Distributed tracking with energy management in wireless sensor networks

Nicholas Roseveare, Balasubramaniam Natarajan

How to cite this manuscript

Published Version Information

Distributed Tracking with Energy Management in Wireless Sensor Networks

Nicholas Roseveare, *Student Member IEEE* and Balasubramaniam Natarajan, *Senior Member IEEE*
Department of Electrical and Computer Engineering
Kansas State University
Manhattan, Kansas, 66506
Email: {nickrose,bala}@ksu.edu

Abstract—We consider a wireless sensor network tasked with tracking a process using a set of distributed nodes. Here, multiple remote sensor nodes estimate the physical process (viz., a moving object) and transmit quantized estimates to a fusion center for processing. At the fusion node a BLUE (Best Linear Unbiased Estimation) approach is used to combine the sensor estimates and create a final estimate of the state. In this framework, the uncertainty of the overall estimate is derived and shown to depend on the individual sensor transmit energy and quantization levels, as well as the Kalman tracker uncertainty at the node. Since power and bandwidth are critically constrained resources in battery operated sensor nodes, we attempt to quantify the trade-off between the lifetime of the network and the estimation quality over time. Three different convex formulations of the underlying non-convex Mixed Integer Non-Linear optimization problem are presented. Unlike previous work, this effort incorporates the operating state of the nodes into the decisions of the optimum bits and transmission power levels based on a heuristic. Simulation results for all formulations demonstrate the quality of the state estimate as well as the extended lifetime of the WSN.

Index Terms—distributed tracking, distributed estimation, wireless sensor networks, convex optimization

I. INTRODUCTION

IN wireless sensor network (WSN) applications, a typical pursuit is to take measurements or estimate the state of a physical dynamical process at each of the independent, spatially dispersed, sensors in the network. In the centralized distributed estimation scheme [1], individual sensors perform local processing and forward the data or estimates through a communication network to a fusion node, which combines the reports to form an estimate of the parameter or process. Distributed estimation via wireless sensor networks presents a variety of interesting challenges including the understandable limits of communication bandwidth which introduce quantization error in addition to the error induced by noise in the wireless channel. Another practical constraint is that of the sensing nodes themselves, which have limited battery power. Thus, a WSN estimation paradigm is desired which offers acceptable performance (in mean squared error (MSE) or a similar metric) while also being energy efficient so as to prolong the lifetime of the WSN.

Distributed estimation incorporating the effects quantization and/or communication has received significant interest in the

recent past, and especially in the last decade. Early works [2], [3] typically consider scenarios with spatially distributed processors utilizing linear measurements with knowledge of the joint distribution of the measurement noise. The authors in [1] generalize distributed estimation to nonlinear observations with the similar assumption of partially known statistics. Early work on quantization [4]–[6] uses joint distributions of the measurement noise for efficient estimation in distributed systems while considering noiseless communication. The work in [7] explores sequential signal encoding with power and delay constraints on the distributed estimation framework. The authors in [8] achieve the optimal quantization for distributed estimation based on a training sample in unknown noise statistics. The spatial correlation among sensor measurements is accounted for in the design of quantizers for distributed estimation in [9]. A class of maximum likelihood (ML) estimators of a parameter are proposed in [10] which achieve the estimation performance of the sample mean merely from observations which are quantized to a single bit. Similarly, in [11], ML estimates of a variance parameter from single bit quantizations utilize a sequentially updated adaptive quantization threshold.

A universal decentralized estimator is designed in [12] which utilizes Best Linear Unbiased Estimation (BLUE) without knowledge of the measurement noise statistics. In [13], a BLU estimator is used while considering of the effect of channel noise and measurement noise on the variance of the estimator. Here, an upper bound for the variance is derived which produces a power and rate efficient estimator. Using the result in [13], the authors in [14] find an upper bound on the variance of the BLUE which is used to design an efficient estimator. A rate-constrained distributed estimation scheme is designed in [15] which trades off the total rate used by the WSN with the number of active sensors. Similarly, [16] investigates the trade-off between number of active sensors and the energy used by each sensor. The same authors (in [17]) introduce function-based network lifetime and optimize it to produce a specified estimation accuracy at the fusion node. The authors in [15], [17] both assume distortion free communication. The joint optimal energy allocation and quantization level to minimize error in a binary symmetric channel with non-zero cross over probabilities is analyzed in [18].

Distributed BLU estimators that are utilized in previous work either consider only measurement noise variance or mea-

surement and quantization noise variances. The only previous works that incorporate measurement, quantization, and channel noise variance for use with a BLU estimator are [19] and [13]. The work in [13] considers *scheduling* of sensor energy transmission and quantization levels for local estimation at the sensors from which other control actions are taken. Distributed estimation is not implemented. This estimator follows from a model [19] used to investigate the effect of channel fading on the accuracy of a sensor node.

Prior efforts in distributed tracking are primarily concerned with sensor scheduling and selection algorithms. In [20], the lifetime of the network is optimized by determining how many sensors to keep active. A detailed energy function is used to construct an energy-usage based cost function which is optimized. The authors in [21] formulate the sensor scheduling problem in terms of disjoint set covers of the observation space. Distributed tracking using WSNs is done in [22] where quantization is accomplished by reducing the dimension of the state variable such that the transmit power budget will be met. Channel-aware distributed tracking is accomplished in [23] by performing particle filtering at the fusion node, however, only centralized tracking and transmission power levels are considered. A unique alternative to the above is found in [24] where sensor scheduling is formulated as approximate dynamic programming problem which chooses a leader node and a subset of observation nodes.However, while power constraints and channel noise are considered, quantization noise is not. Unlike the prior works discussed, our work focuses on optimizing the use of network resources *after* scheduling and selection.

Once a subset of the sensors has been selected/scheduled, how does the fusion center instruct the optimal quantization and transmission of the node data? And how does it continue to do this so as to prolong WSN lifetime? These are the questions this paper attempts to add insights to. We consider a time-based resource management problem and, specifically, this work attempts to extend the authors' concept of fairness [25] (with respect to the operating state of each node) and incorporates a heuristic which extends the lifetime of energy-deficient nodes. Prior efforts have put significant focus on the understanding of dependencies of channel conditions, quantization level and measurement noise on ideal sensors for distributed estimation for a single time instance. While some modeling elements are similar, it is of particular distinction to note the differences of this work from [13], [14], [16]. These works consider a total power constraint (or minimizing total power) with the motivation of preventing nodes from over utilizing power. We do not consider the total power, but the minimization error. In addition to this, a total power constraint is looser than individual node power constraints, which we have implemented. These works, however, do not consider a network total bandwidth constraint, or the decoupling of the power and bit decision variables, as we have. It could even be proposed that a total power constraint and a bandwidth constraint are comparable given the correct antecedents. Furthermore, the resulting methods typically are intuitively water-filling-like algorithms; while our method (given the nodes are already scheduled) has the behavior of a reverse-water-filling-
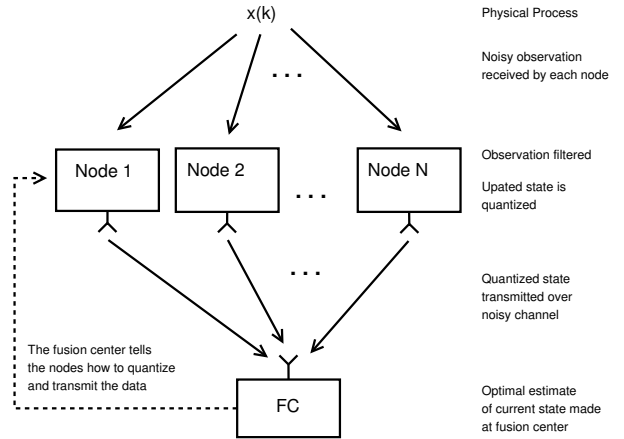

Fig. 1. An illustration of the specified distributed estimation system in which sensor nodes send Kalman filter updates.

like algorithm. An additional novel pursuit of this research with respect to prior similar work is that it accounts for the operating state of the sensor through time and explores the effect of its consideration on WSN lifetime.

We formulate an optimization problem in which distributed estimation takes into account the power and bandwidth constraints of wireless sensors nodes for repeated time instances. This method uses a scalar BLU estimator which is adapted for vector quantities and whose variance represents the effects of process, measurement, quantization, and channel noise. The minimization of this variance objective under the aforementioned constraints is a non-convex Mixed-Integer Non-Linear Program (MINLP). The relaxed (continuous) problem is identified as a difference of convex functions (DC) and approximated by a convex function [26] and solved using Sequential Convex Programming (SCP). An additional convex formulation (not DC) is also detailed, as well as a "worst case" formulation which results in a more rapid solution.

Section II details the setup of the problem of interest while Section III chronicles the formulation of the optimization problem and introduces the heuristic scaling parameter. As simulation results in Section IV demonstrate, the lifetime of the WSN is extended with the inclusion this heuristic, maintaining some menial loss of estimation performance. The trade-off between consistent estimation accuracy and network lifetime is investigated using the heuristic scaling developed herein. We show that network lifetime can be extended to over 250% of the original by an appropriate choice of the heuristic scaling parameter.

II. SYSTEM MODEL

The considered paradigm contains multiple remote sensor nodes with power and bandwidth constraints which filter state estimates and transmit them to a fusion center for processing, as illustrated in Figure 1. It is the responsibility of the fusion node to instruct each node on how to send its update at each time instance via some orthogonal signaling scheme[1]. It is assumed that the fusion node is not energy constrained in its

---

[1]For example, if FDMA is used then a fixed total bandwidth is divided among the nodes. Thus, our method can be thought of as an adaptive bandwidth assignment FDMA scheme.

transmissions and that the energy of receiving a transmission at a sensor node is negligible. As power is not constrained at the fusion node we assume practically noiseless feedback. Parameters which the fusion node uses to control the optimal reporting strategy or that it must take into account include: the accuracy of the state estimate of each node, the level of quantization, any effects of the noisy communications channel, as well as the remaining transmit energy (battery life) of each node. We consider a WSN with N spatially distributed nodes. These nodes take measurements of a dynamic process and update their local state estimates. The state estimate is quantized and transmitted to the fusion node, which estimates the state from the received reports.

A. Sensor Level Kalman Filtering

In these systems, the generalized state space model for the nth sensor is of the form

$$
\begin{align}
\mathbf{x}(k+1) &= \mathbf{f}(\mathbf{x}(k), \mathbf{w}(k+1), k) \tag{1} \\
\mathbf{z}_n(k) &= \mathbf{h}_n(\mathbf{x}(k), k) + \mathbf{v}_n(k) \tag{2}
\end{align}
$$

where $\mathbf{x}(k)$ is the d-dimensional state of the true system at time instance k. The functions $\mathbf{f}$ and $\mathbf{h}_n$ are the generalized state transition and observation functions. The process noise vector $\mathbf{w}(k) \sim \mathcal{N}(0, \mathbf{Q}(k))$ is assumed to be due to disturbances and modeling errors, $\mathbf{z}_n(k)$ is the observation vector and $\mathbf{v}_n(k) \sim \mathcal{N}(0, \mathbf{R}_n(k))$ is the measurement noise at node n, where $\mathbf{Q}(k)$ and $\mathbf{R}_n(k)$ represent the process and measurement noise covariance matrices, respectively. We make the normal simplifying assumptions about the noise processes being zero mean, white, and uncorrelated. The goal for each local sensor is to estimate $\mathbf{x}(k)$.

In this paper, the state transition function modeled in the simulations of Section IV is "nearly constant velocity" (NCV) propagation. The measurement model is a simple position-only observation function. The next section details the estimation of the state from the received Kalman updates.

B. Optimal Estimation from Kalman Updates

After a sensor node has measured the dynamical process and updated its local estimate, it will have obtained a state vector and covariance pair, $\{\hat{\mathbf{x}}_n(k|k), \mathbf{P}_n(k|k)\}$. In practice, both the state estimate and its covariance matrix would be quantized and transmitted to the fusion node where a final estimate is formed. However, to maintain simplicity in presentation, we study the effect of quantizing and transmitting the state information while assuming that the fusion center has error-free knowledge of the covariance structure; this is a common assumption adopted by other authors [13]. While the extension of the analysis to additionally transmitting the covariance structure simply involves sending more data (corruptible by quantization and transmission), we delay this discussion for future work. Instead, we evaluate our approach by testing the sensitivity of it to the knowledge of the covariance information. This is done by adding a random perturbation to the covariance information and executing the forthcoming methods, as shown in Section IV-D. Given the state estimate at the node, the state vector is then quantized as

$$
\mathcal{Q}\{\hat{\mathbf{x}}_n(k|k)\} = \hat{\mathbf{x}}_n(k|k) + \mathbf{n}_n^q(k) \quad \forall \ n = 1, \ldots, N \tag{3}
$$

where $\mathbf{n}_q^n(k) \sim \mathcal{N}(0, \mathbf{R}_n^q(k))$ is the quantization noise of the nth sensor at time k, with $\mathbf{R}_n^q(k)$ the quantization noise covariance. The quantized data is mapped to a bit stream or other form suitable for transmission. Each bit of the data stream consisting of b_n bits is transmitted independently (by means of some orthogonal signaling scheme) through noisy wireless fading channels to the fusion node. The final information received at the sensor fusion node is

$$
\begin{align}
\tilde{\mathbf{x}}_n(k|k) &= \mathcal{Q}\{\hat{\mathbf{x}}_n(k|k)\} + \mathbf{n}_n^c(k) \\
&= \hat{\mathbf{x}}_n(k|k) + \mathbf{n}_n^q(k) + \mathbf{n}_n^c(k), \tag{4}
\end{align}
$$

$\forall \ n = 1, \ldots, N$, where $\mathbf{n}_n^c(k) \sim \mathcal{N}(0, \mathbf{R}_n^c(k))$ is the channel noise of the nth sensor at time k due to imperfect communication, with $\mathbf{R}_n^c(k)$ the channel noise covariance. We make the standard assumption that the internal noise of the state estimate, and the quantization and channel noises are all uncorrelated, since each of these noises could be considered as coming from independent sources.

At the fusion node, received state updates, affected by various noise sources, are combined linearly to form an estimate of the actual process state. The predecessor of this estimation paradigm was introduced in [25] where scalar measurements of a deterministic source are used to make a single estimate at the fusion node. The issue of combining Kalman states by filtering is complicated by common process and measurement noise [27] of the sequential state estimate reports from a particular sensor. The methods typically employed are often practically infeasible [28] and it should be noted that they are avoided in our work by simply estimating the current state from the only the most recently reported state estimates. We extend the simple scalar estimator of our previous work to send the elements of a vector state. In this case we consider each element of the state vector independently and this results in the centralized BLU-like estimator of the elements of $\mathbf{x}(k)$ which are $x^i(k), \ i = 1, \ldots, d$. The BLUE at the fusion node then determines an optimal number of bits and transmitting power for each of the sensor nodes in order to minimize the total variance of the elements of the state vector it is estimating.

We extend the simple scalar estimator in order to send the elements of a vector state. The BLUE method for vector estimates (cf. [29]) assumes the "measurements" (or reports in this case) of the true state are of the form

$$
\tilde{\mathbb{X}}(k) = \mathbf{H}(k)\mathbf{x}(k) + \mathbf{U}(k) \tag{5}
$$

where $\mathbf{H}(k)$ is a linear combining matrix and $\mathbf{U}(k) \sim \mathcal{N}(\mathbf{0}, \mathbb{P}(k))$ represents additive noise. In the notation we have given so far

$$
\begin{align}
\tilde{\mathbb{X}}(k) &= [\tilde{\mathbf{x}}_1^T(k|k) \ \tilde{\mathbf{x}}_2^T(k|k) \ \cdots \ \tilde{\mathbf{x}}_N^T(k|k)]^T \tag{6} \\
\mathbf{H}(k) &= [\mathbf{I}_d \mid \mathbf{I}_d \mid \cdots \mid \mathbf{I}_d]_{d \times Nd}^T \\
\mathbf{U}(k) &= [\mathbf{u}_1^T(k) \ \mathbf{u}_2^T(k) \ \cdots \ \mathbf{u}_N^T(k)]^T
\end{align}
$$

where $\mathbf{I}_d$ is an identity matrix of size $d \times d$. To maintain unbiasedness, the linear combining matrix $\mathbf{W}(k)$ must satisfy $\mathbf{W}(k)\mathbf{H}(k) = \mathbf{I}$. The resulting vector BLU estimate is given by

$$
\begin{align}
\hat{\mathbf{x}}_{BLUE}(k) &= \mathbf{W}(k)\tilde{\mathbb{X}}(k) \tag{7} \\
&= [\mathbf{H}^T(k)\mathbb{P}^{-1}(k)\mathbf{H}(k)]^{-1}\mathbf{H}^T(k)\mathbb{P}^{-1}(k)\tilde{\mathbb{X}}(k)
\end{align}
$$

where $\mathbb{P}(k)$ is a composite covariance matrix represented as

$$\mathbb{P}(k) = diag\{[\mathbf{P}_1(k|k)+\mathbf{R}_1^c+\mathbf{R}_1^q], \ldots, [\mathbf{P}_N(k|k)+\mathbf{R}_N^c+\mathbf{R}_N^q]\}, \tag{8}$$

i.e., block diagonal, for the case where the individual reports are uncorrelated[2]. We define $\mathbf{P}_n^{(i,i)}(k|k)$ as the variance of the ith element of the update vector at time k from sensor node n. We assume these noise processes are uncorrelated in time, as well as spatially across vector elements. Noting the lack of a time index, this is reflected by $\mathbf{R}_n^q = E[\mathbf{n}_n^q \mathbf{n}_n^{q\,H}] = \sigma_q^2 \mathbf{I}_d$, $\mathbf{R}_n^c = E[\mathbf{n}_n^c \mathbf{n}_n^{c\,H}] = \sigma_c^2 \mathbf{I}_d \ \forall \ i = 1, \ldots, d, \ n = 1, \ldots, N$, where $E[\cdot]$ denotes the expectation operation. We represent the scalar variance terms as $r_n^{q,i}(k) = [\mathbf{R}_n^q(k)]_{i,i}$ and $r_n^{c,i}(k) = [\mathbf{R}_n^c(k)]_{i,i} \ \forall \ i = 1, \ldots, d$, which are the variances of the elements of the noise vectors. Note that the channel and quantization noise variances are functions of the power transmission level and bits used for quantization.

We use the mean squared error associated with this BLU estimator (also the variance), denoted D, as the metric of uncertainty to be minimized. The trace of the BLU estimate error covariance is taken to obtain a scalar quantity as

$$\begin{aligned} D(k) &= tr\{\text{cov}(\hat{\mathbf{x}}_{BLUE}(k) - \mathbf{x}(k))\} \\ &= tr\left\{ \left(\mathbf{H}^T(k)\mathbb{P}^{-1}(k)\mathbf{H}(k)\right)^{-1} \right\} \\ &= tr\left\{ \left(\sum_{n=1}^N [\mathbf{P}_n(k|k) + \mathbf{R}_n^c + \mathbf{R}_n^q]^{-1}\right)^{-1} \right\} \\ &= \sum_{i=1}^d \left(\sum_{n=1}^N \frac{1}{\mathbf{P}_n^{(i,i)}(k|k) + r_n^{c,i}(k) + r_n^{q,i}(k)} \right)^{-1} \\ &= \sum_{i=1}^d D_i(k), \end{aligned} \tag{9}$$

where $D_i(k)$ is the BLUE error variance for the ith element of the state estimate. The above definition of $D(k)$ follows since i) the second equality produces, by definition, the matrix form of the BLUE error covariance, ii) the third equality holds when the reports from the nodes are uncorrelated, and iii) the fourth equality is true when the noise terms are spatially uncorrelated. Since we have made these assumptions[2], (9) is our total uncertainty metric. Under these assumptions, the calculation of the optimal parameters for quantization and transmission allow the elements of the state vector to be estimated individually as

$$\begin{aligned} \hat{x}_{BLUE}^i &= \left(\sum_{n=1}^N \frac{1}{E[(\tilde{x}_n^i(k|k) - x^i(k))^2]} \right)^{-1} \\ &\times \sum_{n=1}^N \frac{\tilde{x}_n^i(k|k)}{E[(\tilde{x}_n^i(k|k) - x^i(k))^2]}, \end{aligned} \tag{10}$$

$\forall \ i = 1, \ldots, d$. Let $x_i \in [-W, W]$ with $[-W, W]$ the dynamic range of the measurement source. Then for a scalar element of the state vector,

$$r_n^{q,i}(k) = \frac{W^2}{3(2^{b_n^i(k)} - 1)^2} \tag{11}$$

is the uniform quantization noise variance. Each $b_n^i(k) \in [1, BW]$ is the number of bits used to quantize the ith element of the state estimate from sensor n to the fusion node. BW

is the rate constraint for the entire system. The quantization scheme is homogeneous across sensors for a quantization level, with a set dynamic range for components of position, velocity, et cetera. If the nth sensor node communicates using BPSK modulation for a Rayleigh fading channel then a modulation scheme that produces a probability of error of $P_n^{i,\ell}(error|k)$ for the ℓth bit of the transmission, then the noise due to the imperfect channel is

$$n_n^{c,i}(k) = \begin{cases} \pm 2^\ell \Delta_n^{i,\ell} & P_n^{i,\ell}(error|k) \\ 0 & 1 - \sum_{\ell=0}^{b_n^i(k)-1} P_n^{i,\ell}(error|k) \end{cases} \tag{12}$$

where $\Delta_n^{i,\ell} = \frac{2W}{2^{b_n^i(k)}-1}$ is the quantizer step size. We now assume: a) the bits in the transmit sequence have independently distributed probability of error, i.e., $P_n^{i,\ell}(error|k) = P_n^i(error|k)$, c) there is as most one bit error in each transmit sequence, and b) the channel noise variance is unchanged during a transmission period. All of these are reasonable assumptions for a slow fading channel with adequate channel coding. Thus the noise variance contributed from the channel is

$$\begin{aligned} r_n^{c,i}(k) &= \sum_{\ell=0}^{b_n^i(k)-1} (\pm 2^\ell \Delta_n^{i,\ell})^2 P_n^{i,\ell}(error|k) \\ &= P_n^i(error|k) \times (\Delta_n^{i,\ell})^2 \sum_{\ell=0}^{b_n^i(k)-1} 4^\ell \\ &\approx \frac{4W^2}{3} P_n^i(error|k), \end{aligned}$$

which simplifies to

$$r_n^{c,i}(k) \approx \frac{4W^2}{3} \left(1 - \sqrt{\frac{0.5\Gamma_n^i}{1 + 0.5\Gamma_n^i}} \right). \tag{13}$$

Where $\Gamma_n^i = \frac{2p_n^i \overline{|h_n|^2}}{N_0}$ represents the average received signal-to-noise-ratio, $p_n^i \in [p_n^{min} \ p_n^{max}]$ is the transmit energy level for the ith element of the nth sensor. The minimum power level per bit is p_n^{min} which is necessary to achieve a minimum system SNR. The maximum power per bit in a transmission is p_n^{max}. The power level p_n^i considers only the RF power required at the node, and none of the power consumed by other circuits in the device, which are considered negligible for simplicity in our analysis. The average power of the Rayleigh fading channel coefficient is $\overline{|h_n|^2}$ and $N_0/2$ is the channel noise power spectral density.

The previous discussion has introduced an objective for distributed estimation, which is optimal in the sense of minimizing an uncertainty metric which is a function of the variance of the estimate. The next section nestles this objective into the formal optimization statement which includes the constraints on the control variables.

III. OPTIMIZATION PROBLEM FORMULATION

It is desirable that optimization of quantization and transmit energy levels produce a balanced trade-off between estimate uncertainty and network lifetime. This trade-off is the subject of the following discussion. The optimization problem initially considers only the minimization of estimate variance under the given constraints, while lifetime is considered when we discuss energy-aware optimization. Assuming a subset of sensors has

---

[2]With the uncorrelated reports assumption, iii), we recognize that not all process noise models possess the quality of being spatially uncorrelated. When not true, then the total uncertainty in (9) becomes an approximation.

already been selected, we want to find the optimal of number of bits and transmit power levels that produce the best linear unbiased estimate of the process state, given the maximum resources allowed to be utilized. We perform the optimization of the following method for all time steps in the scenario or until the network has no active nodes (non-zero remaining energy). The formal expression of the minimization problem for a single dimension state vector is

$$
\begin{aligned}
minimize \quad & D(k) \\
subject\ to \quad & \\
& \sum_{n=1}^{N} b_n(k) \leq BW \\
& \Lambda_n(k) p_n(k) b_n(k) \leq p_n^{rem}(k) \\
& -b_n(k) + 1 \leq 0 \\
& p_n(k) - p_n^{max} \leq 0, \quad p_n^{min} - p_n(k) \leq 0 \\
& \forall\ n = 1, \dots, N.
\end{aligned}
\tag{14}
$$

Equivalently, the objective can be written as

$$
minimize \quad - D^{-1}(k),
$$

which results in a simpler objective function (by removing the inverse operating on the sum of inverted variance terms). The total power resources expended by each node is $p_n(k)b_n(k)$. Here $p_n(k)$ and $b_n(k)$ denote the power and bits used by the nth sensor node at the kth time instance; $\Lambda_n(k) \in [1, \frac{1}{\alpha}]$ is the weighting parameter, with fixed α a frugality parameter. $\Lambda_n(k)$ is best defined as a weighting that reflects the resource policy of each sensor based on its operating state, i.e., its remaining energy in the battery (p_n^{rem} in the above formulation). The weighting adjustment parameter, α, determines how $\Lambda_n(k)$ is updated. Low battery power would result in a large value for $\Lambda_n(k)$ and vice versa, the role of $\Lambda_n(k)$ is discussed next in Section III-C. The requirement that every node transmit at least one bit (the $-b_n(k) + 1 \leq 0$ constraint) reflects the status that selection and scheduling has already happened. The maximum and minimum constraints defined for the $p_n^i(k)$ and $b_n^i(k)$ henceforth shall be referred to as "box-constraints". The above problem must be expanded to account for the information transmitted for each of the elements of a multi-dimensional state vector. Thus, the altered form for vector quantities is written as

$$
\begin{aligned}
minimize \quad & -\sum_{i=1}^{d} D_i^{-1}(k) \\
subject\ to \quad & \\
C1: \quad & \sum_{i=1}^{d} \sum_{n=1}^{N} b_n^i(k) \leq BW \\
C2: \quad & \sum_{i=1}^{d} \Lambda_n^i(k) p_n^i(k) b_n^i(k) \leq p_n^{rem}(k) \\
C3: \quad & -b_n^i(k) + 1 \leq 0 \\
C4: \quad & p_n^i(k) - p_n^{max} \leq 0 \\
C5: \quad & p_n^{min} - p_n^i(k) \leq 0, \\
& \forall\ n = 1, \dots, N, \quad i = 1, \dots, d.
\end{aligned}
\tag{15}
$$

This formulation is by nature non-convex in the variables p_n^i and b_n^i, and in reality is a mixed-integer non-linear program (MINLP) with respect to the discrete values of the bits. We convert this to a difference of convex (DC) functions problem and solve the relaxed epigraph version of the problem by introducing new "uncertainty" variables $u_n^i(k)$, $n = 1, \dots, N$, $i = 1, \dots, d$, where

$$
u_n^i(k) = \frac{1}{\mathbf{P}_n^{(i,i)}(k|k) + r_n^{c,i}(k) + r_n^{q,i}(k)},
$$

each of which is a scalar quantity. Recalling the dependency of the channel and quantization noise on the power level and number of bits variables, the new epigraph form of the optimization problem is

$$
\begin{aligned}
minimize \quad & -\sum_{i=1}^{d} \sum_{n=1}^{N} u_n^i(k) \\
subject\ to \quad & \\
& u_n^i(k) - \frac{1}{\mathbf{P}_n^{(i,i)}(k|k) + r_n^{c,i}(k) + r_n^{q,i}(k)} = 0 \\
& \forall\ n = 1, \dots, N, \quad i = 1, \dots, d,
\end{aligned}
\tag{16}
$$

in addition to constraints C1-C5. We make additional simplifications to the uncertainty constraint by rewriting it as

$$
\mathbf{P}_n^{(i,i)}(k|k) + r_n^{c,i}(k) + r_n^{q,i}(k) - \frac{1}{u_n^i(k)} \leq 0
$$

(noting that the function decreases in $u_n^i(k)$ while the equality constraint increases with respect to it, the substitution is therefore adequate since the inequality introduced is strictly active at the minimum). This form of the constraint however, contains a convex function of power and bits, and concave function of the introduced uncertainty variables (i.e., the u's). A first order Taylor approximation of the concave reciprocal uncertainty term can be used to transform the difference of convex (DC) functions constraint into an approximate convex constraint. Thus the final convex approximation formulation (abbreviated as CVX) is

$$
\begin{aligned}
minimize \quad & -\sum_{i=1}^{d} \sum_{n=1}^{N} u_n^i(k) \\
subject\ to \quad & \\
& \mathbf{P}_n^{(i,i)}(k|k) + r_n^{c,i}(k) + r_n^{q,i}(k) - \left(\frac{2\tilde{u}_n^i(k) - u_n^i(k)}{(\tilde{u}_n^i(k))^2} \right) \leq 0 \\
& \forall\ n = 1, \dots, N, \quad i = 1, \dots, d,
\end{aligned}
\tag{17}
$$

still subject to constraints C1-C5, where $\tilde{u}_n^i(k)$ is the iterated point about which the Taylor approximation is taken. The Sequential Convex Programming (SCP) iterations find a stable upper bound to the original non-convex MINLP. Using SCP to obtain an approximation to the DC program comes attached to an increased computational effort as we must execute $O(nm^2)$ operations to solve a SQP at each iteration (where m is the number of constraints and n is the number of variables and $n \leq m$). A total computational cost of $O(nm^2L)$ results, where L is the number of SCP iterations. Since the KKT analysis does not provide any additional useful information concerning the problem behavior [25], we use simulation results to quantify the estimation performance of our relaxed convex approximation. Although this Taylor approximation is convex, the formulation experiences convergence issues because of the curvature of reciprocal function in the additional constraint. Next we investigate the use of a linear function in the added constraint to form a convex approximation which bypasses the need for SCP.

A. Low Complexity Formulation

A simpler formulation that matches this problem and circumvents sequential convex programming is presented. If, for
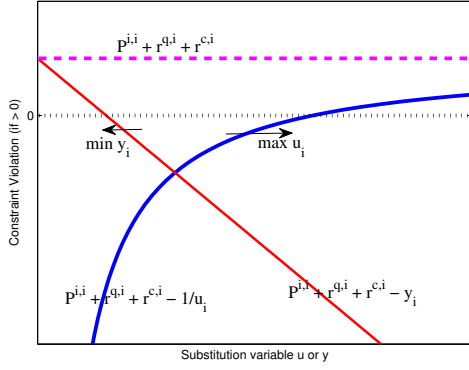
Fig. 2. An illustration of how the value produced by a negative linear function minimized produces a reciprocal value corresponding to the negative reciprocal function.

a moment, we relax some of the notation for the sake of clarity in explanation, consider the objective obtained above (17): Minimizing the negative of the sum of the uncertainty terms implies minimizing each individually. The statement

$$(min \ -u) \ \& \ (P \ + \ r^q + r^c \leq \frac{1}{u})$$
$$\Rightarrow max \ u \quad \rightarrow \quad min \ \{P + r^q + r^c\} \qquad (18)$$

maintains the truthfulness of what the substituted variable is meant to force. Now a parallel is drawn for a different substitutionary variable, y, playing an inverse role to u. The objective is simply to minimize positive uncertainty (or $-\frac{1}{u}$ in the current notation).

$$(min \ y) \ \& \ (P \ + \ r^q + r^c \leq y)$$
$$\Rightarrow min \ y \quad \rightarrow \quad min \ \{P + r^q + r^c\} \qquad (19)$$

which has the same resulting effect as the reciprocal substituted term. Thus u corresponds to inverse uncertainty in (18), and y simply to uncertainty in (19). This would now imply that the new uncertainty variable should be

$$y_n^i(k) = \mathbf{P}_n^{(i,i)}(k|k) + r_n^{c,i}(k) + r_n^{q,i}(k). \qquad (20)$$

To draw out a little more intuition, we plot the constraints involving the substitutionary variables for the two formulations. Maximizing the u's will reduce the final objective of the original non-convex problem written as $(\sum_n u_n^i)^{-1}$ is minimized for maximum u_n^i. Similarly, the y's represent the noise variance directly (and not the reciprocal) which will relates the original objective when written as $(\sum_n \frac{1}{y_n^i})^{-1}$, which is minimized when y_n^i's are minimum. Figure 2 illustrates that the graph of the two functions pursue reciprocal values of each other and that u and y are both valid substitutionary variables for the convexification of the problem. After inserting all of the notable changes above, the final problem formulation is

$$minimize \quad \sum_{i=1}^d \sum_{n=1}^N y_n^i(k)$$
$$subject \ to$$
$$\mathbf{P}_n^{(i,i)}(k|k) + r_n^{c,i}(k) + r_n^{q,i}(k) - y_i^n(k) \leq 0 \qquad (21)$$
$$\forall \ n = 1, \ldots, N, \quad i = 1, \ldots, d,$$

again still subject to constraints C1-C5. We call this new formulation in (21) the Linear Constraint Convex Approximation (LCVX) of the MINLP.

There are some convergence issues associated with the reciprocal constraint convex form (CVX). These issues can be mediated by tightening the tolerances which dictate the stopping criterion for the optimization routine. In addition to this, because the slope of the first-order Taylor approximation is dependent on the noise variance offset in the constraint, the difficulties associated with convergence vary from instance to instance so that in many cases the reciprocal constraint convex approximation achieves equal results than linear constraint convex form. However, the swift convergence of linear constraint form provides the equivalent solution with more consistency. In some sense, these comparisons are a moot point since both approximations equally form convex upper bounds to the original problem. For the sake of comparison, the value of the objective functions of the previous convex approximations are compared in Section IV-A with the same objective evaluated with integer bits (which are rounded or floored to satisfy the bandwidth constraint) from the continuous optimized values. Thus the relaxed convex approximations can be compared with the original mixed-integer non-convex problem (which is solved using a branch and bound based global solver). Next, the above approximations are further reduced so that the number of variables is decreased and the solutions to such simplifications provide a rapid upper bound to the original MINLP.

B. Worst-Case Formulation

In the present convex formulations, there are d bit and transmit energy parameters for each sensor node. It is desirable to formulate this problem with an objective function D which accounts for a single quantization and transmit energy parameter for each of the sensor nodes. The reduction of these parameters is done as follows. In the above (LCVX) formulation there exist a choice of the number of bits and power transmission level for each element of the state update for each node. If we consider a system which chooses the number of bits and power transmission level once for all elements of the state vector, the we reduce the number of needed variables by $2N(d-1)$. This transformation of the optimization problem is akin to requiring worst-case satisfaction of the chosen solution, i.e., only one of the uncertainty constraints will be active. An additional $N(d-1)$ variables can be removed from the problem by utilizing a single constraint with respect to the uncertainty term per sensor node, instead of d of them. This form is easily obtained by taking the maximum of the state covariances for all state elements, the constraints now include only one bit ($p_n(k)$), transmit energy ($b_n(k)$), and uncertainty ($y_n(k)$) variable per sensor node. This reduced variable "worst-case" problem is written as

$$minimize \quad d\sum_{n=1}^N y^n(k)$$
$$subject \ to$$
$$\max_i\{\mathbf{P}_n^{(i,i)}(k|k)\} + r_c^n(k) + r_q^n(k) - y^n(k) \leq 0$$
$$d\sum_{n=1}^N b^n(k) \leq BW \qquad (22)$$
$$d\Lambda^n(k)p^n(k)b^n(k) \leq p_n^{rem}(k)$$
$$\forall \ n = 1, \ldots, N.$$

TABLE I
CONVEX APPROXIMATION ALGORITHMIC DETAILS.

	WC	LCVX	CVX
Variables	$2N$	$2Nd$	$3Nd$
Constraints	$1 + 7N$	$1 + N(1 + 6d)$	
Runtime	$O(N^3)$	$O(N^3 d^3)$	$O(N^3 d^3 L)$

This representation of the convex program has a significant difference from the previous (LCVX) form: the uncertainty value being minimized is now the dimension of the state vector times the worst MSE term of that state estimate. Thus, the element of the state update vector with the largest variance determines the quantization and power levels for all of the elements of the state estimate for that node. Equivalently, for that node this determines the number of bits all data are encoded with and the transmission power level used. This formulation upper bounds the original MINLP as well.

Continuing the analysis of this simplification we have broken down the number of variables and constraints needed to evaluate the preceding problems. The formulation in equation (21) is the linear constraint convex formulation (LCVX) derived alongside the reciprocal constraint convex approximation (CVX) in (17). Both of these forms use the full number of variables and constraints (with the added uncertainty constraints) from the original problem. The form in (22) is the worst-case (WC) upper bound solution (worst-case in the sense of the sensor node state estimate covariances), and simplifies the problem by using a single bit and transmit energy variable and the maximum covariance value of any of the elements of an estimate at a particular node, thus reducing the number of variables and constraints. Table I shows the number of variables and their respective number of constraints given N sensors, state vector dimension d, and L SCP iterations. From these approximations we can glean the fact that the worst-case upper bound will be considerably more efficient when optimizing wireless sensor networks with a large number of nodes, which might be running Kalman estimators for two or three dimensions with higher order terms (velocity or acceleration of states). Number of variables and constraints for different convex approximate formulations of the optimal power-quantization problem. N is the number of sensors, d is the dimension of the state vector, and L is the number of SCP iterations

C. Energy-Aware Optimization

The above formulations attempt to reduce the variance of the fusion node estimate of the state by minimizing over uncertainty as the free variable. Under this paradigm the system will blindly use resources at each time iteration without consideration of the need for future transmissions, for either the sensor locally or its neighbors in the network. It was also noted that the behavior of this configuration is unchanging for differing channel or measurement noises between sensors, but rather it continues in the presence of unequal noise levels to follow the strategy of maximizing the transmit energy level and bandwidth per transmission, within the permissibility of the constraints. However, when the remaining energy for a subset of the nodes is significantly disadvantaged from the rest of the network, a reduction in the number of active sensors
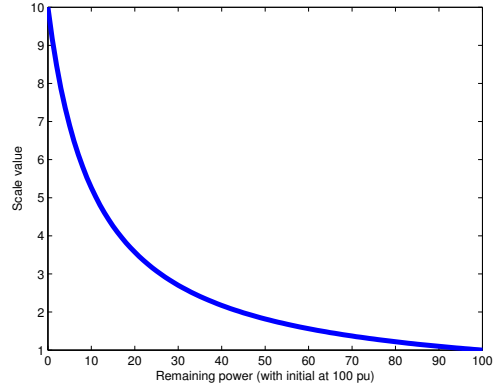

Fig. 3. The energy-aware scaling as a function of remaining energy.

may occur prematurely, i.e., node batteries are depleted. The results section will amply demonstrate this. As mentioned at the beginning of this section, we want to develop a strategy for preserving the energy of each node and thereby prolong the lifetime of the network. We use a heuristic scaling to alter the way in which a disadvantaged (in remaining energy) node is constrained in energy usage. A weighting function, $L(p_n^{rem}(k), p^{init}, k)$, is used to decide the level of transmit energy frugality. This preserves the battery life of nodes with a dwindling operating states. A possible schema for this scaling function, currently implemented, is

$$\Lambda_n(k) = L(p_n^{rem}(k), p^{init}, k) = \frac{1}{\alpha + (1 - \alpha) \cdot \frac{p_n^{rem}(k)}{p^{init}}} \quad (23)$$

$\forall\, n = 1, \ldots, N,\; \forall\, k$, where p_n^{rem} is the remaining power at node n while p^{init} is the average initial power allocated to each node. These weights are updated at each iteration and used to create the constraint

$$\Lambda_n(k) \cdot \sum_i p_n^i(k) b_n^i(k) < p_n^{rem}(k) \quad (24)$$

$\forall\, k, n$, where p_n and b_n are the power per bit and number of bits used in the current transmission interval. This is used in place of the previous remaining energy constraint (C2). A plot of the scaling value versus current energy level for $\alpha = 0.1$ and $p^{init} = 100$ is shown in Figure 3. This illustrates how a lower remaining power will result in a more frugal energy policy. A difficultly which becomes apparent is how to gracefully let a node deplete its energy. The constraints of the problem

$$p^{min} > p_n^i(k) \quad \text{and} \quad \sum_i p_n^i(k) b_n^i(k) \Lambda_n(k) < p_n^{rem}(k) \quad (25)$$

are conflicting for low remaining energy with a large value of $\Lambda_n(k) > 1$, which causes the problem to become infeasible. Thus, it was determined that if

$$p_n^{min} \cdot d \cdot \Lambda_n > p_n^{rem}$$

then Λ_n is set to 1 and node n is allowed to be depleted. This feature is plainly visible in Section IV-B. The next section details the simulation of the above optimization problem with and without this heuristic scaling procedure.

IV. SIMULATION RESULTS

We consider the distributed estimation of an object moving in two dimensions[3]. Each sensor maintains a position-velocity (PV) state estimate of the object. Power, time, and distance values are given in generic pu, tu, and du units, respectively, and are not necessarily equivalent to any standard units, but merely serve as a reference. Likewise, the number of bits should not be reflective of any particular wireless communication standard, but is only a part of the illustration of the optimization strategy. The sensor takes position measurements with uncorrelated measurement noise, which has variance $\sigma_n^2 = 1$ along both axes. The true trajectory starts at $[0, 0, v_x, v_y]^T$ (where $v_x = v_y = 1 du/tu$) and evolves with a nearly constant velocity (NCV) or Velocity Wiener process model [29]. The true process noise variance level is 0.8 as well as the assumed process noise of the local filters. The number of sensors in each of the following scenarios is $N = 4$. The communications parameters are as follows: The total allowable rate is set to $BW = 60$ bits per channel use and the dynamic range of observation is $W = 10$, with an offset region of $[0, 20]$ for the $x-$ and $y-$axis. The Rayleigh fading channel coefficient is $\overline{|h_n|^2} = 1$ and $N_0 = 0.5$ is the channel noise power spectral density coefficient. The maximum and minimum transmit energy are $p^{max} = 40pu$ and $p^{min} = 5pu$, respectively. We first illustrate that all methods perform comparably and their differences are measured against the single-instance global solution. Then, a pair of single run scenarios are shown for the purpose of illustrating the performance of the system for a typical run. Finally, multiple Monte Carlo runs are executed to characterize average performance.

A. Comparison of Convex Formulations to a Single-Instance Global Solution

The results for the linear constraint convex (LCVX) approximation should theoretically be very close if not equal to the reciprocal constraint convex approximation (CVX). This pair of formulations along with the worst-case (WC) approximations of the original problem are shown in the single instance optimization results given in Table II along with the exact global solution. The relaxed and integer value solutions are shown for comparison. This table shows the objective values computed from the original objective in (9) using the final decision variables determined by the convex approximations. Performance is compared to the exact (branch and bound) global solution[4]. In order to make an accurate comparison, the relaxed solution (rlxd) of the optimization problem is rounded (or floored, to maintain the total bandwidth constraint) and the original objective is then computed for comparison. This comparison also illustrates that the effect of integer relaxation is minimal in terms of achieved cost function values. The two

[3]As noted in Section II-B, the objective in (9) becomes an approximation of the MSE for spatially correlated noise. For a moving object in a field of sensors, the process noise among nodes is correlated. The application here was not selected to match the uncorrelated case observed in our objective, but rather for illustration purposes. Other work breaches the correlated process for these scenarios, e.g., in [30].

[4]LINGO® was used to find the globally optimal branch and bound solution. The optimal solution was returned with a dual certificate equal to it with sufficient numerical precision.

TABLE II
OBJECTIVE VALUES FOR CONVEX FORMULATIONS.

	WC		LCVX		CVX		Global
	rlxd	int	rlxd	int	rlxd	int	
Case A	6.533	7.522	5.499	6.036	5.605	5.658	3.588
Case B	10.674	11.211	9.318	9.213	9.372	9.324	5.665

cases considered in this juxtaposition of solutions represent uniform operating states, with readily available energy for each node (Case A); and a highly energy constrained scenario with uneven operating states where improper utilization of resources would cause sensor nodes to become inoperable (Case B). The heuristic parameter is set to $\alpha = 0.1$ for both cases. As expected, the linear constraint convex (LCVX) and reciprocal constraint convex (CVX) formulations both achieve nearly the same value. We note that the reciprocal constraint formulation converges very slowly and convergence can be improved by providing tighter tolerances and more iterations. However, we quickly reach the relative floating point accuracy of the numeric solution. This is a result of the difficulties explained in Section III-A. The worst-case approximation (WC) also does well but with performance inferior to the other convex approximations. This difference is due to non-uniformity in the covariance of the state estimate. For uniform state estimate variance, the LCVX and WC methods will be exactly the same. However, if there are differences in the covariance of the elements of the state estimate the WC will still uniformly split the bandwidth and power chosen for a particular sensor between the elements of the state estimate to be transmitted. Thus, the minimum BLU-like objective is only achieved for the most noisy element of the state estimate. Whereas the LCVX method can increase bits and transmit energy for noisier elements of the state estimate and thus maintain an overall lower BLU-like objective. This could be thought of as a reverse-water-filling-like property, i.e., attempting to match the estimate variance with more precision in the channel and quantization variance. The WC treats all elements of the state estimate covariance equally when they are not equal, resulting in using more bandwidth and power necessary for some estimates and not enough for others. The advantage of WC gained in computational speed is implied by the runtimes shown previously in Table I. We next analyze the behavior some typical single run time-based results common for LCVX.

B. Single Object Tracking Run with Non-uniform Initial Energy Resources

For the following scenarios linear constraint convex (LCVX) and worst-case (WC) approximations are shown here, as the LCVX and CVX methods perform similarly. The initial power allocated for this scenario is $p_n^{rem} = 4000pu$ for $n = 1, 4$ and $p_n^{rem} = 2000pu$ for $n = 2, 3$. The methods are shown both with and without the energy-aware heuristic (i.e. $\alpha = 1$ and $\alpha = 0.1$, respectively). For $\alpha = 1$ this implies $\Lambda_n(k) = 1, \forall k, n$. We additionally compare our methods to some static policies utilizing the minimum and maximum transmit energy levels. The minimum benchmark utilizes half the bandwidth and evenly allocates it amongst all
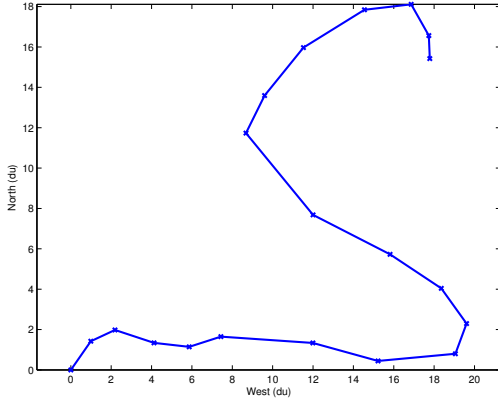
Fig. 4. Simulation scenario truth trajectory for single object tracking run.

active sensors. The maximum benchmark utilizes the entire bandwidth evenly allocated to all sensors. These are denoted "Min Bench", and "Max Bench" in the legends. The plot in Figure 4 shows a two dimensional plot of the object true trajectory. Figure 5 shows the individual power usage, number of bits used by a healthy and unhealthy sensor node in the system, in this case, sensors 1 and 2, respectively. Note that for the LCVX and CVX methods, the maximum bits and energy levels are selected at each time instant until the unhealthy nodes (2 and 3) deplete their energy resources and stop functioning.

After this, the remaining collaborating nodes have twice as much available bandwidth with which to transmit their states. This being the case, the remaining sensors use their energy reserves quickly since the optimization greedily uses all available bandwidth, also maximizing transmit energy levels, where possible. It is plain that the sensors deplete their energy resources quickly and stop tracking at about $9tu$. The benchmark methods provide us an intuitive upper and lower bound on resource usage and possible sensor node lifetimes. While the LCVX and WC follow the usage patterns of the maximum usage benchmark, the allocation is not optimal because it does not consider the influences of any noise sources, as we will see in the next section and the comparison of the error performance. When the energy-aware heuristic is applied to LCVX and WC (denoted in the legend with an appended "EA") for $\alpha = 0.1$, nodes 2 and 3 no longer deplete their energy resources, but rather these nodes function to almost $20tu$. It is easy to see the conditions near the depletion of sensor battery which result in setting $\Lambda_n(k) = 1$, as mentioned in Section III-C. We next compare the LCVX and WC methods alongside our benchmark methods for multiple Monte Carlo runs and apply some performance metrics.

C. Monte Carlo Simulation Analysis

The following results are obtained by executing multiple Monte Carlo (MC) runs. The starting energy in the single tracking run is the same here but results are averaged over 50 runs. We will now define several metrics by which to compare the methods of interest. In the case of a static policy, the *lifetime* is simply the number of iterations that can be run before the battery is depleted, i.e., $LT = \left\lfloor \frac{p_{rem}}{\sum_{i=1}^{d} p_n^i b_n^i} \right\rfloor$.

(a) Individual sensor node quantization levels.

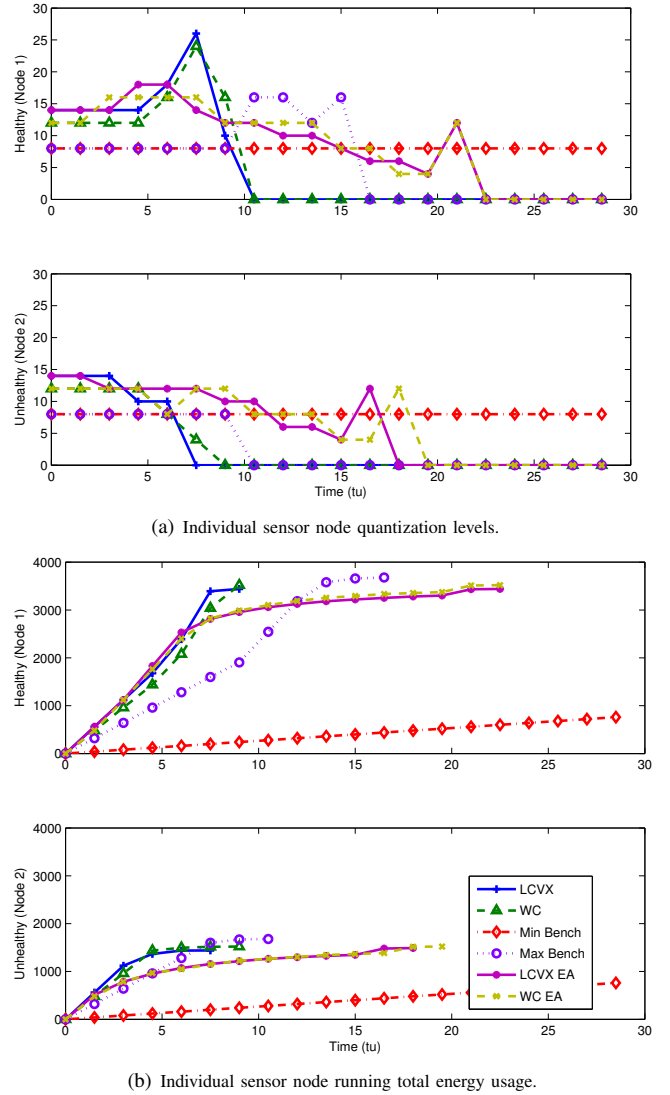(b) Individual sensor node running total energy usage.

Fig. 5. Single object tracking run: Transmit energy and quantization levels for a healthy and unhealthy node of the network, with and without the energy-aware heuristic ($\alpha = 1$ vs. $\alpha = 0.1$).

However, when power and quantization are determined dynamically based on the noise levels, then the lifetime is a probabilistic quantity, and no explicit function can be offered. Instead the lifetime value is determined during simulations and its probabilistic occurrence represented by Monte Carlo averages. We define the lifetime as

$$LT = arg\max_k \left\{ \frac{1}{N} \sum_{n=1}^{N} I(p_n^{rem}(k)) > \epsilon \right\}, \quad (26)$$

where

$$I(p) = \begin{cases} 0 & p < p_{min} \\ 1 & p \geq p_{min} \end{cases} \quad (27)$$

and ϵ is the node outage threshold. We also define the error measure by which we will approve our methods. We use a normed measure since we are consider a vector state. Define the Root Mean Square Normed Error (RMSNE) to be

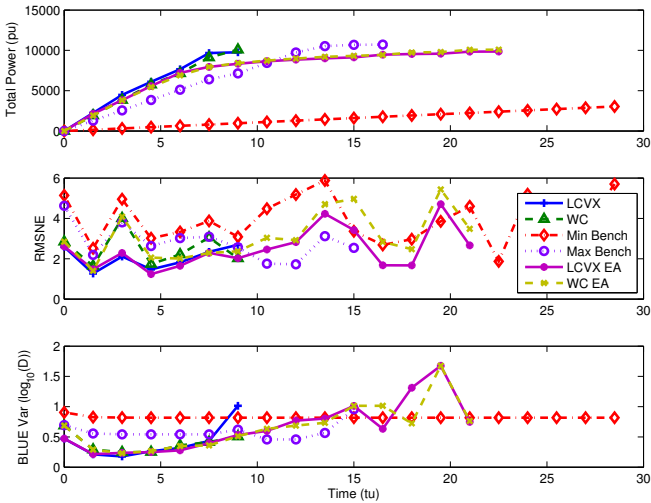$$RMSNE(k) = \sqrt{E[||\mathbf{x}(k) - \hat{\mathbf{x}}_{BLUE}(k)||^2]}, \quad (28)$$

Fig. 6. Single object tracking run: Comparison of total consumed energy, RMSNE, and analytic BLUE variance for the various methods.

where $||\cdot||$ is the ℓ_2-norm. The performance of the optimization methods versus the benchmark methods introduced in the previous section in terms of total power usage and RMSNE is shown in Figure 6. We clearly see that the LCVX and WC methods are comparable with the maximum benchmark, with LCVX having slightly better error performance than WC, as expected. The methods are nearly equivalent with and without the energy-aware heuristic up to about $9tu$ when the methods not employing the heuristic deplete their energy reserves. While the energy-aware enabled methods continue tracking at a somewhat reduced error performance, since they are conserving energy resources. The minimum benchmark maintains a higher error throughout the simulations, as it uses the fewest resources, with the exception of the very end of the energy-aware methods as they are forced to use fewer bits than even the minimum benchmark method. We have also include the analytic value of D from equation (9), which is dependent direct on the various noise variances determined by the choice of transmission power and quantization. While not directly proportional, this metric instructs us on what we should expect in the relative values of the RMSNE for the different methods.

Our evaluation of the energy-aware optimization can be performed by observing how average RMSNE and lifetime are affected by the heuristic scaling parameter, α. To do this, we find the RMSNE(k), and then average across time $k = 1, \ldots, K$. For these comparisons we also find the average lifetime for each α value. Figure IV-C clearly shows that the lifetime is greatly extended by using reasonably small values of α. It also demonstrates that the increase in error is moderate relative to the naïve approach. As expected, the LCVX always has much lower error than the WC method, while the WC method maintains slightly longer lifetimes for the same values of α. The probability of outage is the probability that the remaining percentage of nodes at any given time instance falls below a threshold. In our case, the running lifetime of each node is recorded and the WSN is considered inoperable when the fraction of nodes with remaining energy drops below the
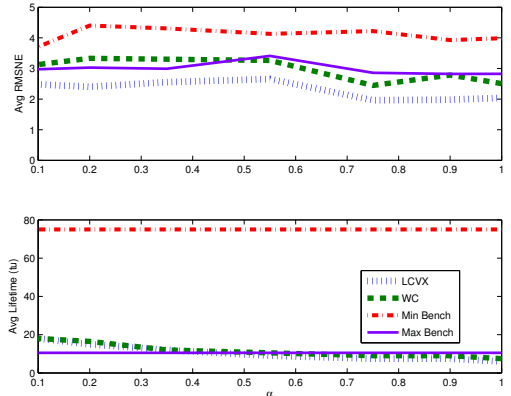
(a) WSN lifetime and error plotted for varying frugality parameter, α; the unequal initial energy scenario (1 & 2) with $N = 4$ is considered for the LCVX and WC formulations.
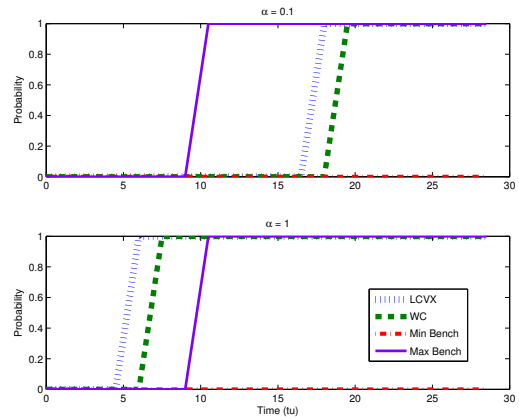

(b) System probability of outage for each of the formulation types, considering a threshold of 75% of the operating sensors still functioning. With (top) $\alpha = 0.1$ and (bottom) $\alpha = 1$.
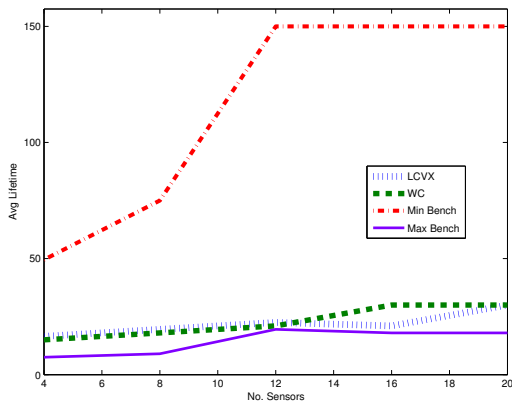
Fig. 7. The trade-off of lifetime versus average error performance and its affect on the system probability of outage.

threshold number of nodes ($\epsilon = 0.75$ in this case). For a single run the WSN is either operable or not (0 or 1) at each time instance. However, this is averaged over 50 MC runs, thus this provides a probability of outage, $Pr(k) \in [0, 1]$ at each time instance, k. We note in Figure 7(b) that the probability of outage for the network is extended for $\alpha = 0.1$ to 350% for LCVX and to 250% for WC. The LCVX and WC formulation have differing probability of outage performance for the reason given in Section IV-A, and it follows that slightly worse error performance would indicate a more energy-conservative network.

At this juncture it is important to point out that we do not explicitly investigate the scalability problem because of the assumption that sensor scheduling has been completed. The scalability of distributed estimation in wireless sensor networks in terms of either number of sensors or percentage of active sensors has been extensively explored in other works (e.g., [12]–[14]). These metrics of performance for a scheduling algorithm are indicators of the energy savings offered by the algorithm. As a thought experiment, if we assume the lifetime of a single node to be the number of (not necessarily consecutive) operational time instances, then it is intuitive if we schedule one sensor at each time step then we
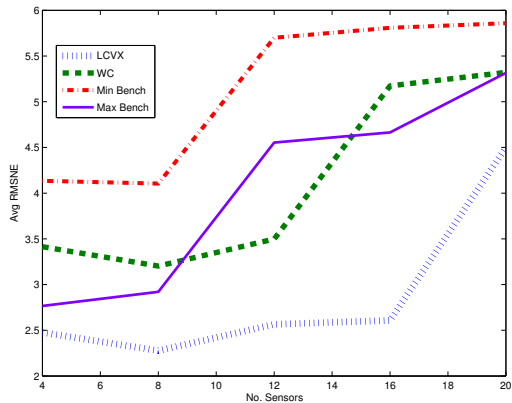
(a) RMSNE vs. number of sensors.

(b) Lifetime vs. number of sensors.

Fig. 8. The scalability of the scheduled subset size.

(a) LCVX Sensitivity to Covariance data perturbation.

(b) WC Sensitivity to Covariance data perturbation.

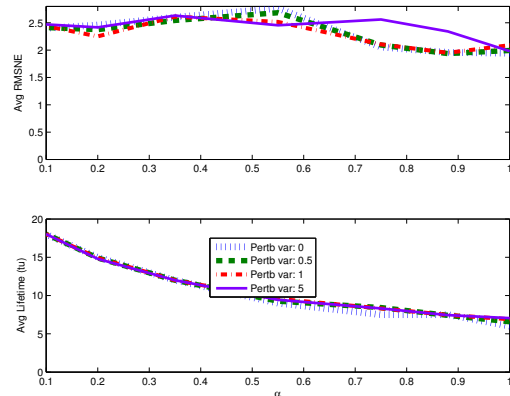Fig. 9. Covariance sensitivity tests for the LCVX and WC approximations for several values of α.

would have the network lifetime proportional to the product of the number of sensors and the lifetime of a single node. For our illustrations, if we assume a schedule that cycles between sets of sensor nodes of the same size and LT is the lifetime of a subset of $N < M$ sensors in a M node network, the relationship of overall WSN lifetime to that of the subset will be
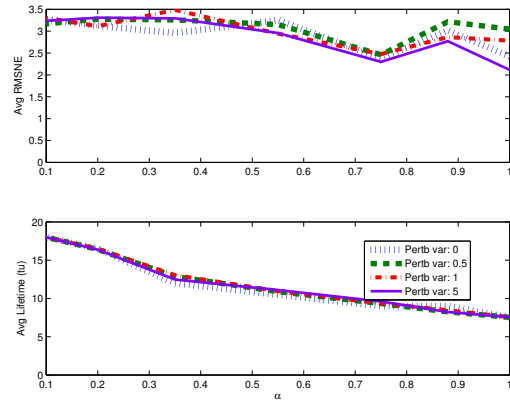
$$LT_{WSN} \propto \frac{M}{N} LT. \tag{29}$$

This is actually a lower bound since smarter scheduling could further increase the WSN lifetime by preventing disadvantaged nodes from being depleted (e.g. [31]). The vital issue with respect to the scalability of the network in this case is chiefly that of how to divide up the limited bandwidth. Indeed, there is a maximum number of scheduled nodes such that each node can transmit all components of the local updated state vector. As the network increases in size, the optimization is constrained is a smaller feasible space, as each element of each state vector must have at least one bit (since the nodes have been scheduled). The maximum network size based on single bit (minimum) quantization is

$$N_{max} = \frac{BW}{d}, \tag{30}$$

where d is the dimension of the state vector. In addition to this upper bound on scheduled node subset size, single bit

quantization can introduce significant errors, especially for a large dynamic range. Thus, we expect with increasing network size that the errors increase until no more sensors can be utilize in the scheduled set. We have included an illustration of this in Figure 8, where we have shown results for the LCVX and WC methods (since LCVX and CVX have equivalent results for the appropriate optimizing parameters). The setup is the same as the basic scenario described in Section IV-B (but with the adjustment of $BW = 100$). Figure 8 demonstrates the increasing average RMSNE for increasing number of sensors, particularly as the number of sensors approaches the limit in (30), which is 25 for Figure 8.

D. Sensitivity Analysis for Unknown Covariance

The nature of the problem set forth in this work demands knowledge of the covariances of the estimates at each node. This extra information is not considered in the transmission of data from the sensor nodes to the fusion center. We therefore wish to determine the sensitivity of the quantization and transmit energy level decision variables to random perturbations in the covariances, which are used to determine them. These perturbations are generated from Chi-squared distribution with one DOF, i.e., $\chi^2(1, \rho) = \sum_{i=1}^{k} Z_i^2$ for $k = 1$ with independent $Z_i \sim \mathcal{N}(0, \rho)$. In Figure 9 the optimization is carried out with randomly perturbed covariance

information and the results are averaged over 50 Monte Carlo runs. The variance parameter (ρ) of the perturbation is swept from zero (no perturbation) to five. Figure 9 clearly illustrates the small average error which is introduced for the linear constraint convex (LCVX) and worst-case (WC) approximations at various values of α. The same also demonstrates that the lifetime remains largely unaffected by perturbations to the covariance information. For all the methods the original objective (9) decreases for non-zero covariance perturbation. This is because more bits and higher transmit energy levels are selected to reduce noise and compensate for the additional noise introduced by the variance perturbation. However, the final estimation accuracy depends on the actual state covariance. Thus, more accurate estimates are obtained for perturbed covariance values, while lifetimes for such perturbations are decreased.

V. CONCLUSIONS

The distributed estimation scheme presented in this work utilizes a resource constrained uncertainty objective which is a non-convex MINLP. Formulations from the relaxed, convex approximated, scalar case were applied in estimating a state vector from distributed nodes. It was found that the linear constraint convex approximation and its worst-case upper bound are well suited for problem, both in performance and relative runtime. The energy-aware heuristic introduced allowed for the extension of network lifetime while delivering adequate estimation results, clearly expressing the trade-off between lifetime and estimation accuracy. Monte Carlo runs demonstrated that lifetime increases for decreasing heuristic parameter, which is also reflected in the probability of outage. While error also increases, it is only minute and does not significantly affect the estimation results.

REFERENCES

[1] D. A. Castanon and D. Teneketzis, "Distributed estimation algorithms for nonlinear systems," *IEEE Transactions on Automatic Control*, vol. 30, no. 5, pp. 418–425, May 1985.
[2] J. L. Speyer, "Computation and transmission requirements for a decentralized linear-quadratic-Gaussian control problem," *IEEE Transactions on Automatic Control*, vol. AC-24, no. 2, pp. 266–269, 1979.
[3] A. Willsky, M. Bello, D. A. Castanon, B. Levy, and G. Verghese, "Combining and updating of local estimates and regional maps along sets of one-dimensional tracks," *IEEE Transactions on Automatic Control*, vol. 27, no. 4, pp. 799–813, Aug. 1982.
[4] E. Ayanoğlu, "On optimal quantization of noisy sources," *Information Theory, IEEE Transactions on*, vol. 36, no. 6, pp. 1450–1452, Nov. 1990.
[5] J. A. Gubner, "Distributed estimation and quantization," *IEEE Transactions on Information Theory*, vol. 39, no. 4, pp. 1456–1459, 1993.
[6] W.-M. Lam and A. R. Reibman, "Design of quantizers for decentralized estimation systems," *Communications, IEEE Transactions on*, vol. 41, no. 11, pp. 1602–1605, Nov. 1993.
[7] H. Papadopoulos, G. Wornell, and a.V. Oppenheim, "Sequential signal encoding from noisy measurements using quantizers with dynamic bias control," *IEEE Transactions on Information Theory*, vol. 47, no. 3, pp. 978–1002, Mar. 2001.
[8] V. Megalooikonomou and Y. Yesha, "Quantizer design for distributed estimation with communication constraints and unknown observation statistics," *Communications, IEEE Transactions on*, vol. 48, no. 2, pp. 181–184, Feb. 2000.
[9] S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Magazine*, vol. 19, no. 2, pp. 51–60, Mar. 2002.
[10] A. Ribeiro, Z.-q. Luo, and G. B. Giannakis, "Distributed Compression-Estimation using wireless sensor networks," *IEEE Signal Processing Magazine*, no. July, pp. 27–41, 2006.
[11] J. Fang and H. Li, "Adaptive Distributed Estimation of Signal Power from One-Bit Quantized Data," *IEEE Transactions On Aerospace And Electronic Systems*, vol. 46, no. 4, 2010.
[12] Z.-Q. Luo and S. Member, "An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network," *IEEE Journal on Selected Areas in Communications*, vol. 23, no. 4, pp. 735–744, Apr. 2005.
[13] J.-J. Xiao, S. Cui, Z.-Q. Luo, and A. J. Goldsmith, "Joint estimation in sensor networks under energy constraints," *2004 First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004.*, vol. 00, no. c, pp. 264–271, 2004.
[14] V. Aravinthan, S. K. Jayaweera, and K. Al Tarazi, "Distributed estimation in a power constrained sensor network," in *IEEE Vehicular Technology Conference*. IEEE, 2006, pp. 1048–1052.
[15] J. Li and G. Al Regib, "Rate-Constrained Distributed Estimation in Wireless Sensor Networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 5, pp. 1634–1643, May 2007.
[16] ——, "Energy-constrained distributed estimation in Wireless Sensor Networks," in *Military Communications Conference, 2007 - MILCOM 2007*, 2007, pp. pp. 1 – 7.
[17] ——, "Maximizing Network Lifetime for Estimation in Multi-Hop Wireless Sensor Networks," *2008 Proceedings of 17th International Conference on Computer Communications and Networks*, vol. 15, pp. 533–536, Aug. 2008.
[18] A. Sripad and D. Snyder, "A necessary and sufficient condition for quantization errors to be uniform and white," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 25, no. 5, pp. 442–448, Oct. 1977.
[19] Y. Mostofi and R. Murray, "Effect of time-varying fading channels on the control performance of a mobile sensor node," *2004 First Annual IEEE Communications Society Conference on Sensor and Ad Hoc Communications and Networks, 2004. IEEE SECON 2004.*, vol. 00, no. c, pp. 317–324, 2004.
[20] S. Balasubramanian, S. Jayaweera, and K. Namuduri, "Energy-aware, collaborative tracking with ad-hoc wireless sensor networks," *IEEE Wireless Communications and Networking Conference, 2005*, pp. 1878–1883, 2005.
[21] M. Cardei and D.-Z. Du, "Improving Wireless Sensor Network Lifetime through Power Aware Organization," *Wireless Networks*, vol. 11, no. 3, pp. 333–340, May 2005.
[22] H. Zhu, I. Schizas, and G. Giannakis, "Power-efficient dimensionality reduction for distributed channel-aware kalman tracking using WSNs," *IEEE Transactions on Signal Processing*, vol. 57, no. 8, pp. 3193–3207, Aug. 2009.
[23] O. Ozdemir, R. Niu, and P. K. Varshney, "Channel Aware Target Localization With Quantized Data in Wireless Sensor Networks," *IEEE Transactions on Signal Processing*, vol. 57, no. 3, pp. 1190–1202, Mar. 2009.
[24] J. L. Williams, J. W. Fisher, and A. S. Willsky, "Approximate Dynamic Programming for Communication-Constrained Sensor Network Management," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4300–4311, 2007.
[25] R. Krishnan and B. Natarajan, "Joint Power and Quantization Optimization for Target Tracking in Wireless Sensor Networks," *IEEE GLOBECOM 2008*, vol. 2, no. 785, pp. 1–5, 2008.
[26] R. Horst and N. V. Thoai, "DC programming: overview," *J. Optim. Theory Appl.*, vol. vol. 103, no. No. 1, pp. pp. 1—43, 1999.
[27] Y. Bar-shalom and L. Campo, "The effect of common process noise on the two-sensor fused-track covariance," *IEEE Transactions On Aerospace and Electronic Systems*, no. 10, pp. 803–805, 1986.
[28] O. Drummond, "Tracklets and a hybrid fusion with process noise," in *Proc. 1997 SPIE Conf. Signal and Data Processing of Small Targets*, vol. vol. 3163, 1997, pp. pp. 512 – 524.
[29] J. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications, and Control*. New Jersey: Prentice Hall, 1995.
[30] A. D'Costa, V. Ramachandran, and A. Sayeed, "Distributed Classification of Gaussian Space-Time Sources in Wireless Sensor Networks," *IEEE Journal on Selected Areas in Communications*, vol. 22, no. 6, pp. 1026–1036, Aug. 2004.
[31] V. Gupta, T. H. Chung, B. Hassibi, and R. M. Murray, "On a stochastic sensor selection algorithm with applications in sensor scheduling and sensor coverage," *Automatica*, vol. 42, no. 2, pp. 251–260, Feb. 2006.