On Data Acquisition and Field Reconstruction in Wireless Sensor Networks^{*}

Carla-Fabiana Chiasserini, Alessandro Nordio, and Emanuele Viterbo

Dipartimento di Elettronica, Politecnico di Torino C. Duca degli Abruzzi 24, I-10129 Torino (Italy) e-mail: <name>@polito.it

Abstract. Wireless sensor networks are often used for environmental monitoring applications. Sampling and reconstruction of a physical field is therefore one of the most important problems to solve. We focus on band-limited fields and investigate the relationship between the random topology of a sensor network and the quality of the reconstructed field. By reviewing irregular sampling theory, we derive some guidelines on how sensors should be deployed over a spatial area for efficient data acquisition and reconstruction. We analyze the problem using random matrix theory and show that even a very irregular spatial distribution of sensors may lead to a successful signal reconstruction, provided that the number of collected samples is large enough with respect to the field bandwidth.

Key words: sensor networks, field reconstruction, irregular sampling.

1 Introduction

One of the most popular applications of wireless sensor networks is environmental monitoring. In general, a physical phenomenon (hereinafter also called sensor or physical field) may vary over both space and time. In this work, we address the problem of sampling and reconstruction of a one-dimensional, spatial field at a particular time instant. We focus on a band-limited field (e.g., pressure and temperature), and assume that sensors are randomly deployed over the area of interest. Also, nodes can represent each sample with a sufficient number of bits, so that the quantization error is negligible.

Data are transfered from the sensors to a common data-collecting unit, the so-called sink node. In this work, however, we are concerned only with acquisition and reconstruction of the sensor field, and we do not address issues related to information transport. Thus, although studying the effect of errors and losses due to data transfer is of great interest, we assume that all data is correctly received at the sink. Furthermore, we assume that the sensors position are known. This implies that nodes are either located at pre-defined positions, or, if randomly deployed, their location can be acquired.

^{*} This work was supported through the PATTERN project

Our objective is to investigate the relation between the network topology and the quality of the reconstructed field. More specifically, we aim at identifying the topology characteristics and, hence, a sample distribution that allows the sink node to reconstruct the signal of interest with the desired precision. The main contributions of this paper are the following:

- (i) by reviewing irregular sampling theory, we derive some guidelines on the number of sensors to be deployed and on how they should be spatially spaced so as to successfully reconstruct the measured field;
- (ii) by analyzing the problem using random matrix theory, we show that even a very irregular spatial distribution of sensors may lead to a successful signal reconstruction, provided that the number of collected samples is large enough with respect to the field bandwidth;
- (*iii*) we identify the theoretical basis to estimate the required number of active sensors, given the field bandwidth.

2 Related work on data acquisition in sensor networks

To the best of our knowledge, few works have addressed the problem of sampling and reconstruction in sensor networks. Efficient techniques for spatial sampling in sensor networks are proposed in [5,6]. In particular [5] presents an algorithm to determine which sensor subsets should be selected to acquire data from an area of interest and which nodes should remain inactive to save energy. The algorithm chooses sensors in such a way that the node positions can be mapped into a blue noise binary pattern. In [6], an adaptive sampling is described, which allows the central data-collector to vary the number of active sensors, i.e., samples, according to the desired resolution level. The problem of data acquisition is also addressed in [4], where the authors consider a one-dimension field, uniformly sampled at the Nyquist frequency by low precision sensors. The authors show that the number of sensors (i.e., samples) can be traded-off with the precision of sensors. Finally, the work in [7] proposes to use synthetic data generation techniques to generate irregular data topology from some available experimental data. The objective there is to obtain a field model to evaluate sensor network algorithms.

3 Irregular sampling of band-limited signals

Let us consider the one-dimensional model where r sensors, randomly located in the interval [0, 1), measure the value of a band-limited signal p(t). Let $t_j \in [0, 1)$ for $j = 1 \dots, r$ be the locations of the sampling points ordered increasingly and $p(t_j)$ the corresponding samples.

A strictly band-limited signal over the interval [0, 1) can be written as the weighted sum of M' harmonics in terms of Fourier series:

$$p(t) = \sum_{k=-M'}^{M'} a_k \exp(2\pi i k t) \tag{1}$$

Note that for real valued signals the Fourier coefficients satisfy the relation $a_k^* = a_{-k}$ so that

$$p(t) = \sum_{k=-M'}^{M'} \rho_k \cos(2\pi kt + \phi_k)$$

where $a_k = \rho_k \exp(i\phi_k)$.

The reconstruction problem can be formulated as follows: given r pairs $[t_j, p(t_j)]$ for j = 1, ..., r find the band-limited signal in (1) uniquely specified by the sequence of its Fourier coefficients a_k .

Let the reconstructed signal be

$$\hat{p}(t) = \sum_{k=-M}^{M} \hat{a}_k \exp(2\pi i k t)$$
(2)

where the \hat{a}_k are the corresponding Fourier coefficients up to the *M*-th harmonic. In general, the reconstruction procedure will minimize $||p(t) - \hat{p}(t)||^2$ if M < M'and gives $p(t) = \hat{p}(t)$ if M = M'.

Consider the $(2M + 1) \times r$ matrix **F** whose (k, q)-th element is defined by

$$(\mathbf{F})_{k,q} = \frac{1}{\sqrt{r}} \exp(2\pi i k t_q) \quad \begin{array}{l} k = -M, \dots, M\\ q = 1, \dots, r \end{array}$$

the vector $\hat{\mathbf{a}} = [a_{-M}, \dots, a_0, \dots, a_M]^{\mathrm{T}}$ of size 2M + 1 and the vector $\mathbf{p} = [p(t_1), \dots, p(t_r)]^{\mathrm{T}}$. We have the following linear system

$$\mathbf{F}\mathbf{F}^{\dagger}\hat{\mathbf{a}} = \mathbf{F}\mathbf{p} \tag{3}$$

where $(\cdot)^{\dagger}$ is the conjugate transpose operator. In the following we will denote $\mathbf{T} = \mathbf{F}\mathbf{F}^{\dagger}$ and $\mathbf{b} = \mathbf{F}\mathbf{p}$.

When the samples are equally spaced in the interval [0, 1), i.e., $t_q = (q-1)/r$, we observe that the matrix \mathbf{F} is a unitary matrix $(\mathbf{FF}^{\dagger} = \mathbf{T} = \mathbf{I}_{2M+1})^{-1}$ and its rows are orthonormal row vectors of an inverse DFT matrix. In this case (3) gives the first M Fourier coefficients of sample sequence \mathbf{p} .

When the samples t_q are not equally spaced, the matrix **F** is no longer unitary and the matrix **T** becomes a $(2M + 1) \times (2M + 1)$ Hermitian Toeplitz matrix

$$\mathbf{T} = \mathbf{T}^{\dagger} = \begin{pmatrix} r_0 & r_1 \cdots & r_{2M} \\ r_{-1} & r_0 \cdots & r_{2M-1} \\ & \ddots & \\ r_{-2M} & \cdots & r_0 \end{pmatrix}$$

where

$$(\mathbf{T})_{k,m} = r_{k-m} = \frac{1}{r} \sum_{q=1}^{r} \exp(2\pi i(k-m)t_q) \qquad k,m = -M\dots,M$$
(4)

¹ The symbol \mathbf{I}_n represents the *n* by *n* identity matrix

It follows that the Toeplitz matrix \mathbf{T} is uniquely defined by the 4M + 1 variables

$$r_{\ell} = \frac{1}{r} \sum_{q=1}^{r} \exp(2\pi i\ell t_q) \qquad \ell = -2M, \dots 2M$$

The solution of (3), which involves the inversion of \mathbf{T} , requires some attention if the condition number of \mathbf{T} (or equivalently of \mathbf{F}) becomes large. We recall that the condition number of \mathbf{T} is defined as

$$\kappa = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}}$$

where λ_{max} and λ_{min} are the largest and the smallest eigenvalues of **T**, respectively. In practice, matrix inversion is usually performed by algorithms which are very sensitive to small eigenvalues, especially when smaller than the machine precision. For this reason in [1] a preconditioning technique is used to guarantee a bounded condition number when the maximum separation between consecutive sampling points is not too large. More precisely, by defining $w_q = (t_{q+1} - t_{q-1})/2$ for q = 1..., r, where $t_0 = t_r - 1$ and $t_{r+1} = 1 + t_1$, and by letting $\mathbf{W} = \text{diag}(w_1, \ldots, w_r)$, the preconditioned system becomes

$$\mathbf{T}_w \hat{\mathbf{a}} = \mathbf{b}_w$$

where $\mathbf{T}_w = \mathbf{FWF}^{\dagger}$ and $\mathbf{b}_w = \mathbf{FWp}$. By defining the maximum gap between consecutive sampling points as

$$\delta = \max(t_q - t_{q-1}),$$

when $\delta < 1/2M$, it is shown in [1] that

$$\kappa(\mathbf{T}_w) \le \left(\frac{1+2\delta M}{1-2\delta M}\right)^2$$

This result generalizes the Nyquist sampling theorem to the case of irregular sampling, but only gives a sufficient condition for perfect reconstruction when the condition number is compatible with the machine precision.

In Figure 1 and 2, we give an example of the reconstruction from irregular samples of a band-limited signal, using (3). In Figure 1, we chose M = 10 and r = 26 and the samples have been randomly selected over the interval [0, 0.8). The signal has been perfectly reconstructed even if large gaps are present $(\delta > 0.2)$. In Figure 2, r = 21 samples of the same signal of Fig. 1 have been taken randomly over the entire window. Due to the bad conditioning of the matrix **T** (i.e., very low eigenvalues) the algorithm failed to reconstruct the signal due to machine precision underflow.

4 The random matrix approach

4.1 Some statistics on r_{ℓ}

The above results are based on deterministic locations of the sampling points. In this section we discuss instead the case where the sampling points t_q are i.i.d.



Fig. 1. Example of a reconstructed signal from irregular sampling

random variables with uniform distribution $\mathcal{U}[0,1]$. The Toeplitz matrix **T** is now defined by the complex random variables

$$r_{\ell} = \frac{1}{r} \sum_{q=1}^{r} \exp(2\pi i \ell t_q) \qquad \ell = -2M, \dots 2M$$

where the complex random variable $\xi = \exp(2\pi i \ell t_q)$, for some k, q, with $t_q \sim \mathcal{U}[0,1]$ has $\mathbb{E}[\xi] = 0$ and $\mathbb{E}[|\xi|^2] = 1$, where $\mathbb{E}[\cdot]$ is the expectation operator.

Given a set of p random variables $r_{l_1} \dots, r_{l_p}$ from **T** we are now interested in computing the mixed moments

$$\mu_p(\ell_1, \dots, \ell_p) = \mathbb{E}[r_{\ell_1} \cdots r_{\ell_p}] \tag{5}$$

These statistics are useful for analyzing the eigenvalue distribution of the random matrix \mathbf{T} , as shown later in Section 4.2. For this purpose we define the vectors of integers $\mathbf{l} = [\ell_1, \ldots, \ell_p]$ where $\ell_i = -2M, \ldots, 2M$, and $\mathbf{q} = [q_1, \ldots, q_p]$ where $q_i = 1, \ldots, r$, and the real vector $\mathbf{u}_{\mathbf{q}} = [u_{q_1}, \ldots, u_{q_p}] \in [0, 1]^p$. From (5) we have

$$\mu_p(\boldsymbol{l}) = \frac{1}{r^p} \sum_{\mathbf{q}} \int_{[0,1]^p} \exp(2\pi \mathbf{i} \ \boldsymbol{l} \mathbf{u}_{\mathbf{q}}^{\mathrm{T}}) \ \mathrm{d} \mathbf{u}_{\mathbf{q}}$$
(6)



Fig. 2. Example of a badly reconstructed signal due to numerical instability

In order to evaluate the above equation we need to consider the set of all partitions of the integer p

$$\mathcal{P} = \left\{ \underbrace{(\underline{p_1 \dots p_1}_{m_1}, \cdots, \underline{p_L \dots, p_L}_{m_L})}_{m_L} \middle| \sum_{i=1}^L m_i p_i = p \right\}$$

Then we can write,

$$\mu_p(l) = \frac{1}{r^p} \sum_{\mathcal{P}} \sum_{\pi} r^{\sum_{i=1}^L m_i} \prod_{i=1}^L \prod_{h=1}^{m_i} \delta(\sum_{k=1}^{p_i} \ell_{\pi(h,k)})$$
(7)

where $\delta(x) = 1$ for x = 0 and $\delta(x) = 0$ for $x \neq 0$ and the sum over π considers the distinct permutations of the multiset of $\sum_{i=1}^{L} m_i$ distinct elements of multiplicities

$$\underbrace{p_1 \dots, p_1}_{m_1}, \underbrace{p_2 \dots, p_2}_{m_2}, \cdots \underbrace{p_L \dots, p_L}_{m_L}$$

up to permutations of the subsets of the same size p_i . For a given partition, the number of terms resulting from the sum over π is equal to

$$F = \frac{1}{\prod_{i=1}^{L} (m_i!)} \begin{pmatrix} p \\ p_1 \cdots p_1 \ p_2 \cdots p_2 \ \cdots \ p_L \cdots p_L \end{pmatrix}$$

where the term in parenthesis denotes the multinomial coefficient. Note that in (7) for each \mathbf{l} only one term is non null in the sums over \mathcal{P} and π . For example we have

$$\mu_{1}(\ell_{1}) = \begin{cases} 1 \quad \ell_{1} = 0 \\ 0 \quad \text{otherwise} \end{cases}$$
$$\mu_{2}(\ell_{1}, \ell_{2}) = \begin{cases} 1 \quad \ell_{1} = 0, \ell_{2} = 0 \\ 1/r \quad \ell_{1} \neq 0, \ell_{2} = -\ell_{1} \\ 0 \quad \text{otherwise} \end{cases}$$
$$\mu_{3}(\ell_{1}, \ell_{2}, \ell_{3}) = \begin{cases} 1 \quad \ell_{1} = \ell_{2} = \ell_{3} = 0 \\ 1/r \quad \ell_{1} = 0, \ell_{2} \neq 0, \ell_{3} = -\ell_{2} \\ 1/r \quad \ell_{2} = 0, \ell_{1} \neq 0, \ell_{3} = -\ell_{1} \\ 1/r \quad \ell_{3} = 0, \ell_{1} \neq 0, \ell_{2} = -\ell_{1} \\ 1/r^{2} \quad \ell_{1}, \ell_{2}, \ell_{3} \neq 0, \ell_{1} + \ell_{2} + \ell_{3} = 0 \\ 0 \quad \text{otherwise} \end{cases}$$

4.2 Toeplitz distribution theorem [2]

Consider a $(2M + 1) \times (2M + 1)$ Hermitian Toeplitz matrix **T** defined by the sequence of r_{ℓ} for $\ell = -2M \dots, 2M$. Let

$$G(f) = \sum_{\ell=-2M}^{2M} r_\ell \exp(2\pi \mathrm{i} f \ell)$$

be the Fourier transform of the sequence $\{r_\ell\}_{-2M}^{2M}$, then

- the eigenvalues $\lambda_1, \ldots, \lambda_{4M+1}$ of **T** are samples not necessarily equidistant of G(f),
- for any function $\gamma(\cdot)$ we have,

$$\lim_{M \to \infty} \frac{1}{4M+1} \sum_{k=1}^{4M+1} \gamma(\lambda_k) = \int_{-1/2}^{1/2} \gamma(G(f)) \, \mathrm{d}f \tag{8}$$

We will use the above results to evaluate the asymptotic eigenvalue distribution of **T**. We recall that the asymptotic eigenvalue distribution of a \mathbf{HH}^{\dagger} matrix where **H** is a $K \times N$ matrix with independent zero-mean complex random variables with variance 1/N and fourth moments of order $O(1/N^2)$ was given by Marčenko and Pastur [3]. In particular, as $K, N \to \infty$ and $K/N \to \beta$ the empirical distribution of \mathbf{HH}^{\dagger} converges almost surely to a nonrandom limiting distribution with density

$$f_{\beta}(x) = \left(1 - \frac{1}{\beta}\right)^{+} \delta(x) + \frac{\sqrt{(x-a)^{+}(b-x)^{+}}}{2\pi\beta x}$$

where $a = (1 - \sqrt{\beta})^2$, $b = (1 + \sqrt{\beta})^2$ and $(x)^+ = \max(0, x)$. Unfortunately, this result does not apply to $\mathbf{T} = \mathbf{F}\mathbf{F}^{\dagger}$ due to the dependence among the elements of \mathbf{F} , nevertheless it is useful for comparison.

As an example, we plot in Figures 3–5 the experimental eigenvalue distribution of \mathbf{T} , with r = 600 and M = 100, 150, 180 obtained by Monte Carlo simulation. We compare it with the Marčenko–Pastur asymptotic eigenvalue distribution. We observe that both distributions have a bounded support but have significantly different shape as β increases. The bin width is set to 0.1 and prevents from seeing the behavior of the distribution around zero. However, in all cases, the experiments showed that the minimum eigenvalue is bounded away from zero but less than a in the Marčenko–Pastur distribution. This is critical for the condition number of \mathbf{T} , therefore we are interested in evaluating the probability that the minimum eigenvalue is greater than the machine precision, i.e., the probability of correct field reconstruction.



Fig. 3. Comparison between the Marčenko–Pastur distribution and the Monte Carlo experimental distribution for $\beta = 0.34$

By using the Toeplitz distribution theorem [2], we evaluate the moments of the eigenvalue distribution of **T**. In order to compute the *p*-th moment of λ , $\mathbb{E}[\lambda^p]$, we employ (8) where we choose $\gamma(x) = x^p$. Recalling the definition of



Fig. 4. Comparison between the Marčenko–Pastur distribution and the Monte Carlo experimental distribution for $\beta = 0.50$

 $\mu_p(\mathbf{l})$ given in (5) and defining $s_l = \sum_{i=1}^p \ell_i$ we have,

$$\mathbb{E}[\lambda^{p}] = \mathbb{E}\left[\int_{-1/2}^{1/2} G(f)^{p} df\right]$$
$$= \mathbb{E}\left[\int_{-1/2}^{1/2} \sum_{l} r_{\ell_{1}} \cdots r_{\ell_{p}} \exp(2\pi i f s_{\ell}) df\right]$$
$$= \sum_{l} \mathbb{E}\left[r_{\ell_{1}} \cdots r_{\ell_{p}}\right] \int_{-1/2}^{1/2} \exp(2\pi i f s_{l}) df$$
$$= \sum_{l} \mu_{p}(l) \frac{\sin(\pi s_{l})}{2\pi s_{l}} = \sum_{l} \mu_{p}(l) \delta(s_{l})$$
(9)

In order to evaluate this sum we need to count the number $q_n(M)$ of integer vectors with n components ranging from -2M to -1 and from 1 to 2M adding up to zero. Using the generating functions method we find $q_n(M)$ for $n \ge 2$, as



Fig. 5. Comparison between the Marčenko–Pastur distribution and the Monte Carlo experimental distribution for $\beta=0.60$

the constant term of the rational function

$$\left(\sum_{k=1}^{2M} x^k + x^{-k}\right)^n$$

The general formula cannot be found in closed form and needs to be evaluated explicitly for each \boldsymbol{n}

$$q_2(M) = 4M$$

$$q_3(M) = 12M^2 - 6M$$

$$q_4(M) = (128/3)M^3 - 16M^2 + (28/3)M$$

$$q_5(M) = (460/3)M^4 - 60M^3 + (65/3)M^2 - 15M$$

We further define $q_1(M) = 1$. We can finally write

$$\mathbb{E}[\lambda^{p}] = \frac{1}{r^{p}} \sum_{\mathcal{P}} Fr^{\sum_{i=1}^{L} m_{i}} \prod_{i=1}^{L} q_{p_{i}}(M)^{m_{i}}$$
(10)

and in particular

$$\mathbb{E}[\lambda] = 1$$

$$\begin{split} \mathbb{E}[\lambda^2] &= 1 + \frac{4M}{r} \\ \mathbb{E}[\lambda^3] &= 1 + \frac{12M}{r} + \frac{12M^2 - 6M}{r^2} \\ \mathbb{E}[\lambda^4] &= 1 + \frac{24M}{r} + \frac{96M^2 - 24M}{r^2} + \frac{(128/3)M^3 - 16M^2 + (28/3)M}{r^3} \end{split}$$

In the limit for $M\to\infty$ and $r\to\infty$ with constant $\beta=2M/r$ we get the following central moments

$$\mathbb{E}[(\lambda - 1)^2] = 2\beta \\ \mathbb{E}[(\lambda - 1)^3] = 3\beta^2 \\ \mathbb{E}[(\lambda - 1)^4] = 12\beta^2 + (16/3)\beta^3 \\ \mathbb{E}[(\lambda - 1)^5] = 60\beta^3 + (115/12\beta^4)$$

We observe how the eigenvalues are concentrated around 1 when β is small. Recall that M is the field bandwidth and r is the number of samples, i.e., the number of active sensors over the observation area. We are interested in deriving the eigenvalue distribution as $M \to \infty$ and $r \to \infty$ with constant $\beta = 2M/r < 1$. Indeed, given the eigenvalue distribution we could determine the condition number of the matrix \mathbf{T} , hence the probability of correct field reconstruction. This will be the next step in our work.

5 Conclusions and Future Work

We considered a large-scale wireless sensor network sampling a physical field, and we investigated the relationship between the random network topology and the quality of the reconstructed field. We employed random matrix theory, and gave the basis to derive the ratio of the field bandwidth to the number of samples necessary for a successful reconstruction.

In our future research, we need to address several issues. First of all, we would like to obtain the asymptotic eigenvalue distribution of the matrix employed for the field reconstruction, so that the condition number, i.e., the probability of correct field reconstruction can be determined. A two-dimensional field should be analyzed, and several aspects should be taken into account. For instance, the fact that sensors can represent the detected information with a limited number of bits, data (i.e., sample) losses can occur during the information transfer to the sink, and the sensor locations can be unknown, should all be included in the analysis.

References

 H.G. Feichtinger, K. Gröchenig, T. Strohmer, "Efficient numerical methods in nonuniform sampling theory," *Numerische Mathematik*, vol. 69, pp. 423–440, 1995.

- 2. U. Grenander and Szegö, *Toeplitz forms and their Applications*, Bekeley, CA, Univ. of California Press.
- V.A. Marčenko and L.A. Pastur, "Distributions of eigenvalues for some sets of random matrices," *Math. USSR-Sbornik*, vol. 1, pp. 457–483, 1967.
- P. Ishwar, A. Kumar, K. Ramchandran, "Distributed sampling for dense sensor networks: a bit-conservation principle," 3rd International Symposium on Information Processing in Sensor Networks (IPSN 2003), Apr. 2003.
- M. Perillo, Z. Ignjatovic, W. Heinzelman, "An energy conservation method for wireless sensor networks employing a blue noise spatial sampling technique," 3rd International Symposium on Information Processing in Sensor Networks (IPSN 2004), Apr. 2004.
- R. Willett, A. Martin, R. Nowak, "Backcasting: adaptive sampling for sensor networks," 3rd International Symposium on Information Processing in Sensor Networks (IPSN 2004), Apr. 2004.
- Y. Yu, D. Ganesan, L. Girod, D. Estrin, R. Govindan, "Synthetic data generation to support irregular sampling in sensor networks," *Geo Sensor Networks 2003*, Portland, Maine, Oct. 2003.