

Designing Optimal Sampling Schemes

Johan Swärd

Centre for Mathematical Sciences
Lund University, Sweden
Email: js@maths.lth.se

Filip Elvander

Centre for Mathematical Sciences
Lund University, Sweden
Email: filipelv@maths.lth.se

Andreas Jakobsson

Centre for Mathematical Sciences
Lund University, Sweden
Email: aj@maths.lth.se

Abstract—In this work, we propose a method for finding an optimal, non-uniform, sampling scheme for a general class of signals in which the signal measurements may be non-linear functions of the parameters to be estimated. Formulated as a convex optimization problem reminiscent of the sensor selection problem, the method determines an optimal sampling scheme given a suitable estimation bound on the parameters of interest. The formulation also allows for putting emphasis on a particular set of parameters of interest by scaling the optimization problem in such a way that the bound to be minimized becomes more sensitive to these parameters. For the case of imprecise *a priori* knowledge of these parameters, we present a framework for customizing the sampling scheme to take such uncertainty into account. Numerical examples illustrate the efficiency of the proposed scheme.

I. INTRODUCTION

Determining how to suitably sample a signal is an important problem in many signal processing applications, such as sensor positioning and selection in network monitoring [1], [2], localization and tracking [3], and selecting the temporal sampling [4]. As an example, in nuclear magnetic resonance (NMR) spectroscopy, one is typically interested in sampling a multi-dimensional field containing partly known signal components. For high-dimensional data, it quickly becomes infeasible to sample the field uniformly, especially when examining living cells, which have limited lifetimes. This has caused an interest in formulating sampling schemes for NMR signals, allowing for notable improvements [4]–[8]. For example, a recent study of 4-D NMR measurements that would have taken about 2.5 years to perform using regular sampling was shown to be possible to construct in merely 90 hours using a non-uniform sampling scheme [9].

Among the developed schemes are some exploiting a compressive sensing framework, allowing for an accurate signal reconstruction using fewer samples than the Nyquist-Shannon sampling theorem necessitates for uniformly sampled signals (see, e.g., [7], [8]). However, the developed schemes typically do not optimize the sampling scheme with respect to the expected signals, even though these are often fairly well known. In this work, we strive to exploit this knowledge in order to design a sampling scheme that would allow for an optimal estimation accuracy given the assumed prior knowledge.

Lately, for the related problem of optimal sensor placement, there has been several methods proposed in which the combinatorial problem of selecting a subset of sensors is relaxed using convex optimization. In [10], the authors consider the

case when signal measurements are linear in the unknown parameters and propose a sensor selection scheme based on solving a convex optimization problem inspired by the determinant criterion (D-optimality) of experimental design [11]. This work was then developed in [12], [13], wherein the authors consider non-linear measurement equations, as well as replacing D-optimality with the average variance criterion (A-optimality) as a performance measure. Specifically, as A-optimality can be interpreted as the sum of the diagonal elements of the Cramér-Rao lower bound (CRLB) for the signal parameters, the problem was formulated as to minimize the number of required sensors subject to an upper bound on the resulting diagonal sum of the CRLB. Assuming that the bound is tight, the method thus finds a sparse set of sensors, i.e., activates a few out of a set of candidate sensors, while keeping the variance of the estimated parameters below a fixed level.

In this paper, we expand on this idea, proposing a method for finding an optimal sampling scheme in order to estimate the parameters for signal models where, in general, the signal measurements are non-linear functions of the unknown parameters. By taking the available prior information of the signal into consideration, we propose a sampling scheme that is found by solving a convex optimization problem that guarantees a bound on the worst case CRLB. The sampling pattern is selected via a variable vector, corresponding to the available sample positions, which is penalized using the ℓ_1 -norm, resulting in a sampling scheme that is limited in the number of samples. In general, when estimating a set of parameters, it might be that the scale of the parameters, as well as the accuracy with which they can be estimated, are significantly different. Also, some of the unknown parameters might be of greater interest than the others; again, using NMR as an example, the signal decay is often of more interest than the signal frequencies, the latter often being relatively well known for a given substance, whereas the former measures the sought interactions. We here propose to use a weighting scheme in order to allow for a relative balancing of the variances of the different parameters, allowing for designing sampling schemes specifically tailored to yield good estimation accuracy for the parameters of interest.

In some applications, one may assume some prior knowledge of the signal of interest, such as, for example, knowledge of the subspace where the signal parameters are to be found. Again using NMR as an illustrative example, the signals of

interest consist of decaying modes, being well modeled as a sum of damped sinusoids. These modes are, as noted, often well known in frequency, at least within some reasonably well defined frequency band, whereas the uncertainty of, and the interest in, the signal decays is often more significant. Typically, the problem of interest is thus to specify the damping parameter as accurately as possible using as few samples as possible. To allow for this case, we herein propose using a gridding of the parameter space in order to guarantee performance within certain bounds, allowing for uncertainty in the parameters.

II. PROBLEM STATEMENT AND PROPOSED SAMPLING SCHEME

Consider a measured signal $y(\mathbf{t}_n)$, defined on a D -dimensional space with N potential D -dimensional sampling points, \mathbf{t}_n , $n = 1, 2, \dots, N$. It is assumed that the probability density function (pdf) of $y(\mathbf{t}_n)$, here denoted with $p(y(\mathbf{t}_n); \boldsymbol{\theta})$, is parametrized by the parameter vector $\boldsymbol{\theta}$ and that two samples $y(\mathbf{t}_n)$ and $y(\mathbf{t}_m)$ are independent if $\mathbf{t}_n \neq \mathbf{t}_m$. The Fisher information matrix (FIM) for sample $y(\mathbf{t}_n)$ may then be defined as

$$\mathbf{F}(\mathbf{t}_n; \boldsymbol{\theta}) = \mathbb{E} \left\{ \nabla_{\boldsymbol{\theta}} \log(p(y(\mathbf{t}_n); \boldsymbol{\theta})) \nabla_{\boldsymbol{\theta}}^H \log(p(y(\mathbf{t}_n); \boldsymbol{\theta})) \right\} \quad (1)$$

where $\mathbb{E}\{\cdot\}$, $\nabla_{\boldsymbol{\theta}}$, and $(\cdot)^H$ denote the statistical expectation, the gradient with respect to $\boldsymbol{\theta}$, and the conjugate transpose, respectively. The here proposed sampling scheme is designed such that it is optimal in the sense of either minimizing the CRLB of the parameters of interest, given that M of the N potential uniform samples are used, or conversely, to minimize the number of samples used given a desired upper bound on the CRLB of the parameters. Note that, as the potential signal samples are assumed to be independent, for any set of samples indices Ω , it holds that

$$\sum_{n \in \Omega} \mathbf{F}(\mathbf{t}_n; \boldsymbol{\theta}) \quad (2)$$

is the corresponding FIM using this sample scheme. Let the N -dimensional vector \mathbf{w} denote the possible sampling points in the D -dimensional sampling space, such that if the n :th index, w_n , is set to one, this sampling point is used, whereas if it is set to zero, it is not. Reminiscent of the case of optimal sensor selection, the resulting sampling design problem may then be formulated as (see also [12])

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && \|\mathbf{w}\|_1 \\ & \text{subject to} && \text{tr} \left(\left(\sum_{n=1}^N w_n \mathbf{F}(\mathbf{t}_n; \boldsymbol{\theta}) \right)^{-1} \right) \leq \lambda \\ & && w_n \in \{0, 1\}, \quad n = 1, 2, \dots, N \end{aligned} \quad (3)$$

where $\lambda > 0$ and $\text{tr}(\cdot)$ denotes the trace operator. The use of the trace constraint corresponds to the so-called A-optimality criterion from design of experiments [11]. As the trace of the inverse FIM corresponds to the sum of the CRLB of the signal

parameters in $\boldsymbol{\theta}$, the parameter λ constitutes an upper bound on the sum of the lower bounds of the variances of the elements of $\boldsymbol{\theta}$. The sampling design scheme (3) is not convex due to the restriction that w_n , for $n = 1, \dots, N$, is defined over a non-convex set. A convex approximation to this problem may be found by relaxing the binary constraint and instead allowing w_n to take any value in the range $[0, 1]$ (see, e.g., [13]), resulting in

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && \|\mathbf{w}\|_1 \\ & \text{subject to} && \text{tr} \left(\left(\sum_{n=1}^N w_n \mathbf{F}(\mathbf{t}_n; \boldsymbol{\theta}) \right)^{-1} \right) \leq \lambda \\ & && w_n \in [0, 1], \quad n = 1, 2, \dots, N \end{aligned} \quad (4)$$

Given a solution $\hat{\mathbf{w}}$ to (4), we define the FIM for the corresponding sampling pattern as

$$\mathcal{I}(\hat{\mathbf{w}}; \boldsymbol{\theta}) = \sum_{\ell \in \Omega} \mathbf{F}(\mathbf{t}_\ell; \boldsymbol{\theta}), \quad \Omega = \{\ell \mid \hat{w}_\ell > \xi\} \quad (5)$$

where $\xi \geq 0$ is a threshold determining whether a sample weight \hat{w}_ℓ should be rounded toward one or zero, i.e., whether the sampling point should be included or not. This formulation allows for the minimization of the sample size, given an upper bound on the sum of the CRLBs of the parameters, or, conversely, to find the sample scheme yielding the lowest sum of the CRLBs given an upper bound on the number of samples used.

The sampling design in (4) does not allow for the case when one is primarily interested in a subset of the available parameters, as, for example, is the case in NMR. Neither does the formulation take into account that the different parameters might have significantly different variances. For example, for a sum of damped sinusoids, the trace constraint in (4) will clearly be dominated by the CRLB for the amplitudes, as these are orders of magnitude larger than those of the frequencies, and the optimization will therefore put an emphasis on minimizing the CRLB of the amplitude parameter. In order to allow for sampling schemes that put an emphasis on a selection of the parameters of interest, we propose to introduce a weighting matrix, $\mathbf{A}(\boldsymbol{\theta})$, acting upon the FIM. Specifically, instead of minimizing the cost function using the FIM, we instead introduce a weighted FIM in the optimization. Let

$$\tilde{\mathbf{F}}(\mathbf{t}_n; \boldsymbol{\theta}) = \mathbf{A}(\boldsymbol{\theta}) \mathbf{F}(\mathbf{t}_n; \boldsymbol{\theta}) \mathbf{A}^T(\boldsymbol{\theta}) \quad (6)$$

This weighting corresponds to performing a linear transformation of the parameters and letting the constraint be formed on the sum of the CRLBs for the transformed parameters $\tilde{\boldsymbol{\theta}} = \mathbf{A}(\boldsymbol{\theta})\boldsymbol{\theta}$. For example, the weighting matrix $\mathbf{A}(\boldsymbol{\theta})$ may be formed as a diagonal loading, where the diagonal elements are set to be the roots of the CRLB corresponding to the case when all N samples are used. Furthermore, one may not only be interested in designing a sampling scheme for a single parameter vector $\boldsymbol{\theta}$, but rather for a set of parameter vectors. For example, consider the case when the parameters in $\boldsymbol{\theta}$ are only partly known, such that one may assume that $\boldsymbol{\theta}$

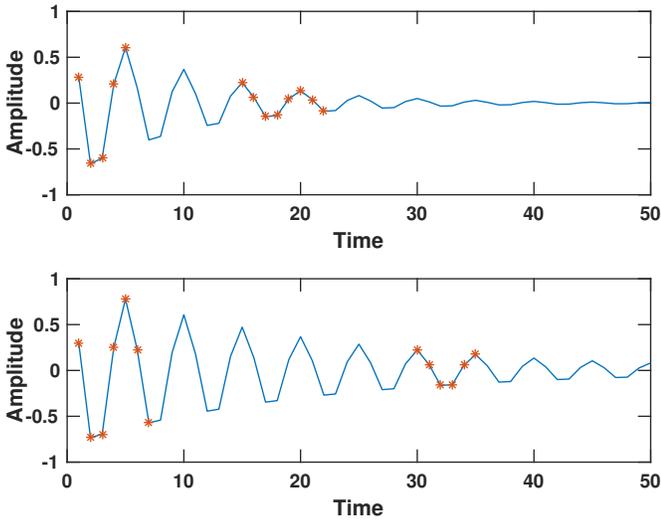


Fig. 1. The resulting sample scheme for two different values of β plotted against the real part of the signal. The upper most figure details the sampling scheme for $\beta = \frac{1}{10}$ and the bottom figure the sampling scheme for $\beta = \frac{1}{20}$.

instead lies in a set of possible parameters, Θ . In this case, one may, e.g., treat some of the parameters as known, whereas others are only partly known, within some set of uncertainty. To allow for this, as well as taking the weighting into account, we generalize (4) such that the sampling scheme is designed as

$$\begin{aligned} & \underset{\mathbf{w}}{\text{minimize}} && \|\mathbf{w}\|_1 \\ & \text{subject to} && \text{tr} \left(\left(\sum_{n=1}^N w_n \tilde{\mathbf{F}}(t_n; \boldsymbol{\theta}) \right)^{-1} \right) \leq \lambda, \forall \boldsymbol{\theta} \in \Theta \quad (7) \\ & && w_n \in [0, 1], n = 1, 2, \dots, N \end{aligned}$$

To illustrate the proposed sampling scheme, we consider the NMR signal model, as noted being formed as a sum of damped sinusoids (for ease of notation, we focus on the 1-D case), such that

$$y(t_n) = \sum_{k=1}^K \alpha_k \exp\{2i\pi f_k t_n - \beta_k t_n + i\phi_k\} + \epsilon(t_n) \quad (8)$$

for $n = 1, \dots, N$, where α_k , f_k , β_k , and ϕ_k are the frequency, damping, and phase of the k :th component, respectively, and where ϵ is an additive noise term, here assumed to be well modeled as a white Gaussian noise, with N being the number of samples, and t_n the time at sample n . For simplicity, we consider uniformly sampled candidate sampling times, t_n . As an illustration, Figure 1 shows an example of sampling schemes found by solving (7) for two different levels of decay for a single damped sinusoid such that $\beta = 1/10$ for the top figure, and $\beta = 1/20$ for the bottom figure, but otherwise identical signal parameters. In both cases, λ has been chosen as to select $M = 13$ sample points out of $N = 50$ possible candidates. As can be seen, the placing of the samples are determined by the damping parameter. As may be expected, for both values of β , some samples are placed in the beginning

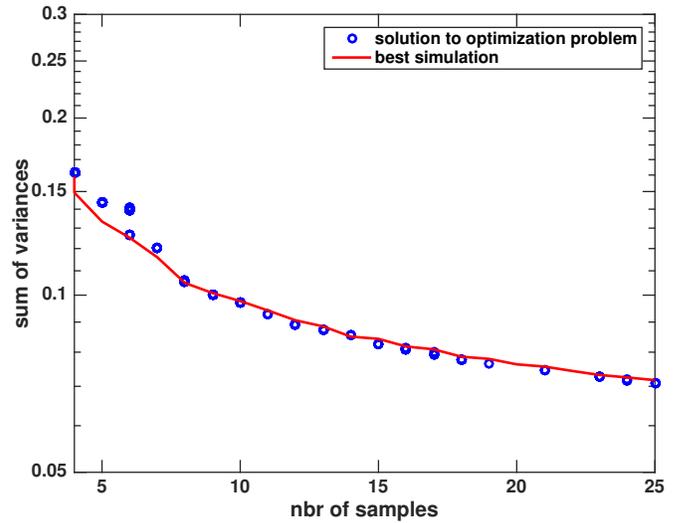


Fig. 2. Sum of CRLBs for the parameters, i.e., $\text{tr}(\mathcal{I}(\hat{\mathbf{w}}; \boldsymbol{\theta})^{-1})$ for the sampling patterns given by the optimization problem and the best simulation, respectively, for different number of sampling points.

of the signal, where the signal to noise ratio (SNR) is at its maximum. To allow for an accurate estimation of the damping constant, one can also note that a further set of samples are selected later in the signal, with the more strongly decaying signal selecting them earlier than the less damped version, agreeing with the intuition that the more rapidly decaying signal contains less information at later sampling times.

III. NUMERICAL RESULTS

A. Optimization vs simulation

In Figure 2, we motivate that solving (7) is indeed a reasonable approach to determine optimal sampling patterns. The figure shows the obtained sum of the CRLBs for the parameters, i.e., $\text{tr}(\mathcal{I}(\hat{\mathbf{w}}; \boldsymbol{\theta})^{-1})$, where the sampling pattern is obtained by solving (7) for the case of $K = 1$ using the model (8), for a singleton set Θ and identity weighting $\mathbf{A}(\boldsymbol{\theta}) = \mathbf{I}$. This is done for varying values of λ such that the number of samples used vary between $M = 5$ and $M = 25$. As a comparison, for each sample size M , we carry out 10^6 Monte Carlo simulations, in which we randomly decide on which M sampling points to use. We then compute which of these 10^6 sampling patterns that results in the lowest sum of CRLBs. As can be seen from the figure, the randomized approach achieves better results for small sample sizes, this as the simulations then become an exhaustive search, i.e., the simulations will with high likelihood find the solution to (3). However, as the sample size increases, so does the number of possible sampling patterns, which is $\binom{N}{M}$. As can be seen from the figure, the sampling scheme determined by (7) is then able to achieve an optimal performance as the sample size increases.

B. Weighting

In Figures 3 and 4, we proceed to examine the effect of using the weighted FIM in (7). This is

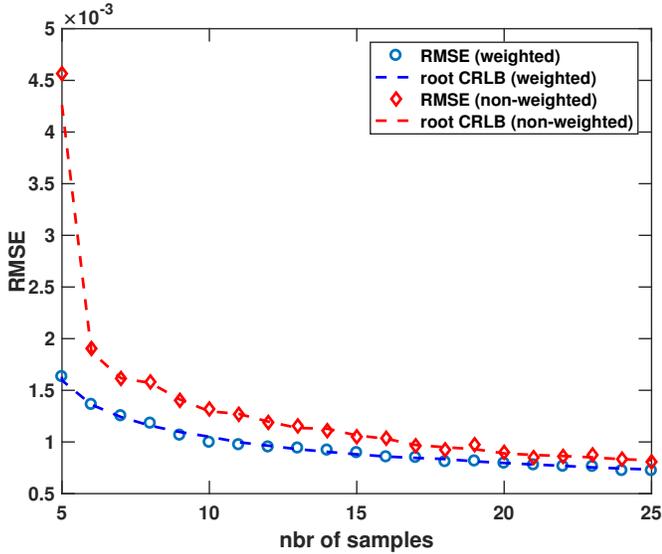


Fig. 3. Obtained RMSE for the frequencies, when using the sampling patterns for the weighted and non-weighted cases, respectively.

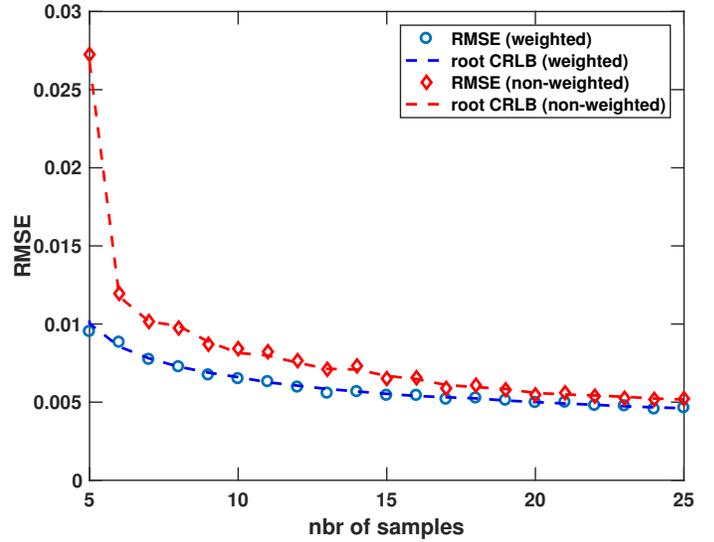


Fig. 4. Obtained RMSE for the damping, when using the sampling patterns for the weighted and non-weighted cases, respectively.

done for a signal consisting of two damped sinusoids with parameters $(\alpha_1, f_1, \beta_1, \varphi_1) = (1, 0.2, 1/12, 0.5)$ and $(\alpha_2, f_2, \beta_2, \varphi_2) = (1, 0.65, 1/20, \pi/5)$. The noise variance was $\sigma^2 = 0.01$ and $N = 50$. Assuming that we are interested only in the frequencies f_1, f_2 , and the damping factors β_1, β_2 , but not in the amplitudes or the phases, the weighting matrix $\mathbf{A}(\theta)$ should be designed to put emphasis on the former parameters. To construct $\mathbf{A}(\theta)$, which for simplicity is chosen to be a diagonal matrix, we first compute the full FIM corresponding to the sample pattern where all N samples are used. The diagonal elements of $\mathbf{A}(\theta)$ corresponding to the amplitudes and phases are then chosen as the roots of the corresponding diagonal elements of the inverse full FIM, while the diagonal elements corresponding to the frequencies and the dampings are set to unity. Thus, the sought sampling pattern will be designed to increase the accuracy for the frequency and damping parameters at the expense of the amplitude and phase parameters. The resulting root CRLB, as a function of the number of samples used, for the frequencies f_1 and f_2 and the dampings β_1 and β_2 are shown in Figures 3 and 4, respectively. The root CRLB for the frequencies f_1 and f_2 is here defined as the root of the sum the individual CRLBs, and correspondingly for the dampings, β_1 and β_2 . For comparison, the figures also present the root CRLBs corresponding to the optimal sampling patterns obtained for the case when no weighting is applied to the FIM, i.e., with $\mathbf{A}(\theta)$ being the identity matrix. As can be seen, the weighting scheme results in sampling patterns that decreases the CRLB for the parameters of interest, in this case the frequencies and dampings. Also plotted is the obtained root mean squared error (RMSE) for the frequency and damping parameters, respectively, obtained when estimating these parameters using non-linear least squares (NLS) applied to simulated signals. The RMSE is here defined as the root of the sum of the individual MSEs for the frequencies and

dampings, respectively. As can be seen, the RMSE coincides with the root CRLB, implying that the bound is tight.

C. Gridding

Figures 5 and 6 show the effect of finding an optimal sampling pattern for a set of parameters $\theta \in \Theta$ when solving (7). The results are obtained for a single decaying sinusoid. Here, we let $\Theta = \{\theta_\ell\}_{\ell=1}^L$ express uncertainty in only the damping parameter β by fixing α, f , and φ and letting Θ be a gridding over the damping parameter β , such that the parameter vectors constituting Θ are $\theta_\ell = (\alpha, f, \beta_\ell, \varphi)^T$ where

$$\beta_\ell = \beta_{\text{lower}} + \frac{\ell - 1}{L} \Delta_\beta \quad (9)$$

with Δ_β denoting the grid spacing, in effect letting β reside in the uncertainty interval

$$\mathcal{J}_\beta = \left[\beta_{\text{lower}}, \beta_{\text{lower}} + \frac{L - 1}{L} \Delta_\beta \right] \quad (10)$$

The parameters used are $\alpha = 1$, $\varphi = 0.5$, $\sigma^2 = 0.1$, $\beta_{\text{lower}} = 0.1$, $\Delta_\beta = 0.022$, and $L = 10$. Using this, we solve (7) to get optimal sampling patterns as the number of samples grows. To evaluate the performance of the obtained sampling schemes, we then randomly sample the parameter vectors θ where β is sampled uniformly on \mathcal{J}_β , i.e., on the interval covered by the grid Θ , but not on the grid points β_ℓ , $\ell = 0, 1, \dots, L - 1$. We then estimate θ using NLS and compute the RMSE for the parameters θ . The figures show the obtained MSE using 5000 Monte Carlo simulations for the frequency f and the damping β , respectively. Also presented are the best and worst case root CRLBs found on the grid Θ for each parameter. The obtained RMSE lies between the lowest and highest on-grid root CRLB for both parameters and for all considered sample sizes, suggesting that (7) indeed yields

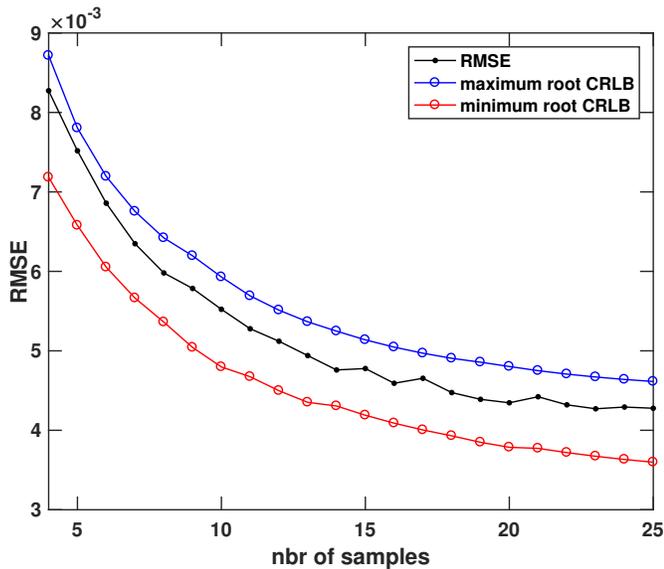


Fig. 5. Obtained RMSE for the frequency f , when estimating θ for the sampling pattern obtained for a grid of damping parameters β .

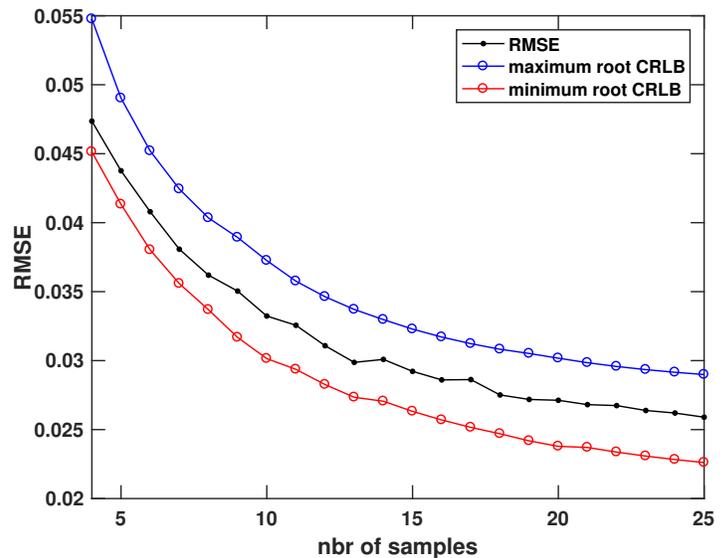


Fig. 6. Obtained RMSE for the damping β , when estimating θ for the sampling pattern obtained for a grid of damping parameters β .

sampling schemes with a guaranteed worst case performance, as well as a lower limit on the possible RMSE.

REFERENCES

- [1] S. Liu, M. Fardad, E. Masazade, and P. K. Varshney, "Optimal Periodic Sensor Scheduling in Networks of Dynamical Systems," *IEEE Trans. Sig.*, vol. 62, no. 12, pp. 3055–3068, December 2014.
- [2] H. Jamali-Rad, A. Simonetto, X. Ma, and G. Leus, "Distributed Sparsity-Aware Sensor Selection," *IEEE Trans. Signal Process.*, vol. 63, no. 22, pp. 5951–5964, November 2015.
- [3] S. P. Chepuri, G. Leus, and A. J. van der Veen, "Sparsity-Exploiting Anchor Placement for Localization in Sensor Networks," in *21st European Signal Processing Conference*, 9–13 September 2013, pp. 1–5.
- [4] P. Schmieder, A. S. Stern, G. Wagner, and J. C. Hoch, "Application of nonlinear sampling scheme to COSY-type spectra," *Journal of Biomolecular NMR*, vol. 3, pp. 569–576, 1993.
- [5] S. G. Hyberts, K. Takeuchi, and G. Wagner, "Poisson-Gap Sampling and Forward Maximum Entropy Reconstruction for Enhancing the Resolution and Sensitivity of Protein NMR Data," *J Am Chem Soc.*, vol. 132, pp. 2145–2147, 2010.
- [6] P. J. Sidebottom, "A new approach to the optimisation of non-uniform sampling schedules for use in the rapid acquisition of 2D NMR spectra of small molecules," *Magn Reson Chem*, vol. 54, no. 8, pp. 689–694, August 2016.
- [7] K. Kazimierczuk and V. Y. Orekhov, "Accelerated NMR Spectroscopy by Using Compressed Sensing," *Angewandte Chemie International Edition*, vol. 50, no. 24, June 2011.
- [8] K. Kazimierczuk and V. Y. Orekhov, "A comparison of convex and non-convex compressed sensing applied to multidimensional NMR," *J. Magn. Reson.*, vol. 223, pp. 1–10, 2012.
- [9] K. Kazimierczuk, A. Zawadzka-Kazimierczuk, and W. Kozmiński, "Non-uniform frequency domain for uniform exploitation of non-uniform sampling," *J. Magn. Reson.*, vol. 205, pp. 286–292, 2010.
- [10] S. Joshi and S. Boyd, "Sensor Selection via Convex Optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, February 2009.
- [11] F. Pukelsheim, *Optimal design of experiments*, Wiley series in probability and mathematical statistics. Wiley, New York, 1993.
- [12] S. P. Chepuri, *Sparse Sensing for Statistical Inference - Theory, Algorithms, and Applications*, Ph.D. thesis, Delft University of Technology, 2015.
- [13] S. P. Chepuri and G. Leus, "Sparsity-Promoting Sensor Selection for Non-Linear Measurement Models," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 684–698, February 2015.