SPATIAL FILTERING IN SENSOR NETWORKS WITH COMPUTATION CODES

Anand D. Sarwate, Bobak Nazer, Michael Gastpar

Department of Electrical Engineering and Computer Sciences, UC Berkeley

ABSTRACT

A data collection problem in sensor networks is formulated in which the number of channel uses per source sample is greater than one. An example of this problem is given in which the objective of the data collector is to compute a filtered and downsampled version of the sensor field. In this regime, it is shown that uncoded transmission is not appropriate and that strategies based on separating source and channel coding perform poorly. By using a novel coding strategy based on computation codes, the power-distortion tradeoff becomes more favorable than that from separation.

Index Terms— sensor networks, spatial filtering, joint source-channel coding, distributed refinement

1. INTRODUCTION

Sensor network applications provide a rich set of problems in distributed signal processing. In many proposed frameworks, a central agent is interested in estimating a processed version of the sensor's observations. The quality of the estimate is limited by constraints on the communication channels between the sensors and the agent. Broadly speaking, two different approaches have been proposed to perform this kind of reconstruction. In the first approach, *separation*, the data gathering is decoupled from the computation. The sensors compress their observations individually and transmit them as independent messages over a shared communication channel and the collector uses the compressed observations to compute the function. The second approach, *uncoded transmission*, exploits two facts: the computation of interest is often linear and the communication channel is additive. The sensors perform a linear transformation of their observations to meet a transmit power constraint and transmit the analog values over the communication channel, and the collector performs some post-processing on the summed analog signals to arrive the final estimate.

We are interested in the case where there is a *source-channel bandwidth mismatch*, so that we have βk channel uses for k source symbols, with $\beta > 1$. Both separation and uncoded transmission have limitations in this setting – because separating compression and communication is generally suboptimal in networks, the distortion-power tradeoff for separation is often quite poor compared to joint source-channel strategies [1, 2]. Although uncoded transmission can sometimes achieve the optimal distortion-power tradeoff, it cannot exploit the bandwidth mismatch and hence becomes inefficient in this regime. In the paper we propose the use of a novel coding strategy based on *computation codes* as developed by Nazer and Gastpar [3].

As a motivating example, we will examine a toy model of a sensor network inspired by seismic monitoring. Sensors are placed at regular intervals closer than required by the spatial bandwidth of the seismic waves. Every day a plane flies over the sensor field to collect the data from the sensors. The collector wishes to create a summary of the sensor field in the form of a spatially low-pass filtered and downsampled version of the data at each time instant. One approach to this problem would be to gather all of the data and perform the processing and filtering off-line. This has the benefit of allowing further flexibility in processing at the data collector, but may be wasteful in terms of the energy expended by the sensors, since each sensor must quantize and transmit all of its observed data. By employing a computation code, we can directly compute the desired filter in an energy efficient fashion.

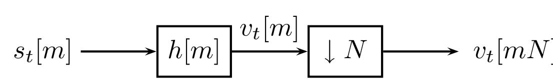

Fig. 1. Data collector's goal.

Our main result is that the tradeoffs provided by computation codes are beneficial in applications which require a coarse but low-distortion version of the sensor field. In the next section we describe our simple network model, followed by an analysis of the collect-then-filter in Section 3.1 and computation coding schemes in Section 3.2 for computing the filter-downsample operation. An example is given in Section 5 to illustrate the distortion-energy tradeoff.

2. PROBLEM FORMULATION

For simplicity, we will consider a network comprised of evenly spaced sensors along a line, as shown in Figure 2. We will assume the m-th sensor is at position $m\Delta$. The sensor at

B. Nazer was supported by an NSF Graduate Fellowship, A. Sarwate by NSF Grant CCR-0347298, and M. Gastpar by the ARO-MURI grant W911NF-06-1-0076. The first two authors contributed equally to this work.

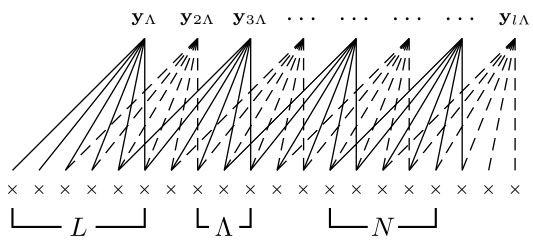

Fig. 2. A network on the line. Here $L = 6$, $\Lambda = 2$, and $N = 4$.

position m observes a source vector $\mathbf{s}[m]$ of length k. We can write the observations as a matrix

$$\begin{pmatrix} \cdots & s_1[1] & s_1[2] & \cdots & s_1[m] & \cdots \\ \cdots & s_2[1] & s_2[2] & \cdots & s_1[m] & \cdots \\ \cdots & \vdots & \vdots & \cdots & \vdots & \cdots \\ \cdots & s_k[1] & s_k[2] & \cdots & s_k[m] & \cdots \end{pmatrix} . \quad (1)$$

For simplicity of exposition, we will assume that the source vectors $\mathbf{s}[m]$ are iid (across time) and Gaussian with variance σ_s^2. However, for each time instant t, the realization of the sensor field $\{s_t[m] : m \in \mathbb{Z}\}$ is wide-sense stationary process with autocorrelation $R_{ss}[m]$ and power spectral density $S_{ss}(e^{j\omega})$.

The data collector is interested in computing a filtered and downsampled version of $s_t[m]$ for $t = 1, 2, \ldots k$, as shown in Figure 1. We will assume the filter $h[m]$ is FIR with transfer function

$$H(z) = a_0 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_M z^{-M}. \quad (2)$$

The output of the filter is $v_t[m] = (h * s_t)[m]$ and the collector would like to compute $v_t[mN]$, the output of the filter at integer multiples of the downsampling factor N.

As we face communication constraints and our sources are continuous-valued, we cannot reconstruct the filter output at arbitrary precision at the receiver. Our objective is to approximate this filter output at the minimum mean-squared error (MSE) or distortion, D, where:

$$D = \mathbb{E}\left[(v_t[m] - \hat{v}_t[m])^2\right] , \quad (3)$$

where $\hat{v}_t[m]$ is the estimate of $v_t[m]$ and expectation is over both time, t, and position, m.

We will assume that a collection agent can scan the network progressively. We can model this as a set of L-user additive white Gaussian noise multiple access channels (AWGN-MACs) containing overlapping subsets of the sensors, so that the MAC at position m has as users the sensors at positions $m - L + 1$ through m. We will assume that the data collector can query at positions Λm, where Λ is an integer greater than 1. Thus each sensor participates in L/Λ multiple access channels. For simplicity, we will assume $L \geq M$ unless otherwise specified.

For the m-th MAC, sensor i will encode its observed vector $\mathbf{s}[i]$ into a codeword $\mathbf{x}_{i,m}$ of length βk, where $\beta > 1$ is the bandwidth expansion factor. We will assume that β is an integer. The sensors then transmit their codewords over the channel to the data collector, which receives

$$\mathbf{y}_m = \sum_{i=m-L+1}^{m} \mathbf{x}_{i,m} + \mathbf{w}_i , \quad (4)$$

where $\mathbf{w}_j$ is iid Gaussian noise with variance σ_w^2. We can write the *average power used by sensor i for collection point m* by

$$P(i, m) = \frac{1}{\beta k} \|\mathbf{x}_{i,m}\|^2 . \quad (5)$$

We will assume that the collection agent samples the network *very infrequently*, and that the nodes collect a large amount of data between samplings, so that k is large.

3. SPATIALLY INDEPENDENT SOURCES

We first consider the case where the sensors are not correlated over space, so that $S_{ss}(e^{j\omega}) = \sigma_s^2$. Although the filter-then-downsample objective is less motivated in this case, the results are cleaner to state and form a point of comparison for the noisy and correlated cases in the next section. We describe the energy-distortion tradeoffs for separation and our algorithm based on computation codes. Because there is no canonical way of adapting the uncoded transmission strategy to the case $\beta > 1$, we will not compute tradeoffs for those schemes. For a related discussion, see [3]. Our results show the benefit gained by using computation codes to both exploit the channel structure as well as the source-channel bandwidth mismatch.

3.1. Collect then filter

We now analyze a coding strategy based on the separation principle. The data collector attempts to collect all of the source vectors from all of the sensors and then apply the desired filter. This scheme is limited by interference as each sensor has to share its channel with other users.

Proposition 1 (Separation with iid sources)**.** *For the sensor network collection problem with source vectors of length k and variance σ_s^2, filter $h[m]$, downsampling factor N, L-user MACs at every Λ sensors, channel noise σ_w^2, and channel blocklength βk, quantizing at the sensors and transmitting using a separation-based scheme results in a distortion-energy tradeoff given by:*

$$D_{\text{sep}} = \|h\|^2 \cdot \sigma_s^2 \left(\frac{\sigma_w^2}{\Lambda P_{\text{sep}} + \sigma_w^2} \right)^{\beta/\Lambda} \quad (6)$$

$$P_{\text{sep}} = \frac{\sigma_w^2}{\Lambda} \left(\left(\frac{\sigma_s^2 \|h\|^2}{D_{\text{sep}}} \right)^{\Lambda/\beta} - 1 \right) . \quad (7)$$

Proof. (Sketch) In the first step, each sensor uses an optimal Gaussian rate-distortion code to quantize their source observation with rate R and transmits these Rk bits using an optimal channel code to the data collector. With this encoding, we can write the quantized source vectors as

$$\tilde{\mathbf{s}}[m] = \mathbf{s}[m] + \mathbf{q}[m] \;, \tag{8}$$

where $\mathbf{q}[m]$ can be thought of as an iid white Gaussian source with variance D. The rate-distortion function per is given by

$$R(D_Q) = \frac{1}{2}\log\left(\frac{\sigma_s^2}{D_Q}\right) \;. \tag{9}$$

Since each sensor participates in L/Λ different MACs and time sharing is optimal for the MAC, we can view each sensor as effectively having $(L/\Lambda)\beta k$ channel uses on an L-user MAC. Since each user must transmit Rk bits, the power for each user must be set via:

$$\frac{R}{(L/\Lambda)\beta} = \frac{1}{2L}\log\left(1 + \frac{PL}{\sigma_w^2}\right) \;. \tag{10}$$

Therefore the power used per sensor can be written in terms of D_Q as

$$P = \frac{\sigma_w^2}{L}\left(\left(\frac{\sigma_s^2}{D_Q}\right)^{\Lambda/\beta} - 1\right) \;. \tag{11}$$

We can now assume that the vectors $\tilde{\mathbf{s}}[m]$ are available to the data collector for filtering and downsampling. The distortion in the filter output is just $D_{\mathrm{sep}} = \|h\|^2 D_Q$, and the energy expended is just $E_{\mathrm{sep}} = (L/\Lambda)n_c P$. Since downsampling will not reduce the average distortion per sample in the output of the filter, the average distortion for this scheme is given by D_{sep}. □

3.2. Using computation codes

In this section we describe a scheme that directly computes the output of downsampling the filter by using the additive property of the channel. Using the channel as part of a distributed estimation scheme was first considered in [1] and extended to a variety of settings in [4–6]. However, there the channel is used in an uncoded fashion and the distortion is driven down by a scaling number of sensors. Here, the number of sensors involved in a single filter output is fixed and the distortion is reduced by distributed refinement over several channel uses.

The key construction we will need are the computation codes developed in [3]. These lattice-based codes can allow the sensors to compute a linear function of their sources over the channel with a better distortion-energy tradeoff than that provided by the separation-based scheme. We will assume here that the downsampling factor N is an integer multiple of the MAC-sampling interval Λ.

In the computation code, the sensors participating in the lN-th MAC will compute the lN-th filter output $\mathbf{v}[lN]$ directly :

$$\mathbf{u}_{lN} = \sum_{i=1}^{M} a_i \mathbf{s}[lN - i] \;. \tag{12}$$

Computation codes use the MAC to compute an approximation $\hat{\mathbf{u}}$ to minimize the distortion $\frac{1}{k}\|\mathbf{u} - \hat{\mathbf{u}}\|^2$. The power-distortion performance of a computation code is given by the following lemma.

Lemma 1 (Computation codes for linear functions)**.** *For an M-user AWGN-MAC with Gaussian sources $\mathbf{s}_m$ at the users of variance σ_s^2, for $n_c = \beta k$ where $\beta \in \mathbb{Z}_+$ and k sufficiently large there exists a code of blocklength n_c for the channel that can compute a vector $\hat{\mathbf{u}}$ approximating $\mathbf{u} = \sum_{m=1}^{M} a_m \mathbf{s}_m$ such that*

$$D_{\mathrm{comp}} \leq \left(\sum_{m=1}^{M} a_m^2\right) \cdot M^{\beta-1}\sigma_s^2 \left(\frac{\sigma_w^2}{\sigma_w^2 + MP}\right)^{\beta} \;.$$

Due to space limitations, we cannot give a full proof of this result. We refer the interested reader to [3] for a detailed study of the computation of a sum from which the linear function case follows. The construction is a generalization of an elegant point-to-point scheme by Kochman and Zamir [7]. In the point-to-point setting, we want to communicate a single Gaussian source to a receiver over an AWGN channel. The receiver has Gaussian side information about the original source. Essentially, the sender quantizes its source to a course enough lattice such that the receiver can (with high probability) quantize the side information to the same lattice point. The *quantization error* is then transmitted uncoded over the channel. Using its observation of this quantization error, the receiver improves its estimate of the source. This process repeats with the new estimate serving as side information until we have run out of channel uses. For our computation code, each sender simultaneously runs this scheme using the same lattice. It follows that the receiver then sees the quantization error for the sum of the sources.

Thus, our computation coding scheme only requires a single vector quantizer on top of whatever hardware is needed to implement uncoded transmission. Note that the distortion given above may not be the best possible using computation codes. In particular, a construction that is adapted to the filter coefficients may achieve a lower distortion for the same power. However, with this result in hand, we can easily evaluate our performance in the proposed sensor network.

Theorem 1 (Computation codes with independent sources)**.** *For the sensor network collection problem with source vectors of length k and variance σ_s^2, filter $h[m]$, downsampling factor N, L-user MACs at every N sensors, channel noise σ_w^2, and channel blocklength βk, using a computation code to com-*

pute the filter output over the channel results in a distortion-energy tradeoff given by

$$D_{\text{comp}} \leq \|h\|^2 \cdot M^{\beta-1}\sigma_s^2 \left(\frac{\sigma_w^2}{\sigma_w^2 + NP_{\text{comp}}}\right)^{\beta} \tag{13}$$

$$P_{\text{comp}} \leq \frac{\sigma_w^2}{N}\left(\left(\frac{\|h\|^2 \cdot M^{\beta-1}\sigma_s^2}{D_{\text{comp}}}\right)^{1/\beta} - 1\right) . \tag{14}$$

The proof of this theorem follows directly from Lemma 1. Recall that we have assumed $M \leq L$, so the entire FIR filter output can be computed using the computation code. Each user participates in M/N MACs on average, so the average power expended per user is $P_{\text{comp}} = (M/N)P$ and by assumption $M/N < L/\Lambda$. The achievable distortion for computation codes in (13) is similar to that from separation in (6), but the exponent on the attenuation term is larger, and there is a penalty term of $M^{\beta-1}$. As we will see in the example, computation codes have a more favorable power-distortion tradeoff for low distortion.

4. NOISE AND SPATIAL CORRELATION

We now turn to two extensions – one for the case where the power spectrum $S_{ss}(e^{j\omega})$ is not constant, so that the sensor measurements are correlated over space, and the other when the sensor observations are noisy. As before, we are interested in computing the filtered and downsampled version of the sensor field. For the distributed source coding problem in the correlated setting, the exact rate region is not known except in some special cases, which means we cannot evaluate the true distortion-power tradeoff for the separation-based scheme. In the noisy case, the performance is limited by the observation noise, but the gap between the best centralized estimator and the decentralized estimator decays in a manner similar to (6). Due to space limitations we will defer evaluating bounds on the performance of separation for the correlated and noisy cases for the full version of this paper.

4.1. Observation Noise

In the noisy setting, we assume that the each terminal now observes an corrupted version of the source:

$$\mathbf{r}[m] = \mathbf{s}[m] + \mathbf{z}[m] , \tag{15}$$

where $\mathbf{z}[m]$ is an iid noise sequence with Gaussian distribution of mean 0 and variance σ_z^2.

In the noisy setting, the sensors first pre-process their information to compute the minimum mean-squared error estimate of $\mathbf{s}$ given $\mathbf{r}$:

$$\tilde{\mathbf{s}}[m] = \frac{\sigma_s^2}{\sigma_s^2 + \sigma_z^2}\mathbf{r}[m] . \tag{16}$$

They then use the computation code from the previous section to compute the filter output. This will efficiently compute the output of the filter with the MMSE estimate of each sensor value in place of the true value. An extra offset term appears due to the observation noise $\mathbf{z}[m]$.

Theorem 2. *For the sensor network collection problem with source vectors of length k and variance σ_s^2 observed through white Gaussian noise of variance σ_z^2 to form observations $\mathbf{r}[m]$ filter $h[m]$, downsampling factor N, L-user MACs at every N sensors, channel noise σ_w^2, and channel blocklength βk, using a computation code to compute the filter output over the channel results in a distortion-energy tradeoff given by*

$$D_{\text{comp}} \leq \|h\|^2 \left(\frac{\sigma_s^2\sigma_z^2}{\sigma_s^2 + \sigma_z^2}\right) \tag{17}$$

$$+ \|h\|^2 M^{\beta-1}\frac{\sigma_s^4}{\sigma_s^2 + \sigma_z^2}\left(\frac{\sigma_w^2}{\sigma_w^2 + NP_{\text{comp}}}\right)^{\beta} . \tag{18}$$

Proof. (Sketch) The constant first term in the distortion is identical to the distortion for the best centralized estimator having access to the values of $\mathbf{r}[m]$ directly. The computation code instead minimizes the distortion of the filter applied to the MMSE estimates of the source samples. Thus the distortion is the same as (13) with the source variance σ_s^2 replaced by $\sigma_s^4/(\sigma_s^2 + \sigma_z^2)$, the variance of the MMSE estimate. □

4.2. Correlated sources

We now take the case where the sensor observations are correlated across space (but not time), and there is no observation noise. We assume the sensor observations are wide-sense stationary with autocorrelation $R_{ss}[m]$. Using the computation code in the correlated case requires a modification of our main Lemma 1.

Theorem 3. *For the sensor network collection problem with source vectors of length k and autocorrelation $R_{ss}[m]$, target filter $h[m]$, downsampling factor N, L-user MACs at every N sensors, channel noise σ_w^2, and channel blocklength βk, using a computation code to compute the filter output over the channel results in a distortion-energy tradeoff given by*

$$D_{\text{comp}} \leq \|h\|^2\sigma_s^2 \left(\frac{\sigma_w^2}{\sigma_w^2 + NP_{\text{comp}}}\right)^{\beta-1}\left(\frac{\sigma_w^2}{\sigma_w^2 + P_{\text{eff}}}\right) , \tag{19}$$

where

$$P_{\text{eff}} = P_{\text{comp}}\left(\frac{1}{\max_m |a_m|^2}\right)\sum_{i,j=1}^{M} a_i a_j \frac{R_{ss}[i-j]}{\sigma_S^2} . \tag{20}$$

Proof. *(Sketch)* In the first phase of the computation code, the sensor observations are forwarded using an uncoded scheme that has effective power P_{eff}, which captures the beamforming gain of the sensor's correlations. In the remaining $\beta - 1$ phases, the lattice-based refinement scheme is agnostic to the correlations in the quantization errors, so the distortion reduction factor is identical to the case of independent source observations. □

The lattices used in this strategy do not take advantage of correlation and structure simultaneously, but we are not aware of structured codes which can outperform the performance shown above.

5. COMPARISON AND EXAMPLE

To illustrate the kinds of tradeoff curves we can achieve using computation codes, we consider a simple example with noiseless sensor observations. The filter h is a 64-tap FIR lowpass filter with transition band between frequencies $0.4 * 2\pi$ and $0.6 * 2\pi$, designed using the Parks-McClellan algorithm. For this example we take $\Lambda = 4$, $N = 16$, $L = 64$, $\sigma_s^2 = 0.5$, and $\sigma_w^2 = 1$. Figure 3 shows the distortion-power tradeoff curves for the collect-then-filter scheme and the scheme based on computation codes. We plot the tradeoff on a log-log scale for $\beta = 2, 4, 6, 8$.

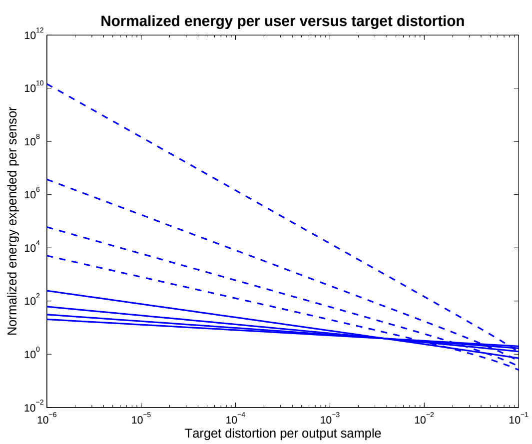

Fig. 3. Tradeoff between average power and distortion for $\beta = 2, 4, 6, 8$ using a 64-tap FIR lowpass filter. The dashed line is for the collect-then-filter protocol. The solid line is for the computation-code protocol.

As the bandwidth expansion factor increases, the power required for both schemes to reach the target distortion decreases. For applications which require low distortion, the savings offered by computation codes are quite significant, but for distortion-insensitive applications the gains are modest to non-existent, as evidenced by the crossing of the tradeoff curves. In the full version of this work we will demonstrate similar tradeoffs in the noisy and correlated settings.

6. DISCUSSION

Many models for linear data processing for wireless sensor networks exploit the additive nature of the wireless channel to improve the estimation performance. Uncoded transmission has been proposed as a power-efficient method of computing low-distortion estimates over multiple-access channels, but cannot exploit extra channel uses per source symbol. Computation codes exploit the additive structure of the channel and the extra channel uses to achieve a more favorable power-distortion tradeoff than separation. To explore this improvement we examined the situation of a network on a line in which the data collector wishes to acquire a filtered and downsampled version of the source observations. In the full version of this work we will derive explicit bounds for the performance of separation in the correlated and noisy settings to which we can compare the results presented here.

7. REFERENCES

[1] M. Gastpar and M. Vetterli, "Source-channel communication in sensor networks," in *2nd Int Workshop on Info Proc in Sensor Networks (IPSN '03)* (L. J. Guibas and F. Zhao, eds.), Lecture Notes in Computer Science, (New York, NY), pp. 167–177, Springer, April 2003.

[2] M. Gastpar, M. Vetterli, and P. L. Dragotti, "Sensing reality and communicating bits: A dangerous liason," *IEEE Signal Processing Magazine*, vol. 23, pp. 70–83, July 2006.

[3] B. Nazer and M. Gastpar, "Computation over multiple-access channels," *IEEE Transactions on Information Theory*, 2007. To appear October 2007.

[4] W. U. Bajwa, A. M. Sayeed, and R. Nowak, "Matched source-channel communication for field estimation in wireless sensor networks," *Proc. 4th Intl. Conf. on Info. Proc. in Sensor Networks (IPSN '05), Los Angeles, CA*, 2005.

[5] W. U. Bajwa, J. Haupt, A. M. Sayeed, and R. Nowak, "Joint source-channel communication for distributed estimation in sensor networks," *IEEE Transactions on Information Theory*, 2007. To appear October 2007.

[6] G. Mergen and L. Tong, "Type based estimation over multiaccess channels," *IEEE Transactions on Signal Processing*, vol. 54, pp. 613–626, February 2006.

[7] Y. Kochman and R. Zamir, "Analog matching of colored sources to colored channels," in *Proceedings of the IEEE International Symposium on Information Theory, Seattle, WA*, 2006.