi_{1,i} + \ldots) C^T
\]

(24)

We will need to further simplify (24) in order to solve for \( \Phi_0 \) and \( \Phi_{1,i}, i = 1, \ldots, M \). First, we have the following lemma, which is a generalization of a result from p.26 of [28]:

**Lemma 2.6:** Suppose that the matrix \( A \) is invertible and

\[
\| \sum_{i=1}^{M} \epsilon_i A^{-1} B_i \| < 1
\]

Then as \( \epsilon_i \to 0, i = 1, \ldots, M \),

\[
(\sum_{i=1}^{M} \epsilon_i B_i)^{-1} = A^{-1} - \sum_{i=1}^{M} \epsilon_i A^{-1} B_i A^{-1} + \sum_{i,j} O(\epsilon_i \epsilon_j)
\]

**Proof**  See Appendix E.

Next, from the asymptotic analysis of the single sensor case, which may be found in [16], we have \( P_{\infty} = P_{\infty}^{bf} + O(\delta N_i) \), and hence

\[
\Sigma_n = \text{diag}(\Sigma_{1,n}, \ldots, \Sigma_{M,n})
\]

\[
= \text{diag}(\delta N_1(C_{1,1} P_{\infty}^{bf} C_{1,1}^T + \Sigma_{1,v}) + O(\delta^2 N_1)), \ldots,
\]

\[
\delta N_M(C_{1,M} P_{\infty}^{bf} C_{1,M}^T + \Sigma_{M,v}) + O(\delta^2 N_M))
\]

\[
= \sum_{i=1}^{M} \delta N_i F_i + \sum_{i=1}^{M} O(\delta^2 N_i)
\]

where \( F_i \) is a diagonal matrix with \( i \)-th diagonal entry equal to \( C_{1,i} P_{\infty}^{bf} C_{1,i}^T + \Sigma_{i,v} \).

Then

\[
(C(\Phi_0 + \sum_{i=1}^{M} \delta N_i \Phi_{1,i} + \ldots) C^T + \Sigma_v + \Sigma_n)^{-1}
\]

\[
= (C\Phi_0 C^T + \Sigma_v + \sum_{i=1}^{M} \delta N_i (C\Phi_0 C^T + F_i) + \ldots)^{-1}
\]

\[
= (C\Phi_0 C^T + \Sigma_v)^{-1} - \sum_{i=1}^{M} \delta N_i (C\Phi_0 C^T + \Sigma_v)^{-1} C(\Phi_0 C^T + F_i)(C\Phi_0 C^T + \Sigma_v)^{-1} \times (C\Phi_1 C^T + F_i)(C\Phi_0 C^T + \Sigma_v)^{-1} - \ldots
\]

where the last line follows from Lemma 2.6. Note that since \( \delta N_i \to 0 \) as \( N_i \to \infty \), the norm condition in Lemma 2.6 is satisfied for \( N_i \) sufficiently large.

We can thus rewrite (24) as

\[
\Phi_0 + \sum_{i=1}^{M} \delta N_i \Phi_{1,i} + \ldots = A(\Phi_0 + \sum_{i=1}^{M} \delta N_i \Phi_{1,i} + \ldots) A^T
\]

\[
+ \Sigma_w - A(\Phi_0 + \sum_{i=1}^{M} \delta N_i \Phi_{1,i} + \ldots) C^T[(C\Phi_0 C^T + \Sigma_v)^{-1}
\]

\[
= -\sum_{i=1}^{M} \delta N_i (C\Phi_0 C^T + \Sigma_v)^{-1} C(\Phi_1 C^T + F_i)(C\Phi_0 C^T + \Sigma_v)^{-1} \times (C\Phi_1 C^T + F_i)(C\Phi_0 C^T + \Sigma_v)^{-1} - \ldots
\]

\[
\Phi_0 = A\Phi_0 A^T - \Sigma_w - A\Phi_0 C^T(C\Phi_0 C^T + \Sigma_v)^{-1} C\Phi_0 A^T
\]

This is the same equation as (7) satisfied by \( P_{\infty}^{bf} \), and thus \( \Phi_0 = P_{\infty}^{bf} \).

Equating the \( O(\delta N_i) \) terms in (25), we have for each \( i \):

\[
\Phi_{1,i} = A\Phi_0 A^T - \Phi_{1,i} A^T(C\Phi_0 C^T + \Sigma_v)^{-1} C\Phi_0 A^T
\]

\[
- A\Phi_0 C^T(C\Phi_0 C^T + \Sigma_v)^{-1} C\Phi_1 A^T
\]

\[
- C(\Phi_0 C^T + F_i)(C\Phi_0 C^T + \Sigma_v)^{-1} C\Phi_0 A^T
\]

\[
= (A - A\Phi_0 C^T(C\Phi_0 C^T + \Sigma_v)^{-1} C\Phi_1 A^T)
\]

\[
\times (A - A\Phi_0 C^T(C\Phi_0 C^T + \Sigma_v)^{-1} C\Phi_0 A^T)
\]

(26)
Hence, asymptotically $P_{\infty}$ behaves like $P_{\infty} = P_{\infty}^b + \sum_{i=1}^{M} \delta N_i \Phi_{1,i} \sum_{i,j} O(\delta N_i \delta N_j)$, where $P_{\infty}^b$ is the unquantized steady state error covariance that can be found numerically by solving the algebraic Riccati equation (7), and $\Phi_{1,i}, i = 1, \ldots, M$ can be found numerically by solving the Lyapunov equations (26).

E. A rate allocation problem

Suppose we are given a total rate $R_{tot}$, where $R_{tot}$ is large. We wish to determine how this total rate is to be allocated amongst the sensors, where the rate of each sensor $R_i$ is defined as $R_i = \log_2(N_i)$. One way to allocate the rates is to minimize the trace of the asymptotic expression $P_{\infty}^b + \sum_{i=1}^{M} \delta N_i \Phi_{1,i}$ derived in the previous subsection, subject to a total rate constraint. We then have for uniform quantization the discrete optimization problem:

$$\min_{R_1, \ldots, R_M} \text{tr}(P_{\infty}^b) + \sum_{i=1}^{M} e_i R_i \text{ s.t. } \sum_{i=1}^{M} R_i = R_{tot}$$

(27)

where $Z^+ = \{1, 2, \ldots\}$ and $e_i \equiv \frac{4n^2}{\delta} \text{tr}(\Phi_{1,i})$. For the case of increasing $M$, such problems are in general NP-hard. On the other hand, consider the case where the number of sensors $M$ is fixed but $R_{tot}$ is large. As an upper bound on the complexity, suppose we solve problem (27) by exhaustive search. Then the number of $M$-tuples $(R_1, \ldots, R_M)$ that need to be tested is given by the number of compositions of $R_{tot}$ into $M$ parts (see e.g. [29]), which is equal to

$$\binom{R_{tot} - 1}{M - 1} \leq \binom{R_{tot} - 1}{M - 1} M^{-1} = O(R_{tot}^M)$$

Thus in the case where the number of sensors $M$ is fixed but $R_{tot}$ is large, the complexity of problem (27) is polynomial in $R_{tot}$.

One can also attempt to relax problem (27) as follows. Let $R_i = \alpha_i R_{tot}$ where $0 \leq \alpha_i \leq 1$, and $R_i$ is not constrained to be integer valued. We then have the relaxed problem:

$$\min_{\alpha_1, \ldots, \alpha_M} \text{tr}(P_{\infty}^b) + \sum_{i=1}^{M} e_i \alpha_i R_{tot} \text{ s.t. } \sum_{i=1}^{M} \alpha_i = 1, \alpha_i \geq 0$$

(28)

However, this problem is still a non-convex optimization problem. Nevertheless, some numerical results for problems (27) and (28) will be presented in Section IV.

III. SYSTEMS WITH VECTOR MEASUREMENTS

In this section we briefly describe how our results can be extended to systems with vector measurements. We will consider lattice quantizers, which can be regarded as the generalization of the uniform scalar quantizer to vector quantizers. To keep the notation simple, we will only treat the single sensor case here, with the extension to multiple sensors along similar lines to Section II.

A. System model

The system is still the discrete time vector linear system $x_{k+1} = Ax_k + w_k$, with $x_k \in \mathbb{R}^n$, but now the sensor makes a vector measurement $y_k = Cx_k + v_k$, where $y_k \in \mathbb{R}^m$, and $v_k$ is i.i.d. zero mean Gaussian with covariance matrix $\Sigma_v > 0$.

B. Quantized filtering scheme

The equations for the quantized filtering scheme are:

$$\hat{x}_{k|-1} = A\hat{x}_{k-1|-1}$$

$$\hat{x}_k = \hat{x}_{k|-1} + K_k y_k - C\hat{x}_{k|-1}$$

(29)

$$P_{k|-1} = AP_{k-1|-1}A^T + \Sigma_v$$

$$P_{k} = P_{k|-1} - K_k C P_{k|-1}$$

where $K_k = P_{k|-1} C^T (CP_{k|-1} C^T + \Sigma_v + \Sigma_{n,k})^{-1}$, with the scaling factors $l_k$ still taken to be scalar. We will use the shorthand $P_{k+1} \triangleq P_{k+1|k}$. Since $y_k - C\hat{x}_{k|-1}$ is now a vector, we will use vector quantizers $q(\cdot)$ with $N$ quantization values. In general, optimal vector quantization (optimal in terms of minimizing the distortion) is a difficult problem where many open questions remain. The LBG algorithm [30] can be used to find locally optimal vector quantizers but requires numerical methods to compute, and the resulting quantizers often lack structure. We thus consider here the case of lattice vector quantizers [31], whose regular structure makes for efficient encoding and implementation. For scalar measurements, lattice quantization reduces to the case of the uniform scalar quantizer.

Under high rate quantization, we will assume that the quantity $y_k - C\hat{x}_{k|-1}$ is approximately $N(0, CP_k C^T + \Sigma_v)$. We will first diagonalize $CP_k C^T + \Sigma_v$ as

$$CP_k C^T + \Sigma_v = U_k \Lambda_k U_k^T$$

where $U_k$ is a unitary (in fact orthogonal) matrix of eigenvectors and $\Lambda_k$ is a diagonal matrix of eigenvalues (we recall that every real symmetric matrix is diagonalizable, and the eigenvalues of a positive definite matrix are positive). Then the distribution

$$N(0, CP_k C^T + \Sigma_v) = U_k \Lambda_k^{1/2} N(0, I)$$

For zero mean multivariate Gaussian distributions with i.i.d. components, asymptotically optimal lattice quantizers have been considered in [32], with analytical expressions derived for the distortion and sizes of the cells in the lattice quantizer. Thus one way to vector quantize $y_k - C\hat{x}_{k|-1}$ is to first multiply it by $(U_k \Lambda_k^{1/2})^{-1}$ to transform into (approximately) $N(0, I)$ random vectors, quantizing this using the asymptotically optimal lattice quantizers from [32], and then multiplying the quantized vector by $U_k \Lambda_k^{1/2}$, i.e.

$$l_k q_k \left( \frac{y_k - C\hat{x}_{k|-1}}{l_k} \right) = S_k l_k \tilde{q} \left( S_k^{-1} \frac{y_k - C\hat{x}_{k|-1}}{l_k} \right)$$

where $S_k \triangleq U_k \Lambda_k^{1/2}$, and $\tilde{q}(\cdot)$ is the lattice quantizer of [32]. Note that multiplication by $U_k \Lambda_k^{1/2}$ is a linear transformation which preserves the number of values in the codebook. For asymptotically optimal lattice quantization of a Gaussian random vector with i.i.d. components, each having variance $\sigma^2$, the distortion per dimension $D_N \triangleq \frac{1}{m} \text{tr}(E[(x - \tilde{x}(x))^T(x - \tilde{x}(x))])$ is given by (see [32]):

$$D_N \approx \frac{M(S_0) V^{2/m} \frac{2}{m} \ln N}{n^2} N^{2/m} \approx \delta_N$$
where $m$ represents the dimension of the vector to be quantized, $N$ the number of quantization values,

$$
\eta = \frac{1}{\sigma} \sqrt{\int \frac{\Gamma(3/2)}{\Gamma(1/2)}} = \frac{1}{\sigma} \sqrt{\frac{\Gamma}{2}} V = \frac{(1/2)^m}{\Gamma(m/2 + 1)} = \frac{\pi^{m/2}}{\Gamma(m/2 + 1)}
$$

$S_0$ is a Voronoi cell of the lattice, $v(S_0)$ is the volume of $S_0$, and

$$
M(S_0) = \frac{1}{v(S_0)} \int_{S_0} |x - y|^{3/2} dx
$$

is the normalized moment of inertia of $S_0$. The asymptotically optimal scaling $a_N$ of the Voronoi cells of the lattice is given by:

$$
a_N \sim \sqrt{\frac{2}{m \eta}} \left( \frac{V}{v(S_0)} \right)^{1/m} \frac{\ln N}{N^{1/m}}.
$$

Since the components are i.i.d., if we assume that the quantization errors are spread evenly amongst all components, then $E[(x - \hat{q}(x)(x - \bar{q}(x))^T) = \delta_N I$, and the term $\Sigma_{n,k}$ in (29) is then defined as

$$
\Sigma_{n,k} = U_k \Lambda_k^{1/2} (\delta_N I) \Lambda_k^{1/2} U_k^T = \delta_N U_k \Lambda_k U_k^T = \delta_N (C P_k CT + \Sigma_n).
$$

Thus the recursion for $P_k$ can also be written as

$$
P_{k+1} = A P_k A^T + \Sigma_w - \frac{A P_k C T (C P_k C T + \Sigma_w)^{-1} C P_k A^T}{1 + \delta_N}
$$

where now $\delta_N = \frac{M(S_0) V^2/m}{\bar{v}(S_0) \ln N}$. In this case, $P_k$ converges to a steady state value $P_\infty$ satisfying the modified algebraic Riccati equation

$$
P_\infty = A P_\infty A^T + \Sigma_w - \frac{A P_\infty C T (C P_\infty C T + \Sigma_w)^{-1} C P_\infty A^T}{1 + \delta_N}
$$

if

$$
\frac{1}{1 + \delta_N} > 1 - \frac{1}{\prod |\lambda_k^y(A)|^2}
$$

where $\lambda_k^y(A)$ are the unstable eigenvalues of $A$, with the condition (31) being tight if $C$ has rank one.

Define also

$$
\kappa(\omega_k) = \begin{cases} 
\beta(N), & \omega_k \in \{ \text{inner points} \} \\
\gamma(N), & \omega_k \in \{ \text{boundary points} \}
\end{cases}
$$

where the “boundary points” are the quantizer points lying on the boundary. $\beta(N)$ is now defined as

$$
\beta(N) = \sqrt{\frac{2}{m \eta}} \left( \frac{V}{v(S_0)} \right)^{1/m} \frac{\ln N}{N^{1/m}} d_{max},
$$

where $d_{max}$ is the maximum distance from any point in a Voronoi cell to its centroid (before applying the scaling $a_N$), and $\gamma(N) = \sqrt{\ln N}$.

The scaling factors are updated as follows:

$$
l_k = |S_k^{-1}|(||C|| |\hat{l}_k + d_v||
$$

$$
l_{k+1} = \left[ |A(I - K_k)C||\hat{l}_k + d_v||AK_k||d_v
\right.
$$

$$
+ \left. |AK_k S_k||S_k^{-1}||(|C||\hat{l}_k + d_v)\kappa(\omega_k)|
\right]
$$

where

$$
d_v = \frac{1}{||S_k^{-1}||} \left( |A(I - K)| - ||AK||S||S^{-1}||C||\beta(N)|| - ||C||d_v
\right)
$$

$$
1 - ||A(I - K)|| + ||AK||C||C||
$$

Remark 3.1: The expressions for the asymptotic distortion and asymptotically optimal scaling clearly depends on the choice of lattice, or equivalently the shape of the Voronoi cell $S_0$. However, the optimal shapes for $S_0$ are generally not known. As an example, suppose $S_0$ is an $m$-dimensional cube of length 1, i.e. a cubic lattice quantizer. Then one can easily compute that $v(S_0) = 1$ and $M(S_0) = 1/12$. Therefore in this case we have (with $\sigma^2 = 1$):

$$
D_N \sim \frac{\pi}{3m(\Gamma(m/2 + 1))^{2/m}} \frac{\ln N}{N^{2/m}},
$$

$$
a_N \sim \frac{2\sqrt{\pi}}{\sqrt{m(\Gamma(m/2 + 1))^{1/m}}} N^{1/m}.
$$

For dimension $m = 1$, these expressions further simplify to $D_N \sim \frac{4\ln N}{N}$ and $a_N \sim \frac{4\sqrt{\ln N}}{N}$, which are the same expressions as for the uniform quantizer in Section II-B.

C. Stability of quantized filtering scheme

In this subsection we prove the stability of the quantized filtering scheme. Define the estimation error $f_k = x_k - \hat{x}_{k|k-1}$. We have the following preliminary result:

**Lemma 3.1:** Let $X$ be a random vector and $L > 0$ a random variable with $\mathbb{E}[||X||^2 + \epsilon < \infty$ for some $\epsilon > 0$. Let $\gamma(N) = \sqrt{\ln N}$ in (32). Then for the lattice vector quantizers of [32],

$$
||X - Ld_k(X/L), L\kappa(\Omega)||_2^* \leq N(2\beta(N)^2 + \frac{c}{(\ln N)^{1/2}}) ||X, L||^2
$$

where $c$ is a constant that depends on $N$ and $m$.

**Proof** See Appendix F. ■

We have the following stability result.

**Theorem 3.2:** Suppose that for the lattice vector quantizer of [32] $N$ is sufficiently large that condition (31) and

$$
||A(I - KC)|| + ||AK||C||C||\beta(N) < 1
$$

are satisfied, where $K$ is the steady state value of $K_k$, and

$$
\beta(N) = \sqrt{\frac{2}{m \eta}} \left( \frac{V}{v(S_0)} \right)^{1/m} \frac{\ln N}{N^{1/m}} d_{max},
$$

Then $\mathbb{E}[||f_k||^2]$ is bounded $\forall k$.

**Proof** The proof is similar to the proof of Theorem 2.3, but making use of Lemma 3.1 instead of Lemma 2.2. The details are omitted for brevity. ■

D. Asymptotic analysis

The technique used for asymptotic analysis of $P_\infty$ satisfying the expression

$$
P_\infty = A P_\infty A^T + \Sigma_w - \frac{A P_\infty C T (C P_\infty C T + \Sigma_w)^{-1} C P_\infty A^T}{1 + \delta_N}
$$

will be the same as in Section II-D, and we thus only state the final result. Asymptotically we have $P_\infty = P_k^{\infty f} + \delta_N \Phi_1 + O(\delta_N^{2}),$ where $\delta_N$ decays to zero at the rate $\frac{1}{\sqrt{m \ln N}}$ for the lattice quantizer of [32]. $P_k^{\infty f}$ can be found by solving numerically the algebraic Riccati equation

$$
P_k^{\infty f} = A P_k^{\infty f} A^T + \Sigma_w - A P_k^{\infty f} C T (C P_k^{\infty f} C T + \Sigma_w)^{-1} C P_k^{\infty f} A^T,
$$

Even for lattice quantization of uniformly distributed random vectors, the optimal cell shapes are only known for dimensions $m = 1, 2, 3$, see [31].
and $\Phi_1$ can be found by solving numerically the Lyapunov equation (with $\Phi_0 = P_\infty^0$).

$$\Phi_1 = (A(I-KC))\Phi_1(A(I-KC))^T + A\Phi_0 C^T (C\Phi_0 C^T + \Sigma_v)^{-1} C\Phi_0 A^T$$

IV. NUMERICAL STUDIES

A. Scalar measurements

We first consider a two sensor situation, with parameters

$$A = \begin{bmatrix} 1.2 & 0 & 0 \\ 0 & 1.1 & 0 \\ 0 & 0 & 0.6 \end{bmatrix}, \quad \Sigma_w = I, \quad C_1 = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix},$$

$$\Sigma_{1,v} = 1, \quad C_2 = \begin{bmatrix} 0 & 1 & 1 \end{bmatrix}, \quad \Sigma_{2,v} = 0.3.$$  

One can easily verify that the sensor pairs $(A, C_i), i = 1, 2$ are not detectable. In the observability decompositions, we choose

$$T_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad A_{1,1} = \begin{bmatrix} 1.2 & 0 \\ 0 & 0.6 \end{bmatrix},$$

$$C_{1,1} = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad D_{1,1} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},$$

$$T_2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad A_{1,2} = \begin{bmatrix} 1.1 & 0 \end{bmatrix},$$

$$C_{1,2} = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad D_{1,2} = \begin{bmatrix} 0 & 1 \end{bmatrix}.$$  

In Fig. 2 we plot the results from Monte Carlo simulations of the trace of the true error covariance $\text{tr}\mathbb{E}[(x_k - \hat{x}_k|k-1)(x_k - \hat{x}_k|k-1)^T]$, together with $\text{tr}(P_\infty)$ and the asymptotic expression for $\text{tr}(P_\infty)$, for different values of $N_1 = N_2 = N$. We see that the asymptotic approximations to $P_\infty$ become more accurate as $N$ increases.

![Fig. 2](image)

Fig. 2. Error covariance and asymptotic expression: Sensor pairs $(A, C_i)$ not detectable.

We next consider the rate allocation problem (27) of Section II-E, with $R_{tot} = 8$. We consider a two sensor situation, with parameters

$$A = \begin{bmatrix} 1.2 & 0.2 \\ 0.2 & 0.5 \end{bmatrix}, \quad \Sigma_w = I, \quad C_1 = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad \Sigma_{1,v} = 1, \quad C_2 = \begin{bmatrix} 1 & 1 \end{bmatrix}, \quad \Sigma_{2,v} = 0.3.$$  

In this case the sensor pairs $(A, C_i), i = 1, 2$ are detectable. In Table I we tabulate the results for some integer combinations of $R_1 = \log_2(N_1)$ and $R_2 = \log_2(N_2)$, with $R_1 + R_2 = 8$. We see that $R_1 = 3, R_2 = 5$ gives the best performance in terms of both the theoretical approximation $P_\infty$ and performance from Monte Carlo simulations. Solving the relaxed non-convex problem (28) using the `NMinimize` routine in Mathematica gives the solution $\alpha_1^* = 0.3436, \alpha_2^* = 0.6564$, corresponding to rates $R_1^* = 2.7486, R_2^* = 5.2514$.

B. Vector measurements

We next consider the case of a single sensor with vector (2-dimensional) measurements, with parameters

$$A = \begin{bmatrix} 1.2 & 0.2 \\ 0.2 & 0.5 \end{bmatrix}, \quad \Sigma_w = I, \quad C = \begin{bmatrix} 1 & 0.2 \\ 0.2 & 1 \end{bmatrix}, \quad \Sigma_v = I.$$  

We give results for the hexagonal lattice quantizer, which is known to be a good quantizer in two dimension (in fact optimal for quantizing uniformly distributed random vectors [31]). If successive lattice points in the hexagonal lattice have distance 1, then the Voronoi region $S_0$ is a regular hexagon of length $1/\sqrt{3}$, and thus $v(S_0) = \sqrt{3}/2$. The normalized moment of inertia can also been computed (see [31]) as $M(S_0) = \frac{360\sqrt{3}}{3}$. In Fig. 3 we give plots of the error covariances for various values of $N$.

![Fig. 3](image)

Fig. 3. Error covariance and asymptotic expression: Hexagonal lattice quantizer.

V. CONCLUSION

In this paper we have considered some quantized filtering schemes for multi-sensor linear state estimation. We have shown their stability under sufficiently high bit rates, and derived asymptotic approximations to the error covariance for...
linear state estimation of discrete time linear systems with quantized innovations, valid when the sensors use high rate quantization. Areas of future research include the study of the effects of random packet losses [33] in addition to high rate quantization, and analyzing the performance of quantized filtering schemes at rates close to the minimum data rates of [5].

APPENDIX

A. Proof of Lemma 2.1

Since \( \omega_{i,k} \in \{2, \ldots, N_i - 1\} \), \( \forall k \), we have \( \kappa(\omega_{i,k}) = \beta(N_i), \forall k \). Hence (11) becomes

\[
\tilde{l}_{i,k+1} = ||A_{i,k}(I - K_{i,k}C_{i,i})||\tilde{l}_{i,k} + d_{i,w} + ||A_{i,k}K_{i,k}||d_{i,v} + ||A_{i,k}K_{i,k}|| ||C_{i,i}||\tilde{l}_{i,k} + d_{i,v}\beta(N_i)
\]

\[
= ||A_{i,k}(I - K_{i,k}C_{i,i})|| + ||A_{i,k}K_{i,k}|| ||C_{i,i}|| \beta(N_i)
\]

Now if

\[
\lim_{k \to \infty} \left( ||A_{i,k}(I - K_{i,k}C_{i,i})|| + ||A_{i,k}K_{i,k}|| ||C_{i,i}|| \beta(N_i) \right)
\]

then \( \tilde{l}_{i,k} \) converges to the solution \( \tilde{l} \) of the equation

\[
\tilde{l} = ((A_{i,k}(I - K_{i,k}C_{i,i})|| + ||A_{i,k}K_{i,k}|| ||C_{i,i}|| \beta(N_i))\tilde{l} + d_{i,w} + ||A_{i,k}K_{i,k}||d_{i,v}(1 + \beta(N_i))
\]

or

\[
\tilde{l}_{i}||C_{i,i}|| + d_{i,v}
\]

Then

\[
l_{i,k}||C_{i,i}|| + d_{i,v}
\]

by using the definition of \( d_{i,v} \) in the last line. Hence

\[
l_{i,k} = \frac{||C_{i,i}|| + d_{i,v}}{\sigma_i} \rightarrow \frac{||C_{i,i}||\tilde{l}_{i} + d_{i,v}}{\sigma_i} = 1.
\]

B. Proof of Lemma 2.2

Let \( \phi = \lambda \kappa(\omega) \) and \( \omega \) be the index of the quantizer point \( q(x/l) \). If \( 2 \leq \omega \leq N - 1 \), then the interval which contains \( x/l \) has length \( 2\kappa(\omega) = 2\phi \). Thus \( |x - \bar{q}(x/l)| = |x/l - \bar{q}(x/l)| < \phi \) for \( \forall x \in [2, \ldots, N - 1] \).

If \( \omega \in \{1, N\} \), then

\[
\mathbb{E} \left[ \frac{X - L\bar{q}(X/L)}{\Phi} \right]^{2+e} \leq \phi^2 + \kappa(\omega)^{-e} \mathbb{E} \left[ \frac{|X|^{2+e}}{L'} \right] \mathbb{E} \left[ \frac{|\omega|}{L'} \right]
\]

where the inequality comes from using similar arguments as in the proof of Lemma 5.2 in [5]. Averaging over \( \Omega \) and \( L \), we obtain

\[
\mathbb{E} \left[ \frac{X - L\bar{q}(X/L)}{\Phi} \right]^{2+e} \leq \mathbb{E}[\Phi^2] + \gamma(N)^{-e} \mathbb{E} \left[ \frac{|X|^{2+e}}{L'} \right]
\]

Next, we have

\[
\mathbb{E}[\Phi^2|l] = \mathbb{E}[(L\kappa(\omega))^2|l]
\]

Finally, using \( \gamma(N) = \sqrt{\ln N} \), we have

\[
||X - L\bar{q}(X/L), L\kappa(\Omega)||^2 = \mathbb{E}[\Phi^2] + \mathbb{E} \left[ \frac{|X - L\bar{q}(X/L)|^{2+e}}{\Phi} \right]
\]

\[
\leq \mathbb{E}[\Phi^2] + \mathbb{E}[\Phi^2] + \gamma(N)^{-e} \mathbb{E} \left[ \frac{|X|^{2+e}}{L'} \right]
\]

\[
\leq 2 \left( \beta(N)^2 \mathbb{E}[L^2] + \frac{\gamma(N)^2}{(2N-2)\sqrt{\ln N}} \mathbb{E} \left[ \frac{|X|^{2+e}}{L'} \right] \right)
\]

by using the definition of \( d_{i,v} \) in the last line. Hence

\[
l_{i,k} = \frac{||C_{i,i}|| + d_{i,v}}{\sigma_i} \rightarrow \frac{||C_{i,i}||\tilde{l}_{i} + d_{i,v}}{\sigma_i} = 1.
\]
where in the third inequality we used the fact that
\[
\frac{1}{(2(N-2))^2} \leq \frac{1}{(2(N-2))^2} \leq 1 \quad \text{for } N \geq 4
\]

C. Proof of Theorem 2.3

First rewrite (8) as
\[
\hat{x}_{i,k+1|k} = A_{i,k} \hat{x}_{i,k|k-1} + A_{i,k} K_{i,k} \sigma_{i,k} l_{i,k} q_i \left( \frac{y_{i,k} - C_i \hat{x}_{i,k|k-1}}{\sigma_{i,k} l_{i,k}} \right)
\]

From (34) and (3) we can then derive that
\[
f_{i,k+1} = A_{i,k} f_{i,k} + D_{i,k} w_{k} - A_{i,k} K_{i,k} \sigma_{i,k} l_{i,k} q_i \left( \frac{C_i f_{i,k} + v_{i,k}}{\sigma_{i,k} l_{i,k}} \right)
\]
\[
= A_{i,k} \left( I - K_{i,k} C_{i,k} \right) f_{i,k} + D_{i,k} w_{k} - A_{i,k} K_{i,k} v_{i,k}
\]
\[
+ A_{i,k} K_{i,k} \sigma_{i,k} l_{i,k} \left[ \frac{C_i f_{i,k} + v_{i,k}}{\sigma_{i,k} l_{i,k}} \right] - l_{i,k} q_i \left( \frac{C_i f_{i,k} + v_{i,k}}{\sigma_{i,k} l_{i,k}} \right)
\]

As stated previously, rather than showing directly that \( E[\|f_{i,k}\|^2] \) is bounded \( \forall k \), we show instead that the upper bound \( \|f_{i,k} + l_{i,k}\|^2 \) is bounded \( \forall k \). From the definition of \( l_{i,k} \) and \( \tilde{l}_{i,k} \), and using some of the pseudo-norm properties (17), we obtain
\[
\|f_{i,k} + l_{i,k}\|^2 \leq \|A_{i,k} \left( I - K_{i,k} C_{i,k} \right) f_{i,k} + l_{i,k}\|^2 + \|A_{i,k} K_{i,k} \|v_{i,k}\| + \|D_{i,k} w_{k}\| + \|A_{i,k} K_{i,k} \|v_{i,k}\|
\]
\[
+ \|A_{i,k} K_{i,k} \|v_{i,k}\| + \|A_{i,k} K_{i,k} \|v_{i,k}\| + \|A_{i,k} K_{i,k} \|v_{i,k}\|
\]
\[
+ A_{i,k} K_{i,k} \sigma_{i,k} l_{i,k} \left[ \frac{C_i f_{i,k} + v_{i,k}}{\sigma_{i,k} l_{i,k}} \right] - l_{i,k} q_i \left( \frac{C_i f_{i,k} + v_{i,k}}{\sigma_{i,k} l_{i,k}} \right)
\]

Since \( \|A_{i,k} (I - K_{i,k} C_{i,k})\| < 1 \), \( K_{i,k} \) converges to a steady state value \( K_i \), \( \beta(N_i) \to 0 \) as \( N_i \to \infty \), and \( \epsilon > 0 \) can be freely chosen, stability is ensured if \( N_i \) is large enough that
\[
\|A_{i,k} (I - K_{i,k} C_{i,k})\| + \|A_{i,k} K_{i,k}\| |C_{i,k}| |\sqrt{2} \beta(N_i) < 1
\]

D. Proof of Theorem 2.4

From (8)-(9) we have
\[
\hat{x}_{i,k+1|k} = AP_k \left[ P_{i,k-1}^{-1} \hat{x}_{i,k|k-1} + \sum_{i=1}^{M} D_{i,k}^T \left( P_{i,k}^{-1} \hat{x}_{i,k|k-1} \right) \right]
\]
\[
+ K_{i,k} \sigma_{i,k} l_{i,k} q_i \left( \frac{y_{i,k} - C_i \hat{x}_{i,k|k-1}}{\sigma_{i,k} l_{i,k}} \right)
\]
\[
= AP_k \left[ P_{i,k-1}^{-1} \hat{x}_{i,k|k-1} + \sum_{i=1}^{M} D_{i,k}^T \left( P_{i,k}^{-1} \hat{x}_{i,k|k-1} \right) \right]
\]
\[
+ P_{i,k}^{-1} \hat{x}_{i,k|k-1} + P_{i,k}^{-1} \hat{x}_{i,k|k-1} + \hat{x}_{i,k|k-1} + P_{i,k}^{-1} \hat{x}_{i,k|k-1} + P_{i,k}^{-1} \hat{x}_{i,k|k-1}
\]
\[
= AP_k \left[ P_{i,k-1}^{-1} \hat{x}_{i,k|k-1} + \sum_{i=1}^{M} D_{i,k}^T \left( P_{i,k}^{-1} \hat{x}_{i,k|k-1} \right) \right]
\]
\[
+ \hat{x}_{i,k|k-1} + P_{i,k}^{-1} \hat{x}_{i,k|k-1} + P_{i,k}^{-1} \hat{x}_{i,k|k-1} + P_{i,k}^{-1} \hat{x}_{i,k|k-1}
\]

By similar derivations to [3], [20] we can obtain
\[
P_k \left[ P_{i,k-1}^{-1} + \sum_{i=1}^{M} D_{i,k}^T \left( P_{i,k}^{-1} \hat{x}_{i,k|k-1} \right) \right] = P_k P_{i,k}^{-1} = I
\]

Using this, and the definitions \( f_k \triangleq x_k - \hat{x}_{i,k|k-1} \), \( f_k \triangleq x_o - \hat{x}_{i,k|k-1} \), and \( D_{i,k} \triangleq D_{i,k} \hat{x}_{i,k|k-1} \), we have
\[
f_{i,k+1} = AP_k \left[ P_{i,k-1}^{-1} + \sum_{i=1}^{M} D_{i,k}^T \left( P_{i,k}^{-1} \hat{x}_{i,k|k-1} \right) \right] f_{i,k}
\]
\[
+ w_{i,k} = AP_k \left[ P_{i,k-1}^{-1} \hat{x}_{i,k|k-1} \right] - AP_k \left[ \sum_{i=1}^{M} D_{i,k}^T \left( P_{i,k}^{-1} \hat{x}_{i,k|k-1} \right) \right] f_{i,k}
\]
\[
+ w_{i,k} = AP_k \left[ P_{i,k-1}^{-1} \hat{x}_{i,k|k-1} \right] - AP_k \left[ \sum_{i=1}^{M} D_{i,k}^T \left( P_{i,k}^{-1} \hat{x}_{i,k|k-1} \right) \right] f_{i,k}
\]

where we have made use of Lemma 2.2 in the second inequality. Note that the terms \( \|D_{i,k} w_{k}\| \) and \( \|v_{i,k}\| \) can be upper bounded by constants when \( v_{i,k} \) and \( w_{k} \) have uniformly bounded \( (2+\epsilon) \)-th absolute moments, see also [17]. For boundedness of \( \|f_{i,k} + l_{i,k}\| \), we require
\[
\|A_{i,k} (I - K_{i,k} C_{i,k})\| + \|A_{i,k} K_{i,k}\| |C_{i,k}| \max \left( \sqrt{2} \beta(N_i), \frac{\sqrt{3}}{(\ln N_i)^{1/4}} \right) < 1
\]

Since \( \|A_{i,k} (I - K_{i,k} C_{i,k})\| < 1 \), and \( E[\|f_{i,k}\|^2] \) is bounded for each \( i \) by assumption, the result follows.
E. Proof of Lemma 2.6

We have
\[
(A + \sum_{i=1}^{M} \epsilon_i B_i)^{-1} = (A(I + \sum_{i=1}^{M} \epsilon_i A^{-1} B_i))^{-1} = (I - \sum_{i=1}^{M} \epsilon_i A^{-1} B_i + \sum_{i,j} O(\epsilon_i \epsilon_j) A^{-1})
\]

where the third equality holds if \(|\sum_{i=1}^{M} \epsilon_i A^{-1} B_i| < 1.

F. Proof of Lemma 3.1

We give a sketch of the proof. The inequality
\[
\mathbb{E}\left[\frac{||X - Lq(X/L)||^{2+\varepsilon}}{\Phi}\right] \leq \mathbb{E}[\Phi^2] + \gamma(N)^{-\varepsilon} \mathbb{E}\left[\frac{||X||^{2+\varepsilon}}{L^\varepsilon}\right]
\]

can be shown with similar methods as in the proof of Lemma 2.2.

Next, from the scaling \(\frac{X}{N^{1/m}}\), the individual cell sizes will increase at the rate \(O(\frac{\ln N}{N^{1/m}})\), while the number of cells in each dimension is of order \(O(N^{1/m})\). Thus, for some constants \(c_1\) and \(c_2\), we have
\[
\mathbb{E}[\Phi^2] \leq l^2 \beta(N)^2 P \left( \frac{||X||}{L} \leq c_1 \frac{\sqrt{\ln N}}{N^{1/m}} \times N^{1/m} \right)
\]

and hence
\[
\mathbb{E}[\Phi^2] \leq l^2 \beta(N)^2 \mathbb{E}[L^2] + \gamma(N)^2 \frac{\mathbb{E}[||X||^{2+\varepsilon}]}{L^{\varepsilon}}
\]