Bandwidth-Constrained Estimation in Wireless Sensor Networks

Nikolaus Hammler

November 19, 2008

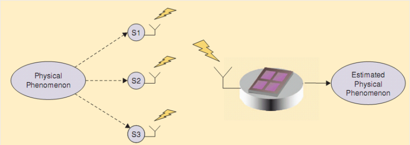
Distributed estimation

- Energy limitation
- Dynamic topology change
- Link quality (wireless communication)

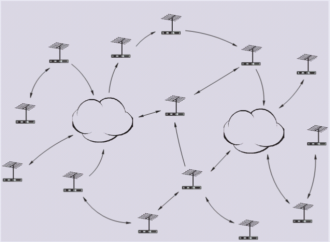
Conclusion

Low rate intersensor communication

WSN with FC



ad-hoc WSN



signal with noise

$$x[n] = f[n, \Theta] + w[n]$$

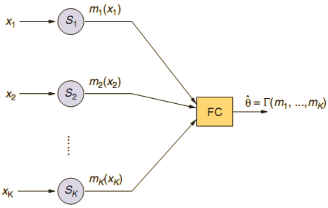
Distributed estimation framework

$$x_k = f_k(\Theta) + w_k$$

Quantization

Only quantized versions of x_k are sent: $m_k(x_k)$

How to quantize observations efficiently?



Estimators

- Estimate unknown parameters given a sampleset
- Function of sampleset: $\hat{\theta} = g(x_1, \ldots, x_N)$
- Classification
  - Classical/Fisher ET (e.g. **ML**, LS, **BLUE**)
  - Bayes ET (e.g. **MAP**, MMSE)

Estimator Performance

- Expected value: $E\{\hat{\theta}\}$ should be θ
- Bias: $b(\hat{\theta}) = E\{\hat{\theta}\} - \theta$
- Variance: $Var\{\hat{\theta}\}$ should be small
- MSE: $mse(\hat{\theta}) := E\{(\hat{\theta} - \theta)^2\} = Var\{\hat{\theta}\} + b(\hat{\theta})^2$
- Consistency: $mse(\hat{\theta}) \rightarrow 0$ for $N \rightarrow \infty$
- MVUE: Minimum Variance Unbiased Estimator
- CRLB: Cramer Raó Lower Bound
- Efficient: attains CRLB

Cramer Raó Lower Bound (CRLB)

CRLB I

If

$$E\left\{\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right\} = 0$$

then

$$Var\left\{\hat{\theta}\right\} \geq \frac{1}{-E\left\{\frac{\partial^2 \ln p}{\partial\theta^2}\right\}}$$

CRLB II

When the factorisation

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} = I(\theta)\,(g(\mathbf{x}) - \theta)$$

can be done, the the MVUE is

$$\hat{\theta}_{MVUE} = g(\mathbf{x}) \quad Var\{\hat{\theta}\} = \frac{1}{I(\theta)}$$

Fisher Information

$I(\theta)$ measure of information content in sampleset

Signal model

$$x_k = \theta + w_k$$

- Noise pdf: $p_k(w) = p(w) \, \forall k$
- $p(w)$ known, e.g. $\sim \mathcal{N}(0, \sigma^2)$
- Fusion center estimates θ from all x_k

The clairvoyant estimator

Assumption

No bandwidth constraints (i.e. sensors can send real values to fusion center)

$$\hat{\theta} = \frac{1}{K} \sum_{k=1}^{K} x_k$$

$$Var(\hat{\theta}) = \frac{\sigma^2}{K}$$

- Attains CRLB: MVUE
- Variance scales with $\mathcal{O}\left(\frac{1}{K}\right)$
- Problem: nothing new ;-)

Bandwidth constraint

Idea

Define a binary threshold for sensors. Use the output as indicator variables.

- Use 1 Bit per sensor: $L_k = 1$
- $m_k(x_k) = \begin{cases} 1 & x_k \geq \tau_c \\ 0 & x_k < \tau_c \end{cases}$
- m_k is bernoulli-distributed with parameter $q = P(x_k \geq \tau_c)$

1 bit estimator I

Probability of seeing a one

$$q = P(x_k \geq \tau_k) = F_w(\tau_c - \theta)$$

Inverting F_w

$$\theta = \tau_c - F_w^{-1}(q)$$

1 bit estimator II

Estimating q

Given all m_k observations, estimate q using the MLE:

$$\hat{q} = \frac{1}{K}\sum_{k=1}^{K} m_k$$

The 1 bit estimator

Using the invariance property, obtain

$$\hat{\theta}_{MLE} = \tau_c - F_w^{-1}\left(\frac{1}{K}\sum_{1}^{K} m_k\right)$$

1 bit estimator III

Alternatively, find by MLE by maximizing the log-likelihood:

$$L(\theta) = \sum_{k=1}^{K} m_k \ln q(\theta) + (1 - m_k) \ln (1 - q(\theta))$$

(observations i.i.d)

Performance I

The CRLB

$$Var(\hat{\theta}) \geq \frac{1}{K}\left(\frac{p^2(\tau_c - \theta)}{F(\tau_c - \theta)[1 - F(\tau_c - \theta)]}\right)^{-1} := \mathcal{B}(\theta)$$

Defining $\Delta_c := \frac{\tau_c - \theta}{\sigma}$

$$\mathcal{B}(\theta) = \frac{\sigma^2}{K}\frac{2\pi Q(\Delta_c)[1 - Q(\Delta_c)]}{e^{-\Delta_c}}$$

- Performance is dependent on distance $\tau_c - \theta$
- Best performance when $\tau_c = \theta$
- Variance increase only $\frac{\pi}{2}$
- Problem: We need to know the unknown parameter

Performance II

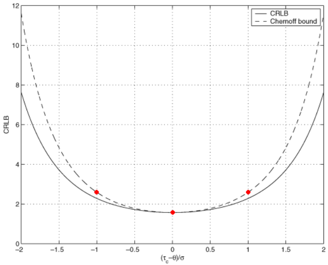
The maximum

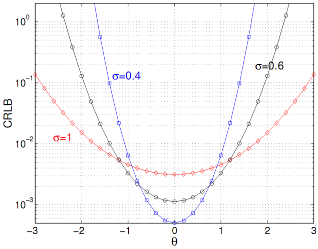
$$\mathcal{B}_{min} = \arg\min_{\Delta_c} \mathcal{B}(\theta) = \frac{2\pi\sigma^2}{4K}$$

Bound on CRLB

Using the Chernoff bound:

$$\mathcal{B}(\theta) \leq \frac{\pi}{2}\frac{\sigma^2}{K} e^{\frac{1}{2}\left(\frac{(\tau_c-\theta)}{\sigma}\right)^2}$$

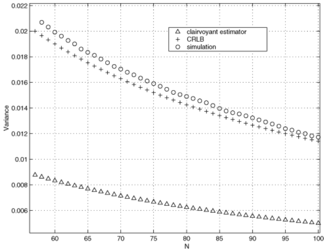




Simulation results

Simulation parameters

$$
\begin{aligned}
\sigma^2 &= 0.5 \\
\theta &= 1 \\
\tau_c &= \theta + \sigma \approx 1.707
\end{aligned}
$$



Iterative Estimator

$$\tau_c^{(i)} = \hat{\theta}^{(i-1)}$$

- $\mathcal{B}(\theta) \to \infty$ as $\tau_c - \theta \to \infty$
- Problem not unique to a particular estimator (CRLB)
- The higher $|\Theta_1 - \Theta_2|$ the more difficult to select τ_c

Q-SNR

$$\gamma := \frac{|\Theta_1 - \Theta_2|^2}{\sigma^2}$$

Conclusion

- Just one bit per sensor can be enough
- $\mathcal{O}(\frac{1}{K})$ remains valid
- Problematic assumption: $\tau_c = \theta$
- Iterative calculation possible
- Dynamic range gives the Q-SNR
- pdf must be fully known (including variance)
- Not for realistic scenarios

Signal model

General

$$x_k = \theta + w_k$$

$$p_w(w; \Psi) \;\; \Psi \in \mathbb{R}^{Lx1}$$

Simplification

Only variance unknown:

$$x_k = \theta + \sigma v_k$$

Idea

Define two thresholds τ_1, τ_2

$$B_k := \begin{cases} (\tau_1, \infty) := B_1 & \text{for } k = 1, \dots, K/2 \\ (\tau_2, \infty) := B_1 & \text{for } k = K/2 + 1, \dots, K \end{cases}$$

$$m_k(x_k) = \begin{cases} 1 & x_k \geq \tau_j \\ 0 & x_k < \tau_j \end{cases} \quad j \in \{1, 2\}$$

Bernoulli distributed q

$$q = \begin{cases} \Pr\{x_k \geq \tau_1\} = F_v\left(\frac{\tau_1 - \theta}{\sigma}\right) & k = 1, \dots, K/2 \\ \Pr\{x_k \geq \tau_2\} = F_v\left(\frac{\tau_2 - \theta}{\sigma}\right) & k = K/2 + 1, \dots, K \end{cases}$$

MLE of observations

$$\hat{q}_1 = \frac{2}{K} \sum_{k=1}^{K/2} m_k \qquad \hat{q}_2 = \frac{2}{K} \sum_{k=K/2+1}^{K} m_k$$

The estimator

Solving nonlinear 2x2-mapping: $(\theta, \sigma) \to (q_1, q_2)$:

$$\hat{\theta}_{MLE} = \frac{F_v^{-1}(\hat{q}_2)\tau_1 - F_v^{-1}(\hat{q}_1)\tau_2}{F_v^{-1}(\hat{q}_2) - F_v^{-1}(\hat{q}_1)}$$

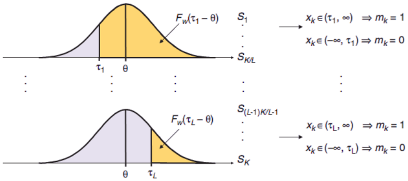
Performance

CRLB

$$Var(\hat{\theta}) \geq \frac{2\sigma^2}{K} \left(\frac{\Delta_1 \Delta_2}{\Delta_2 - \Delta_1} \right)^2 \left[\frac{q_1(1-q_1)}{p_v^2(\Delta_1)\Delta_1^2} + \frac{q_2(1-q_2)}{p_v^2(\Delta_2)\Delta_2^2} \right] := \mathcal{B}(\theta)$$

- $\mathcal{B}(\theta)$ is a linearcombination
- Performance still $\mathcal{O}\left(\frac{1}{K}\right)$
- Q-SNR should be small

L parameters



- $L + 1$ regions and groups
- e.g. Gaussian mixture pdf
- MLE can be found in closed form
- performance penalty small when γ is

Signal model

$$x_k = f_k(\Theta) + w_k$$

- $w_k \sim \mathcal{N}(0, \sigma_k^2)$ are known with pdf $p_k(w)$ and ccdf $F_k(w)$
- May change from sensor to sensor
- $f_k : \mathbb{R}^p \to \mathbb{R}$
- f_k *generally* nonlinear

1 bit estimator I

As before:

- Define 1 bit messages: $m_k(x_k) = \begin{cases} 1 & x_k \geq \tau_k \\ 0 & x_k < \tau_k \end{cases}$
- m_k is bernoulli distributed, as before
- $q_k = Prob(x_k \geq \tau_k) = F_k(\tau_k - f_k(\Theta))$

1 bit estimator II

The likelihoodfunction

Multiply their pdfs (samples i.i.d.):

$$L(\Theta) = \sum_{k=1}^{K} m_k \ln q_k + (1 - m_k) \ln (1 - q_k)$$

The MLE

And find

$$\hat{\Theta}_{MLE} = \arg\max_{\Theta} L(\Theta)$$

Finding the MLE I

- cannot be found in closed form
- numerical solution
- multimodal function
- values close to zero
- numerical ill-conditioning, e.g. saddle-points

Finding the MLE II

Theorem

$L(\Theta)$ is log-concave if

- *noise pdfs are log-concave*
- *f_n is linear*

Proof.

- w_k are log-concave
- regions $\mathcal{R}_k := (\tau_k, \infty)$ and $\bar{\mathcal{R}}_k$ are halflines
- $\mathcal{R}$ and $\bar{\mathcal{R}}$ convex sets
- q_k and $1 - q_k$ are integrals of w_k

□

Performance I

CRLB scalar case: quantized vs. clairvoyant

$$\mathcal{B}(\theta) = \frac{1}{K} \underbrace{\frac{F(\tau-\theta)[1-F(\tau-\theta)]}{p^2(\tau-\theta)}}_{\rho^2} \text{ vs. } Var(\bar{x}) = \frac{1}{K}\sigma^2$$

Idea

Define equivalent noise powers:

$$\rho_k = \frac{F_k(\tau_k - f_k(\Theta))[1 - F_k(\tau_k - f_k(\Theta))]}{p^2(\tau_k - f_k(\Theta))}$$

and

$$\rho := (\rho_1, \ldots, \rho_K)^T$$

Performance II

Equivalent noise increase

Estimating a vector parameter increases noises with factor $\frac{\rho_k^2}{\sigma_k^2}$

Optimally setting thresholds τ_k can lead to small penalty of $\frac{\pi}{2}$:

$$\rho_k^2 = \frac{\pi}{2}\sigma_k^2$$

Vector observations

$$\mathbf{x}_k = \mathbf{f}_k(\Theta) + \mathbf{w}_k$$

Whitening

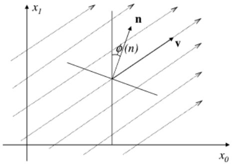
- $\mathbf{w}_k$ must be white: $\mathbf{C} = \sigma^2\mathbf{I}$
- $\mathbf{C}$ and $\mathbf{C}^{-1}$ positive definite
- Cholesky factorisation: $\mathbf{C}^{-1} = \mathbf{D}^T\mathbf{D}$
- $\mathbf{w}' = \mathbf{D}\mathbf{w}$
- Linear model: $\mathbf{x} = \mathbf{H}\Theta + \mathbf{w}$
- $\mathbf{D}\mathbf{x} = \mathbf{D}\mathbf{H}\Theta + \mathbf{D}\mathbf{w} = \mathbf{x}' = \mathbf{H}'\Theta + \mathbf{w}'$

Ex: Estimating a vector flow

Signal model

$$x_k =< \mathbf{v}, \mathbf{n} > + w_k = v_0 \sin \phi_k + v_1 \cos \phi_k + w_k$$

- Place thresholds at $\tau_k = v_0 \sin \phi_k + v_1 \cos \phi_k$
- $\rho_k^2 = (\pi/2)\sigma^2$
- Let $\phi_k \sim \mathcal{U}(-\pi, \pi)$
- $\bar{\mathbf{I}} = \frac{2}{\pi\sigma^2} \begin{pmatrix} N/2 & 0 \\ 0 & N/2 \end{pmatrix}$



Signal model

$$x_k = \theta + w_k \qquad \theta, w_k \in [-U, U]$$

Remarks

- w_k spatially uncorrelated, zero mean, unknown
- channels are orthogonal and distortionless

Remember

- BLUE $\bar{\theta}$ is optimal with mse σ^2/K
- network scales lin. with sensors

The estimator

Idea

1/2 sensors quantize to 1-MSB, 1/4 to 2-MSB etc.

- $mse \leq 4U^2/K$
- $mse \geq U^2/(4K)$
- FC required
- network size K required
- not isotropic

Isotropic estimator

Better idea

Each sensor flips a coin with $\Pr(a = j) = \alpha_j = \frac{1}{2^j}$, $j = 1, \ldots, \infty$

Coding

Local messages are coded by

$$m(x, a) = [b(2U + x; a); a]$$

Message length

$$\ell(m(x, a)) = 1 + \lceil \log_2 a \rceil$$

Set for each bit

$$\mathcal{N}_j = \{k | a_k = j,\ 1 \leq k \leq i\} \quad j = 1, 2, 3, \ldots$$

Estimate

sum up each bits per position

$$y_j = b(2U + x; j) + \sum_{k \in \mathcal{N}_j} b(2U + x_k; a_k)$$

$$\hat{\theta} = f_i(x, m(x_1, a_1), \ldots, m(x_i, a_i)) = -2U + 4U \sum_{j=1}^{\infty} \frac{2^{-j}}{N_j + 1} y_j$$

- $mse \leq 4U^2/(i+1)$
- ad-hoc estimating
- robust and isotropic
- independent of noise pdf

Signal model

$$x_k = \theta + w_k$$

- w_k uncorrelated, zero mean, different variances
- BLUE: $\hat{\theta}_{BLUE} = \left(\sum_{k=1}^{K} \frac{1}{\sigma_k^2}\right)^{-1} \sum_{k=1}^{K} \frac{x_k}{\sigma_k^2}$ with $mse(\hat{\theta}) = \left(\sum_{k=1}^{K} \frac{1}{\sigma_k^2}\right)^{-1}$
- Idea: Message related to local SNR
  - quantize more MSB
  - send more bits
  - different weighting factors

Quantizing

Message length

$$M_k = \left\lceil \log \frac{W}{\sigma_k} \right\rceil$$

Coding

$$m_k(x_k, a_k, M_k) = \sum_{i=1}^{M_k} b_i 2^{-i} + 2^{-M_k} b_{M_k + a_k}$$

with $\Pr(a = i) = 2^{-i}$

Estimator

Estimator

$$\hat{\theta} = \Gamma(m_1, \ldots, m_K) = \left(\sum_{k=1}^{K} 2^{2M_k} \right)^{-1} \sum_{k=1}^{K} 2^{2M_k} W(2m_k - 1)$$

Performance

$$mse(\theta) \leq \frac{25}{8} \left(\sum_{k=1}^{K} \frac{1}{\sigma_k^2} \right)^{-1}$$

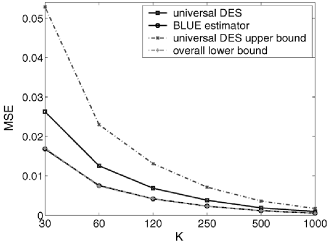
Simulation I

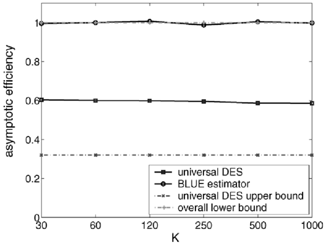
Asymptotic efficiency

$$\text{asymptotic efficiency} := \frac{1}{\text{MSE} \cdot \sum_{k=1}^{K} \frac{1}{\sigma_k^2}}$$

Parameters

$$\begin{aligned} \theta &= 1 \\ V &= 3 \\ U &= 6 \end{aligned}$$





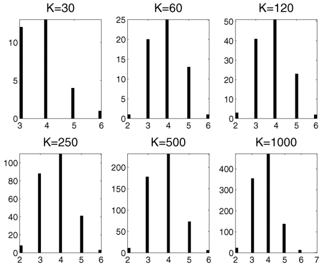
Simulation II

Numer of bits

$$M_k + 1 = \left\lceil \log \frac{W}{\sigma_k} \right\rceil + 1$$

Average message length

$$\bar{M}_k = \{3.8, 3.88, 3.83, 3.78, 3.77, 3.77\}$$



Power consumption

- Each sensor L_k bits
- Sensors to FC
- TDMA
- QAM with 2^{L_k} bits with p_b^k
- Distance d_k with pathloss exponent α: $a_k = d_k^{\alpha}$
- sensor sampling rate B_s

Transmit power

$$P_k = B_s \underbrace{2 N_f N_0 G_d a_k \left(\log \left(\frac{2}{p_b^k} \right) \right) (2^{L_k} - 1)}_{E_k}$$

The optimal L_k

$\mathbf{P}_q$ norm

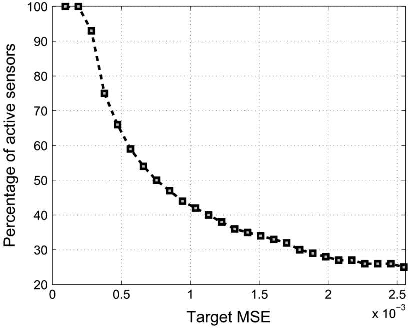
$$\|\mathbf{P}\|_q = \left(\sum_{k=1}^{K} P_k^q \right)^{1/q}$$

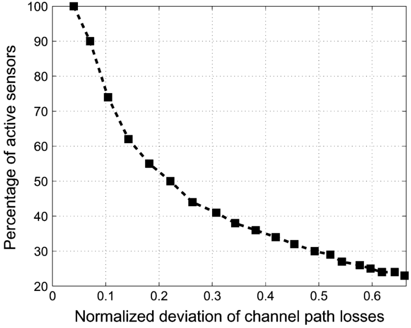
Message length

$$L_k^{opt}(\sigma_k, a_k) = \log \left(1 + \frac{W}{\sigma_k} \sqrt{\left(\frac{\eta_0}{a_k} - 1 \right)^+} \right), \quad \eta_0 = f(D_0, a_k, \sigma_k)$$

- threshold η_0: $L_k = P_k$ if $a_k \geq \eta_0$
- message length proportional to local SNR
- scales by channel path gain

Simulation


Percentage of active sensors
100
90
80
70
60
50
40
30
20
0
0.5
1
1.5
2
2.5
x 10^{-3}
Target MSE


Percentage of active sensors
100
90
80
70
60
50
40
30
20
0
0.1
0.2
0.3
0.4
0.5
0.6
Normalized deviation of channel path losses

MMSE

The MMSE

$$
\begin{aligned}
B_{MSE}\{\hat{\theta}\} &= \int\int (\hat{\theta}-\theta)^2 p(x,\theta)\, d\theta\, dx \\
&= \int \left[\int (\hat{\theta}-\theta)^2 p(\theta|x)\, d\theta\right] \underbrace{p(x)}_{\geq 0} dx \\
0 &\stackrel{!}{=} \frac{\partial}{\partial\hat{\theta}} \int (\hat{\theta}-\theta)^2 p(\theta|x)\, d\theta \\
\Rightarrow \quad \hat{\theta} &= \int \theta p(\theta|x)\, d\theta = \mathrm{E}\{\theta|x\}
\end{aligned}
$$

Exhaustive numerical integration necessary!

MAP

$$\hat{\theta}_{MAP} = \arg\max_{\theta} p(x|\theta)p(\theta)$$

- no closed form
- generalized maximum likelihood estimator
- numerical maximization
- MAP $\rightarrow$ MLE if K large
- concavity?

The estimator

$$
\begin{aligned}
\hat{\theta}_{MAP} &= \arg\max_{\theta} \left[\sum_{k=1}^{K} \log F_w[hm_k(\theta - \mu_0)] + \log p_\theta(\theta) \right] \\
&:= \arg\max_{\theta} L(\theta)
\end{aligned}
$$

Concavity of F_w and $p_\theta(\theta)$?

Performance?

$$
\mathcal{L} := \frac{\mathrm{MSE}(\hat{\theta})}{\mathrm{MSE}(\hat{\theta}_{CV})} \approx \frac{I(\hat{\theta}_{CV})}{I(\hat{\theta})} = \frac{\pi}{2}\sqrt{1+\gamma}
$$

Other topics

- Distributed kalman filtering/target tracking
    - Distributed estimating stochastic processes
    - SOI-KF
    - Target tracking
- Information theoretic aspects

References I

- Jin jun Xiao, Ro Ribeiro, and Zhi quan Luo.
  Distributed compression-estimation using wireless sensor networks.
  IEEE Signal Processing Mag, 23:27–41, 2006.
- S. M. Kay.
  Fundamentals of Statistical Signal Processing: Estimation Theory.
  Prentice Hall, 1997.
- Zhi-Quan Luo.
  An isotropic universal decentralized estimation scheme for a bandwidth constrained ad hoc sensor network.
  IEEE Journal On Selected Areas in Communications, 23:735–744, 2005.

References II

- Zhi quan Luo.
  Universal decentralized estimation in a bandwidth constrained sensor network.
  IEEE Trans. Inf. Theory, 51:2210–2219, 2005.
- Zhi quan Luo and Jin-Jun Xiao.
  Universal decentralized estimation in a bandwidth constrained sensor network.
  IEEE Int. Conf. Acoustic, Speech, Signal Processing, Philadelphia, PA, pages 829–832, 2005.
- Alejandro Ribeiro and Georgios B. Giannakis.
  Bandwidth-constrained distributed estimation for wireless sensor networks, part i: Gaussian case.
  IEEE Trans. Signal Process, 54:1131–1143, 2006.

References III

Alejandro Ribeiro and Georgios B. Giannakis.
Bandwidth-constrained distributed estimation for wireless sensor networks, part ii: Unknown pdf.
IEEE Trans. Signal Process, 54:2784–2796, 2006.

Faisal Ali Shah, Alejandro Ribeiro, and Georgios Giannakis.
Bandwidth-constrained map estimation for wireless sensor networks.
IEEE Trans. Signal Process, 54:413–422, 2006.

Jin-Jun Xiao and Zhi-Quan Luo.
Decentralized estimation in an inhomogeneous sensing enironment.
IEEE Trans. Inf. Theory, 51:3564–3575, 2005.

References IV

- Jin-Jun Xiao, Zhi-Quan Luo, Shuguang Cui, and Andrea Goldsmith.
  Power scheduling of universal decentralized estimation in sensor networks.
  IEEE Trans. Signal Process, 54:413–422, 2006.
- Jin-Jun Xiao, Zhi-Quan Luo, and Georgios Giannakis.
  Performance bounds for the rate-constrained universal decentralized estimators.
  IEEE Trans. Signal Process, 14:47–50, 2007.