LUND UNIVERSITY

Designing Sampling Schemes for Multi-Dimensional Data

J. SWÄRD, F. ELVANDER, AND A. JAKOBSSON

Published in: Elsevier Signal Processing

doi:10.1016/j.sigpro.2018.03.011

Lund 2018

Mathematical Statistics

Centre for Mathematical Sciences

Lund University

Designing Sampling Schemes for Multi-Dimensional Data ☆

J. Swärd[a,*], F. Elvander[a], A. Jakobsson[a]

[a] *Department of Mathematical Statistics, Lund University, P.O. Box 118, SE-221 00 Lund, Sweden*

Abstract

In this work, we propose a method for determining a non-uniform sampling scheme for multi-dimensional signals by solving a convex optimization problem reminiscent of the sensor selection problem. The resulting sampling scheme minimizes the sum of the Cramér-Rao lower bounds for the parameters of interest, given a desired number of sampling points. The proposed framework allows for selecting an arbitrary subset of the parameters detailing the model, as well as weighing the importance of the different parameters. Also presented is a scheme for incorporating any imprecise *a priori* knowledge of the locations of the parameters, as well as defining estimation performance bounds for the parameters of interest. Numerical examples illustrate the efficiency of the proposed scheme.

1. 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], magnetic resonance imaging (MRI) [4], graph signal processing [5, 6], and selecting the temporal sampling [7]. In general, these problems can be viewed as sampling a multi-dimensional field containing partly known signal components. For high-dimensional data, it quickly becomes infeasible to sample the field uniformly, especially, in areas such as nuclear magnetic resonance (NMR) spectroscopy when examining living cells, which have limited lifetimes. 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 [8]. This has caused an interest in formulating sampling schemes for NMR signals, allowing for notable improvements [7, 9–13].

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., [11, 12, 14, 15]). However, the developed schemes typically do not optimize the sampling scheme with respect to the expected signals, even though these are often fairly well

☆This work was supported in part by the Swedish Research Council, Carl Trygger's foundations, and the Swedish strategic research program eSSENCE. Initial work treating the formulation of this paper has been accepted for publication at the EUSIPCO 2017 and Asilomar 2017 conferences.

*Corresponding author. Phone: +46735331037.

Email addresses: `js@maths.lth.se` (J. Swärd), `filipelv@maths.lth.se` (F. Elvander), `aj@maths.lth.se` (A. Jakobsson)

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.

There are many related problems to the herein studied sampling scheme problem. In [16], the problem of how to optimally measure a signal in problems related to propagating wave-fields was studied. More specifically, the authors studied how to best recover the input wave field from noisy measurements of the output field given that each measurement is associated with a cost, where the selected cost was set higher for measurement devices with better resolution. The results were presented as trade-off curves between the error of estimation and the total cost budget. In [17], a framework for joint hypothesis testing and estimation using a minimal sample size was developed. The proposed framework guarantees, under a Bayesian setup, that the overall detection and estimation performance, given the minimization of the sample size, is the best possible. In [18], the optimal placement of phasor measurement units on power grids was studied. Other works have been studying problems related to sampling in random fields [19, 20] and wireless sensor networks [21]. A notable example of the latter category is [21], where the problem of target tracking in wireless sensor networks is studied. The sensors with the most information are found by utilizing a proposed probabilistic sensor management scheme based on the compressed sensing framework. This scheme is determined based on the probability of transmission at each node, found by maximizing the trace of the Fisher information matrix (FIM). Using this approach, sensors with less information can be discarded, implying that fewer sensors need to communicate, thus leading to energy savings.

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 [22], 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 [23]. This work was then further developed in [2, 24–27], 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 a suitable 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. Furthermore, we reformulate the optimization problem into a semidef-

inite programming (SDP) problem that allows for more flexibility and can be used for adding additional constraints on the optimization. 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.

To summarize, the main components of this paper are:

- The development of a unified framework for incorporating strong prior knowledge when designing a sampling scheme, such that one may determine which samples that are most appropriate to select for a signal that may contain the specified components. It should be stressed that any assumed component does not actually need to be present in the signal; the scheme will select samples that are sufficient to estimate signal components at the considered locations well, but these may be absent in an actual measurement.

- The method allows samples to be selected such that some parameters are deemed more important, i.e., of higher interest for the application, than others, and the sample scheme is prioritized accordingly. For instance, one is often more interested in the decay rate of the signal components than their actual frequencies (which are commonly well known), and may then focus the sampling scheme to allow for accurate estimation of these parameters.

- The method allows for uncertainty in the assumed model, allowing a user to indicate regions of interest wherein parameters are assumed to lie. This allows the proposed method to design sampling patterns for signals that are only partly known, relaxing the assumption of the signal modelled being known in detail *a priori*.

- The proposed method is valid for high-dimensional data sets, with the achievable gain growing (rapidly) with the growing dimensionality of the problem.

To illustrate the proposed method, we consider multi-dimensional NMR signals. However, the proposed framework is not limited to this case. Indeed, it is applicable to any signal model having a parametric description that allows for the existence of a Fisher information matrix. Although, we show that the proposed method yields Cramér-Rao lower bounds that are attainable, it should be stressed that this paper is not concerned with parameter estimation as such; any efficient estimator allowing for non-uniformly sampled data may be applied.

This paper is organized as follows. In Section 2, we introduce the problem statement and derive the proposed optimization problem. In Section 3, we present extensive numerical simulations and results that validates our proposed method. Finally, in Section 4, we conclude upon our work.

2. 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, $\mathbf{t}_n$, $n = 1, 2, \ldots, N$. It is assumed that the probability density function (pdf) of $y(\mathbf{t}_n)$, here denoted with $p\left(y(\mathbf{t}_n); \boldsymbol{\theta}\right)$, is parametrized by the parameter vector $\boldsymbol{\theta} \in \mathbb{R}^P$ and that two samples $y(\mathbf{t}_n)$ and $y(\mathbf{t}_m)$ are independent if $\mathbf{t}_n \neq \mathbf{t}_m$. FIM for sample $y(\mathbf{t}_n)$ may then be defined as

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

where $\mathbb{E}\left\{\cdot\right\}$, $\nabla_{\boldsymbol{\theta}}$, 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. It is worth noting that as the potential signal samples are assumed to be independent, for any set of samples indices $\boldsymbol{\Omega}$, it holds that

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

is the corresponding FIM using this sample scheme. Let the N-dimensional vector $\mathbf{w}$ denote an indicator for the possible sampling points in the D-dimensional sampling space, such that if the nth 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 [24])

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

where $\gamma > 0$ and $\operatorname{tr}(\cdot)$ denotes the trace operator. The choice of objective function is related to the so-called A-optimality criterion from design of experiments [23] as the trace of the inverse FIM corresponds to the

sum of the CRLBs of the signal parameters in $\boldsymbol{\theta}$. Here, the parameter γ constitutes an upper bound on the ℓ_1-norm of the sample selection vector. The sampling design scheme (3) is not convex due to the restriction that w_n, for $n = 1, \ldots, N$, 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., [25]), resulting in

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

where $\mathbf{1}$ is a vector of ones of appropriate dimension. It should be noted that we can here replace $||\mathbf{w}||_1$ with simply $\mathbf{1}^T\mathbf{w}$, since each element in $\mathbf{w}$ is equal to or greater than zero. Given a solution $\hat{\mathbf{w}}$ to (4), we define the FIM for the corresponding sampling pattern as

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

where $\xi \geq 0$ is a threshold determining whether a sample weight $\hat{w}_\ell$ 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 sum of the CRLBs given an upper bound on the number of samples used. Note that the problem could alternatively be formulated as minimizing the number of sampling points given an upper bound on the sum of the CRLBs. It is also worth noting that the signal to noise ratio (SNR) will not affect the sampling scheme as such, as the noise power will only affect the optimisation problem by scaling the cost function. Obviously, the SNR will result the performance of any estimator subsequently used to estimate the signal parameters using the selected samples, but thus not the actual selection of these samples.

However, the sampling design in (4) does not allow for the case when one is primarily interested in a subset of the available parameters. 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 for 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 recently proposed to introduce a weighting matrix, $\mathbf{A}(\boldsymbol{\theta})$, acting upon the FIM in [28]. Specifically, instead of minimizing the cost function using the FIM, we proposed to perform the minimization using weighted FIMs

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

i.e., performing a linear transformation of the variables and minimizing the sum of the CRLBs corresponding to the transformed parameters $\tilde{\boldsymbol{\theta}} = \mathbf{A}(\boldsymbol{\theta})\boldsymbol{\theta}$. However, although this formulation allows for shifting emphasis to the parameters of interest, it does not allow for complete disregard of nuisance parameters as $\mathbf{A}(\boldsymbol{\theta})$ has

to be non-singular in order for the matrix inverse to be defined. In order to allow for an arbitrary weighting, we note that the following useful identity holds for any invertible $P \times P$ matrix $\mathbf{B}$;

$$\mathrm{tr}\left(\mathbf{B}^{-1}\right) = \sum_{p=1}^{P} \mathbf{e}_p^T \mathbf{B}^{-1} \mathbf{e}_p, \tag{7}$$

where $\mathbf{e}_p$ denotes the pth canonical basis vector, i.e., a vector with all its elements equal to zero except the pth being equal to one. Furthermore, it is noted that for any positive definite matrix $\mathbf{B}$, scalar μ, and arbitrary vector $\mathbf{a}$, it follows from the Schur complement (see, e.g., [29]) that

$$\mu - \mathbf{a}^T \mathbf{B}^{-1} \mathbf{a} \geq 0 \iff \begin{bmatrix} \mathbf{B} & \mathbf{a} \\ \mathbf{a}^T & \mu \end{bmatrix} \succeq \mathbf{0}, \tag{8}$$

where $\mathbf{X} \succeq \mathbf{0}$ indicates that the matrix $\mathbf{X}$ is positive semidefinite. Thus, it follows that

$$\underset{\mathbf{B} \succ 0}{\text{minimize}} \ \mathbf{a}^T \mathbf{B}^{-1} \mathbf{a} \tag{9}$$

and

$$\begin{aligned} &\underset{\mu, \mathbf{B} \succ 0}{\text{minimize}} \quad \mu \\ &\text{subject to} \quad \begin{bmatrix} \mathbf{B} & \mathbf{a} \\ \mathbf{a}^T & \mu \end{bmatrix} \succeq \mathbf{0} \end{aligned} \tag{10}$$

are minimized by the same matrix $\mathbf{B}$. Here, $\mathbf{B} \succ \mathbf{0}$ indicates that the matrix $\mathbf{B}$ is positive definite. This observation allows us to reformulate (4) as the semidefinite program (SDP) (cf. [2, 18])

$$\begin{aligned} &\underset{\boldsymbol{\mu}, \mathbf{w}}{\text{minimize}} \quad \sum_{p=1}^{P} \psi_p \mu_p \\ &\text{subject to} \quad \begin{bmatrix} \sum_{n=1}^{N} w_n \mathbf{F}(\mathbf{t}_n; \boldsymbol{\theta}) & \mathbf{e}_p \\ \mathbf{e}_p^T & \mu_p \end{bmatrix} \succeq \mathbf{0}, \ p = 1, \ldots, P \\ &\qquad \sum_{n=1}^{N} w_n \mathbf{F}(\mathbf{t}_n; \boldsymbol{\theta}) \succ \mathbf{0} \\ &\qquad \mathbf{1}^T \mathbf{w} \leq \gamma \quad , \quad w_n \in [0, 1], \ n = 1, \ldots, N, \end{aligned} \tag{11}$$

where ψ_p are weight parameters allowing for putting emphasis on different components of the vector $\boldsymbol{\theta}$. For example, if $\psi_q = 1$ and $\psi_p = 0$, $\forall p \neq q$, then the CRLB for the parameter θ_q will be the only one minimized, as μ_q precisely corresponds to this lower bound, whereas the CRLBs for the other parameters θ_p, $p \neq q$ will be disregarded. Similarly, for $\psi_p = 1$, $\forall p$, the problems (4) and (11) are equivalent.

Another benefit of this formulation is that it allows for a straightforward way of incorporating performance constraints in the minimization problem, such as if, for instance, there is some upper tolerance bound λ_p for the CRLB of parameter θ_p. This kind of performance specifications can then be incorporated in the

minimization problem via linear inequality constraints according to

$$\begin{aligned}
\underset{\boldsymbol{\mu},\mathbf{w}}{\text{minimize}} \quad & \sum_{p=1}^{P} \psi_p \mu_p \\
\text{subject to} \quad & \begin{bmatrix} \sum_{n=1}^{N} w_n \mathbf{F}(\mathbf{t}_n;\boldsymbol{\theta}) & \mathbf{e}_p \\ \mathbf{e}_p^T & \mu_p \end{bmatrix} \succeq \mathbf{0}, \ p = 1,\ldots,P \\
& \sum_{n=1}^{N} w_n \mathbf{F}(\mathbf{t}_n;\boldsymbol{\theta}) \succ \mathbf{0} \\
& \mathbf{1}^T \mathbf{w} \leq \gamma \quad , \quad w_n \in [0,1], \ n = 1,\ldots,N \\
& \mu_p \leq \lambda_p \ , \ p = 1,\ldots,P
\end{aligned} \tag{12}$$

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 $\boldsymbol{\theta}$ instead lies in a set of possible parameters, Θ. In such cases, it may be desired to 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 further generalize (12) such that the sampling scheme is designed as

$$\begin{aligned}
\underset{\boldsymbol{\mu},\mathbf{w}}{\text{minimize}} \quad & \sum_{p=1}^{P} \psi_p \mu_p \\
\text{subject to} \quad & \begin{bmatrix} \sum_{n=1}^{N} w_n \mathbf{F}(\mathbf{t}_n;\boldsymbol{\theta}) & \mathbf{e}_p \\ \mathbf{e}_p^T & \mu_p \end{bmatrix} \succeq \mathbf{0}, \ \forall \boldsymbol{\theta} \in \Theta, \ p = 1,\ldots,P \\
& \sum_{n=1}^{N} w_n \mathbf{F}(\mathbf{t}_n;\boldsymbol{\theta}) \succ \mathbf{0} \\
& \mathbf{1}^T \mathbf{w} \leq \gamma \quad , \quad w_n \in [0,1], \ n = 1,\ldots,N \\
& \mu_p \leq \lambda_p \ , \ p = 1,\ldots,P.
\end{aligned} \tag{13}$$

Using this formulation, the optimal μ_p will, assuming that $\psi_p > 0$, now correspond to a worst case CRLB for the pth component of $\boldsymbol{\theta}$, when $\boldsymbol{\theta} \in \Theta$, i.e., for the obtained sampling sampling scheme

$$\mu_p = \arg\max_{\boldsymbol{\theta}\in\Theta} \mathbf{e}_p^T \mathcal{I}(\hat{\mathbf{w}};\boldsymbol{\theta})^{-1} \mathbf{e}_p \tag{14}$$

Thus, the solution to (13) is a sampling scheme minimizing the worst case CRLB for the parameters of interest if the parameter vector $\boldsymbol{\theta}$ is known to be in the set Θ.

Further, one could also consider the case where there is some cost associated with changing sampling points in one of the dimensions. For instance, if one of the sampling dimensions corresponds to a certain setting of a machine, e.g., time delay or magnetic flow, it could be more costly to acquire many different sample points in this dimension. Illustrating this in the 2-D case, one could include such a cost in the

optimization by forming the $N_1 \times N_2$ matrix $\mathbf{W}$ by reshaping the vector $\mathbf{w}$, and adding the constraints

$$\left|\left|\mathbf{W}^T\right|\right|_{2,1} = \sum_{n=1}^{N_1} \left|\left|\mathbf{W}_{(:,n)}\right|\right|_2 \leq \gamma_1 \tag{15}$$

$$||\mathbf{W}||_{2,1} = \sum_{n=1}^{N_2} \left|\left|\mathbf{W}_{(n,:)}\right|\right|_2 \leq \gamma_2 \tag{16}$$

to the optimization in (13). Here, γ_1 and γ_2 are tuning parameters that may be set according to the associated cost. This constraint can easily be omitted simply by setting $\gamma_1 = \gamma_2 = \infty$.

[Figure 1 about here.]

[Figure 2 about here.]

It is also worth noting that when relaxing (3) in favor for (4), we can no longer guarantee that the weights are exactly 0 or 1. In this case, as is noted in (5), we simple choose an appropriate threshold such that values above the threshold are deemed as ones, and the values below are deemed as zeros. However, a better approximation of (3) is found by using re-weighting. This may be done by first solving (13), yielding the estimated $\mathbf{w}^{(1)}$, where the superscript $(\cdot)^{(j)}$ denotes jth iteration. Then, (13) is solved again, but this time with

$$\frac{1}{w_n^{(1)} + \delta} \tag{17}$$

as a scaling factor for each w_n, where δ is a small number added to the denominator to avoid numerical problems. This procedure can then repeated until convergence. The re-weighting is a better approximation of the ℓ_0-norm, and thus is more likely to produce weights with values close to zero or one. As we have empirically found that using re-weighting for the here studied examples offers only a marginal improvement, while significantly increasing the computational cost due to the iterative procedure, we have in our examples chosen to use the simpler thresholding approach.

The SDP formulations in equations (11)-(13) may be minimized using, e.g., off-the-shelf solvers such as SeDuMi [30] or SDPT3 [31]. Typically, these require $\mathcal{O}(J^3)$ operations per iteration, where J is the number of variables. Clearly, this implies that the solution will be computationally demanding for large potential sampling grids and high-dimensional data. However, as noted in the introduction, in many experimental applications, the time required to carry out an experiment is often several orders of magnitude greater than the time needed for determining the optimal sampling scheme, even using standard solvers not exploiting any inherent structure in the problem. As a result, the scheme allows for a substantial overall decrease of the experimental time, even if the forming of the sampling scheme is somewhat cumbersome, as this step only needs to be done once for a given experimental setup, whereas the resulting experiment would typically be done multiple times, with each measurement typically taking substantially longer than the design of the sampling scheme.

3. Numerical results

3.1. Illustration in 1-D

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 initially focus on the 1-D case), such that

$$y(t_n) = \sum_{k=1}^{K} \alpha_k \exp\{2i\pi f_k t_n - \beta_k t_n + i\phi_k\} + \epsilon(t_n), \tag{18}$$

for $n = 1, \ldots, N$, where α_k, f_k, β_k, and ϕ_k are the amplitude, frequency, damping, and phase of the kth component, respectively, and where ϵ is an additive noise term, here assumed to be well modeled as a white circularly symmetric Gaussian noise[1] with variance σ^2, with t_n being the time at sample n. For simplicity, we consider uniformly sampled candidate sampling times, t_n. For the signal model in (18), the FIM corresponding to sampling time t_n is given by

$$\mathbf{F}(t_n; \boldsymbol{\theta}) = \frac{2}{\sigma^2} \begin{pmatrix} \mathbf{F}_{1,1}(t_n) & \mathbf{F}_{1,2}(t_n) & \ldots & \mathbf{F}_{1,K}(t_n) & 0 \\ \vdots & \vdots & \ldots & \vdots & \vdots \\ \mathbf{F}_{K,1}(t_n) & \mathbf{F}_{K,2}(t_n) & \ldots & \mathbf{F}_{K,K}(t_n) & 0 \\ 0 & 0 & \ldots & 0 & 1/\sigma^2 \end{pmatrix} \tag{19}$$

where

$$\mathbf{F}_{k,\ell}(t_n) = \begin{pmatrix} \frac{c_{k,\ell}(t_n)}{\alpha_k \alpha_\ell} & \frac{2\pi t_n s_{k,\ell}(t_n)}{\alpha_k} & \frac{-t_n c_{k,\ell}(t_n)}{\alpha_k} & \frac{s_{k,\ell}(t_n)}{\alpha_k} \\ \frac{-2\pi t_n s_{k,\ell}(t_n)}{\alpha_\ell} & (2\pi t_n)^2 c_{k,\ell}(t_n) & 2\pi t_n^2 s_{k,\ell}(t_n) & 2\pi t_n c_{k,\ell}(t_n) \\ \frac{-t_n c_{k,\ell}(t_n)}{\alpha_\ell} & -2\pi t_n^2 s_{k,\ell}(t_n) & t_n^2 c_{k,\ell}(t_n) & -t_n s_{k,\ell}(t_n) \\ \frac{-s_{k,\ell}(t_n)}{\alpha_\ell} & 2\pi t_n c_{k,\ell}(t_n) & t_n s_{k,\ell}(t_n) & c_{k,\ell}(t_n) \end{pmatrix} \tag{20}$$

with

$$c_{k,\ell}(t_n) = \alpha_k \alpha_\ell e^{-(\beta_k + \beta_\ell)t_n} \cos\left(2\pi(f_k - f_\ell)t_n + (\varphi_k - \varphi_\ell)\right) \tag{21}$$

$$s_{k,\ell}(t_n) = \alpha_k \alpha_\ell e^{-(\beta_k + \beta_\ell)t_n} \sin\left(2\pi(f_k - f_\ell)t_n + (\varphi_k - \varphi_\ell)\right). \tag{22}$$

Here, the parameter vector $\boldsymbol{\theta}$ is defined as

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1^T & \ldots & \boldsymbol{\theta}_K^T \end{bmatrix}^T \tag{23}$$

$$\boldsymbol{\theta}_k = \begin{bmatrix} \alpha_k & f_k & \beta_k & \varphi_k \end{bmatrix}^T. \tag{24}$$

As an illustration, Figure 1 shows an example of sampling schemes found by solving (11) 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

[1] The assumption of white circularly symmetric Gaussian noise is appropriate for the considered application to spectroscopy, where the noise is primarily resulting from thermal (Johnson) noise [32]. It is worth stressing that this is not a limitation for the sampling method as such, and the noise model may be selected to be non-Gaussian if deemed appropriate for the considered application; such a change would only affect the form of the used FIM.

bottom figure, but otherwise identical signal parameters. In both cases, $\gamma = 13$ so that $M = 13$ sample points, out of $N = 50$ possible candidates, are selected. Also, $\psi_p = 1$, $p = 1, \ldots, 4$, i.e., all signal parameters are considered in the minimization. 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 of the signal, where the 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.

As a further example, we next consider an example showing the resulting sample scheme for a signal containing two linear chirp components on the form

$$y(t_n) = \sum_{k=1}^{2} \alpha_k \exp\left\{2i\pi \left(f_k^0 + f_k^1 t_n\right) t_n + i\phi_k\right\} + \epsilon(t_n), \tag{25}$$

where f_k^0 and f_k^1 denote the frequency starting point and the slope of the chirp component k, respectively. Figure 2 shows the three sampling schemes yielded by the proposed method for three different setting on γ, namely $\gamma = 15$, $\gamma = 20$, and $\gamma = 25$. The here used parameters had the values $\alpha_1 = \alpha_2 = 5$, $f_1^0 = 0.1$, $f_2^0 = 0.5$, $f_1^1 = 0.01$, $f_2^1 = -0.003$, and the phases were set to $\phi_1 = \pi/2$, and $\phi_2 = \pi/3$. Due to the linear drift in frequency, it is reasonable to assume that the resulting sample scheme should have at least two clusters; one in the beginning of the signal, and one at the end of the signal. Looking at the sampling schemes in Figure 2 supports this intuition; three clusters are present for all three settings of γ. When γ increases the two first clusters gets bigger, whereas the last cluster remains more or less unchanged.

[Figure 3 about here.]

[Figure 4 about here.]

3.2. Illustration in 2-D

As further illustration of the impact of the choice of weight parameters ψ_p, consider the 2-D case with one damped sinusoid, i.e.,

$$y(t_1, t_2) = \alpha \mathrm{e}^{2i\pi(f_1 t_1 + f_2 t_2) - (\beta_1 t_1 + \beta_2 t_2) + i\phi} + \epsilon(t_1, t_2), \tag{26}$$

with $\alpha = 1$, $f_1 = 0.2$, $f_2 = 0.5$, $\beta_1 = 1/20$, $\beta_2 = 1/10$, $\phi = 1/2$, and noise variance $\sigma^2 = 0.1$. Figure 3 presents the sampling scheme found by solving (11) with $\gamma = 50$, i.e., 50 sampling points are chosen, for the case when $\psi_p = 1$ for all parameters.

As can be seen, the optimal sampling pattern here consists of three clusters of selected sampling points; one close to the origin and two close to the two time axes. Note that this is analogous to the 1-D case as the sampling cluster close to the first time axis is located further from the origin due to the decay in the first dimension being slower.

In contrast, Figure 4 displays the corresponding scheme found when solving (11), again with $\gamma = 50$, but only giving weight to the frequency and damping parameters, i.e., the ψ_p corresponding to the amplitude and phase parameters are set to zero. As can be seen, assigning the amplitude and phase parameters zero weight has the effect of shifting sampling points away from the origin to the clusters close to the t_1 and t_2 axes, in order to put more emphasis on the frequency and damping parameters. Indeed, the sum of the CRLBs for the parameters, as given by the sampling scheme in Figure 3, is $2.31 \cdot 10^{-2}$, whereas it is $3.61 \cdot 10^{-2}$ for the sampling scheme in Figure 4. However, if one considers the sum of the CRLBs for the frequency and damping parameters, these are $6.53 \cdot 10^{-4}$ and $4.42 \cdot 10^{-4}$ for Figures 3 and 4, respectively.

3.3. Simulations in 1-D

3.3.1. Optimization vs simulation

In Figure 5, we motivate that solving (13) is indeed a reasonable approach to determine optimal sampling patterns. The figure shows the obtained sum of the CRLBs for the parameters, i.e., $\mathrm{tr}\left(\mathcal{I}(\hat{\mathbf{w}};\boldsymbol{\theta})^{-1}\right)$, where the sampling pattern is obtained by solving (13) for the case of $K = 1$ using the model (18), for a singleton set Θ. This is done for varying values of γ 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 one of these 10^6 sampling patterns that results in the lowest sum of CRLBs. This will generally result in a sampling scheme significantly preferable to any fixed sampling scheme, such as Poisson gap sampling. 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 exact 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 (13) is then able to achieve an optimal performance as the sample size increases.

[Figure 5 about here.]

3.3.2. Weighting

In Figures 6 and 7, we proceed to examine the effect of using different weightings for the signal parameters when solving (11). This is 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 weight parameters ψ_p are set to one for the frequency and damping parameters, whereas they are set to zero for the amplitudes and phases. 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 6 and 7, 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, i.e., $\psi_p = 1$ for all signal parameters. As can be seen, the weighting scheme results in sampling patterns that decrease 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 NLS estimate is found by solving

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \quad \frac{1}{2}||\mathbf{y} - g(\boldsymbol{\theta})||_2^2 \tag{27}$$

where $\mathbf{y}$ is the data and $g(\boldsymbol{\theta})$ is the (non-linear) data model with parameter $\boldsymbol{\theta}$. In this paper, a minimum of (27) is found by evaluating the cost function over a grid of parameter values $\boldsymbol{\theta}$. The $\boldsymbol{\theta}$ that achieves the lowest value of (27) then becomes the resulting estimate. 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.

[Figure 6 about here.]

[Figure 7 about here.]

3.3.3. Gridding

Figures 8 and 9 show the effect of finding an optimal sampling pattern for a set of parameters $\boldsymbol{\theta} \in \Theta$ when solving (13). The results are obtained for a single decaying sinusoid. Here, we let $\Theta = \{\boldsymbol{\theta}_\ell\}_{\ell=1}^{L}$ 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 $\boldsymbol{\theta}_\ell = (\alpha, f, \beta_\ell, \varphi)^T$, where

$$\beta_\ell = \beta_{\text{lower}} + \frac{\ell - 1}{L}\Delta_\beta \tag{28}$$

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

$$\mathcal{J}_\beta = \left[\beta_{\text{lower}}, \beta_{\text{lower}} + \frac{L-1}{L}\Delta_\beta\right]. \tag{29}$$

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 (13) 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 $\boldsymbol{\theta}$ where β is sampled uniformly on $\mathcal{J}_\beta$, i.e., on the interval covered by the grid Θ, but not on the grid points β_ℓ, $\ell = 0, 1, \ldots, L-1$. We then estimate $\boldsymbol{\theta}$ using NLS and compute the RMSE for the parameters $\boldsymbol{\theta}$. The figures show the obtained RMSE 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 (13) indeed yields sampling schemes with a guaranteed worst case performance, as well as a lower limit on the possible RMSE.

[Figure 8 about here.]

[Figure 9 about here.]

3.4. Simulations in 2-D

3.4.1. Optimization vs simulation

As was seen in the 1-D setting, the optimization scheme was able to outperform the method of randomly selecting sampling points and then choosing the scheme minimizing the sum of the CRLB. In 2-D, this becomes even more apparent as the number of potential sampling points increase rapidly with increasing dimension. An illustration of this is shown in Figure 10, showing the sum of the CRLBs obtained when solving (13) for varying numbers of desired sampling points. The signal considered is the 2-D damped sinusoid in (26) with parameters $\alpha = 1$, $f_1 = 0.2$, $f_2 = 0.5$, $\beta_1 = 1/20$, $\beta_2 = 1/10$, $\phi = 1/2$, and $\sigma^2 = 0.1$. We here let $\psi_p = 1$ for all signal parameters, and consider a sampling space of 50×50 potential sampling times. Also presented is the sum of the CRLBs for the best (defined as the one with smallest sum of CRLBs) among 10^7 sampling scheme obtained by randomly choosing sampling points. As can be seen from the figure, the proposed method outperforms the random sampling for all numbers of selected samples. It is worth noting that the computational time to evaluate the 10^7 sampling schemes was three times longer than solving the proposed problem using a off-the-shelf convex solver [33].

[Figure 10 about here.]

3.4.2. Weighting

We here consider the case of a signal consisting of two 2-D damped sinusoid, i.e.,

$$y(t_1, t_2) = \sum_{k=1}^{K} \alpha_k \mathrm{e}^{i\phi_k} \Pi_{d=1}^{2} \mathrm{e}^{2i\pi f_{k,d} t_d - \beta_{k,d} t_d} + \epsilon(t_1, t_2), \tag{30}$$

for $K = 2$. Let the parameters be $(f_{1,1}, f_{2,1}) = (0.1, 0.2)$ and $(\beta_{1,1}, \beta_{2,1}) = (0.1, 0.1)$ for the first dimension, $(f_{1,1}, f_{2,1}) = (0.1, 0.2)$ and $(\beta_{1,1}, \beta_{2,1}) = (0.1, 0.1)$ for the second dimension, and let $\alpha_1 = 1$, $\alpha_2 = 1.3$, $\phi_1 = \frac{\pi}{3}$, $\phi_2 = \frac{\pi}{3}$, and $\sigma^2 = 0.01$. We then determine optimal sampling schemes by solving (11) for varying numbers of sampling points. This is done for both the equally weighted case, i.e., with $\psi_p = 1$ for all signal parameters, as well as for the case when only the frequency and damping parameters are given weight, i.e., with $\psi_p = 0$ for the amplitude and phase parameters. The results are shown in Figures 11-14. In Figure 11, the root of the sum of the CRLBs for the frequencies in the first dimension, i.e., $f_{1,1}$ and $f_{2,1}$, is shown. Similarly, Figure 12 corresponds to the frequencies in the second dimension, while Figures 13 and 14 correspond to the damping parameters in the first and second dimension, respectively. Also presented is the corresponding RMSE obtained when estimating the parameters using NLS. As can be seen, the obtained RMSEs coincides with the CRLBs for both the weighted and non-weighted case, implying that the bound is tight. Note also that the schemes corresponding to assigning no weight to the amplitude and phase parameters all result in a lower sum of CRLB for the frequency and damping parameters than the non-weighted schemes. This comes

at the price of a larger sum of CRLB for the amplitudes α_1 and α_2, which is illustrated in Figure 15. As can be seen in the figure, the non-weighted sampling scheme here leads to more accurate estimates of the amplitudes.

[Figure 11 about here.]

[Figure 12 about here.]

[Figure 13 about here.]

[Figure 14 about here.]

[Figure 15 about here.]

3.5. Real NMR signal

In order to illustrate the proposed framework's applicability to real, measured data, we first consider a 2-D NMR signal obtained from a ^{15}N-HSQC experiment made on a Histidine sample. The signal is sampled on a uniform 25×25 grid. In this form of experiments, the spectrometer measures the signal resulting from repeatedly pulsing the studied substance, each pulse resulting in a so-called free-induction decay (FID). This measurement is made in the direct dimension, i.e., along time. By modifying the pulse sequence, for instance changing the timing between pulses, one obtain measurements also in a second (indirect) dimension. By further modifications to the pulse sequence, one may similarly obtain higher dimensional data sets, where each dimension corresponds to the range over each of the considered pulse settings [34]. Figure 16 shows the 2-D periodogram of the signal using all available samples. As can be seen, the signal consists of a number of components of varying powers, all which may be fairly well modeled as damped complex sinusoids. Superimposed on the periodogram are estimates obtained using the SEMA method presented in [35], also using all available samples. This method is a sparse reconstruction algorithm requiring no prior knowledge of the number of components constituting the signal. Considering only the four most prominent components estimated by SEMA, we then compute a sampling scheme consisting of 225 samples in total, i.e., a mere 36% of the number of available samples, using the formulation in (11), and estimate the number of signal components, as well as the parameters of the components, using SEMA. The result is presented in Figure 17, displaying the periodogram using all samples superimposed with the obtained SEMA estimates. Comparing Figures 16 and 17, it may be noted that the considerably reduced sample size does not cause any significant degradation in estimation performance.

Proceeding, we examine the generation of a 3-D sampling scheme for a laser spectroscopy measurement containing 4 damped sinusoids. The measured data set contains $40 \times 40 \times 20$ samples, out of which we select a total of 80 samples using (11). Figure 18 shows the resulting sampling scheme. The figure again illustrates, reminiscent to Figures 1 and 3, that samples are selected where the local SNR is highest, to allow for accurate frequency estimation, as well as a smaller number of additional samples selected later in the sequences to better capture the damping behavior.

[Figure 16 about here.]

[Figure 17 about here.]

[Figure 18 about here.]

4. Conclusion

In this work, we have proposed a convex optimization problem for finding suitable sampling schemes for multidimensional data models. The optimization problem is formed as to, from a set of available samples, determine a subset of given cardinality such that linear combinations of the variances of the signal parameters of interest are minimized. Due to the structure of the optimization problem, it is easy to add additional constraints, e.g., adding performance bounds on selected parameters, or putting more emphasize on a subset of the parameters, and to model for the uncertainty in *a priori* assumptions of the parameter values. In the numerical section, we show that solving the proposed optimization problem is a more efficient approach than randomly selecting the sampling points, especially in the multi-dimensional setting. Further, we show that using the sampling schemes found by solving the proposed optimization problem, will provide a lower Cramér-Rao lower bound than that found from using ordinary uniform sampling. By using an efficient parameter estimator on the signal sampled according to the found sampling scheme, we show that these Cramér-Rao lower bounds are, in fact, tight.

5. Acknowledgement

The authors wish to express their gratitude to Prof. Mikael Akke, Prof. Tönu Pullerits, Dr Göran Carlström, and Dr Khadga Jung Karki, all at Lund University, for providing us with the here examined 2-D NMR and 3-D laser spectroscopy measurements.

[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] S. Ravishankar and Y. Bresler, "Adaptive Sampling Design for Compressed Sensing MRI," in *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, Boston, Massachusetts, 30 Aug.-3 Sept 2011, pp. 3751–3755.

[5] F. Gama, A. G. Marques, G. Mateos, and A. Ribeiro, "Rethinking Sketching as Sampling: Linear Transformation of Graph Signals," in *50th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 2016.

[6] A. Anis, A. Gadde, and A. Ortega, "Efficient Sampling Set Selection for Bandlimited Graph Signals Using Graph Spectral Proxies," *IEEE Trans. Signal Process.*, vol. 64, no. 14, pp. 3775–3789, July 2016.

[7] 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.

[8] K. Kazimierczuk, A Zawadzka-Kazimierczuk, and W. Koźmiński, "Non-uniform frequency domain for uniform exploitation of non-uniform sampling," *J. Magn. Reson.*, vol. 205, pp. 286–292, 2010.

[9] 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.

[10] 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.

[11] K. Kazimierczuk and V. Y. Orekhov, "Accelerated NMR Spectroscopy by Using Compressed Sensing," *Angewandte Chemie International Edition*, vol. 50, no. 24, June 2011.

[12] 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.

[13] M. Billeter, "Non-uniform sampling in biomolecular nmr," *Journal of Biomolecular NMR*, vol. 68, no. 2, pp. 65–55, 2017.

[14] S. G. Hyberts, H. Arthanari, S. A. Robson, and G. Wagner, "Perspectives in magnetic resonance: NMR in the post-FFT era," *J. of Magn. Reson.*, vol. 241, pp. 60–73, 2014.

[15] P. C. Aoto, R. B. Fenwick, G. J. A. Kroon, and P. E. Wright, "Accurate Scoring of Non-uniform Sampling Schemes for Quantitative NMR," *Journal of Magnetic Resonance*, vol. 246, pp. 31–35, Sept 2014.

[16] A. Özcelikkale, H. M. Ozaktas, and E. Arikan, "Signal Recovery with Cost-Constrained Measurements," *IEEE Trans. Signal Process.*, vol. 58, no. 7, pp. 3607–3617, July 2010.

[17] Y. Yilmaz, S. Li, and X. Wang, "Sequential Joint Detection and Estimation: Optimum Tests and Applications," *IEEE Trans. Signal Process.*, vol. 64, no. 20, pp. 5311, Oct 2016.

[18] V. Kekatos, G. B. Giannakis, and B. Wollenberg, "Optimal Placement of Phasor Measurement Units via Convex Relaxation," *IEEE Trans. on Power Systems*, vol. 27, no. 3, pp. 1521–1530, Aug 2012.

[19] T. C-Gulcu and H. M. Ozaktas, "Choice of Sampling Interval and Extent for Finite-Energy Fields," *IEEE Trans. Signal Process.*, vol. 65, no. 7, pp. 1741–1751, April 2017.

[20] H. Zhang, J. M. F. Moura, and B. K. Krogh, "Dynamic Field Estimation Using Wireless Sensor Networks: Tradeoffs Between Estimation Error and Communication Cost," *IEEE Trans. Signal Process.*, vol. 57, no. 6, pp. 2383–2395, June 2009.

[21] S. Liu, E. Masazade, and P. K. Varshney, "Temporally Staggered Sensing for Field Estimation with Quantized Data in Wireless Sensor Networks," in *IEEE Statistical Signal Processing Workshop (SSP)*, Ann Arbor, MI, USA, August 2012.

[22] S. Joshi and S. Boyd, "Sensor Selection via Convex Optimization," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 451–462, February 2009.

[23] F. Pukelsheim, *Optimal design of experiments*, Wiley series in probability and mathematical statistics. Wiley, New York, 1993.

[24] S. P. Chepuri, *Sparse Sensing for Statistical Inference - Theory, Algorithms, and Applications*, Ph.D. thesis, Delft University of Technology, 2015.

[25] 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.

[26] S. Liu, S. P. Chepuri, M. Fardad, E. Masazade, and G. Leus P. K. Varshney, "Sensor Selection for Estimation with Correlated Measurement Noise," *IEEE Trans. Signal Process.*, vol. 64, no. 13, pp. 3509–3522, July 2016.

[27] S. P. Chepuri and G. Leus, "Continuous Sensor Placement," *IEEE Signal Process. L*, vol. 22, no. 5, pp. 544–548, May 2015.

[28] J. Swärd, F. Elvander, and A. Jakobsson, "Designing Optimal Sampling Schemes," in *25th European Signal Processing Conference*, Aug 28 - Sep 2 2017.

[29] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[30] J. F. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," *Optimization Methods and Software*, vol. 11-12, pp. 625–653, August 1999.

[31] R. H. Tutuncu, K. C. Toh, and M. J. Todd, "Solving semidefinite-quadratic-linear programs using SDPT3," *Mathematical Programming Ser. B*, vol. 95, pp. 189–217, 2003.

[32] J. B. Johnson, "Thermal Agitation of Electricity in Conductors," *Phy. Rev.*, vol. 32, pp. 97–109, July 1928.

[33] Inc. CVX Research, "CVX: Matlab Software for Disciplined Convex Programming, version 2.0 beta," `http://cvxr.com/cvx`, Sept. 2012.

[34] J. Keeler, *Understanding NMR Spectroscopy*, Chichester, U.K. : John Wiley and Sons, 2010.

[35] J. Swärd, S. I. Adalbjörnsson, and A. Jakobsson, "High Resolution Sparse Estimation of Exponentially Decaying N-dimensional Signals," *Elsevier Signal Processing*, vol. 128, pp. 309–317, Nov 2016.

List of Figures

1 The resulting sampling 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 $\beta = \frac{1}{20}$. 19
2 The resulting sample scheming for three different settings of γ, namely $\gamma = 15$, $\gamma = 20$, and $\gamma = 25$, where the signal contains two linear chirps. 20
3 The resulting sampling scheme consisting of 50 selected samples for a signal consisting of a 2-D damped sinusoid as found when solving (11) with all $\psi_p = 1$. 21
4 The resulting sampling scheme consisting of 50 selected samples for a signal consisting of a 2-D damped sinusoid as found when solving (11) with all $\psi_p = 1$ except for the amplitude and phase parameters, for which $\psi_p = 0$. 22
5 Sum of CRLBs for the parameters, i.e., tr $\left(\mathcal{I}(\hat{\mathbf{w}}; \boldsymbol{\theta})^{-1}\right)$, for the sampling patterns given by the optimization problem and the best simulation, respectively, for different number of sampling points. 23
6 Obtained RMSE for the frequencies, when using the sampling patterns for the weighted and non-weighted cases, respectively. 24
7 Obtained RMSE for the damping, when using the sampling patterns for the weighted and non-weighted cases, respectively. 25
8 Obtained RMSE for the frequency f, when estimating $\boldsymbol{\theta}$ for the sampling pattern obtained for a grid of damping parameters β. 26
9 Obtained RMSE for the damping β, when estimating $\boldsymbol{\theta}$ for the sampling pattern obtained for a grid of damping parameters β. 27
10 The sum of variances of the parameters of interest as a function of the number of selected samples. 28
11 Obtained RMSE for the frequencies in the first dimension, when using the sampling patterns for the weighted and non-weighted cases, respectively. 29
12 Obtained RMSE for the frequencies in the second dimension, when using the sampling patterns for the weighted and non-weighted cases, respectively. 30
13 Obtained RMSE for the dampings in the first dimension, when using the sampling patterns for the weighted and non-weighted cases, respectively. 31
14 Obtained RMSE for the dampings in the second dimension, when using the sampling patterns for the weighted and non-weighted cases, respectively. 32
15 Obtained RMSE for the amplitudes, when using the sampling patterns for the weighted and non-weighted cases, respectively. 33
16 Estimates obtained using SEMA when applied to a real NMR signal using all available samples, superimposed on the corresponding periodogram estimate. 34
17 Estimates obtained using SEMA when applied to a real NMR signal 36% of the available samples as selected by (11), superimposed on the periodogram estimate using all available samples. 35
18 The resulting sampling scheme for a 3-D laser spectroscopic signal containing 4 damped sinusoids. 36

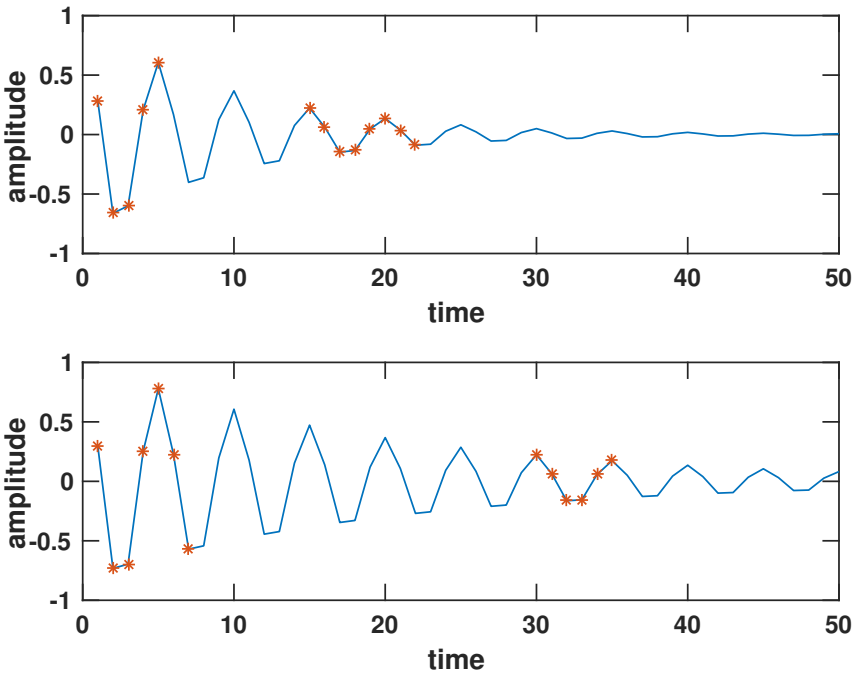

Figure 1: The resulting sampling scheme for two different values of β plotted against the real part of the signal. The upper most figure details the sampling scheme for $\beta = \frac{1}{10}$ and the bottom figure the sampling scheme for $\beta = \frac{1}{20}$.

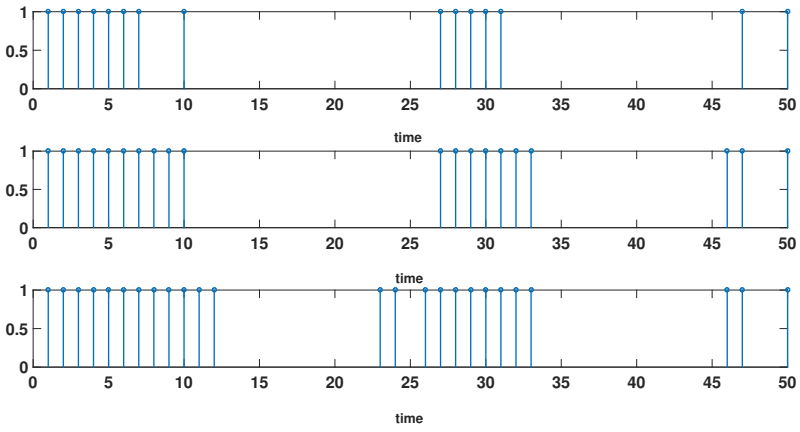

Figure 2: The resulting sample scheming for three different settings of γ, namely $\gamma = 15$, $\gamma = 20$, and $\gamma = 25$, where the signal contains two linear chirps.

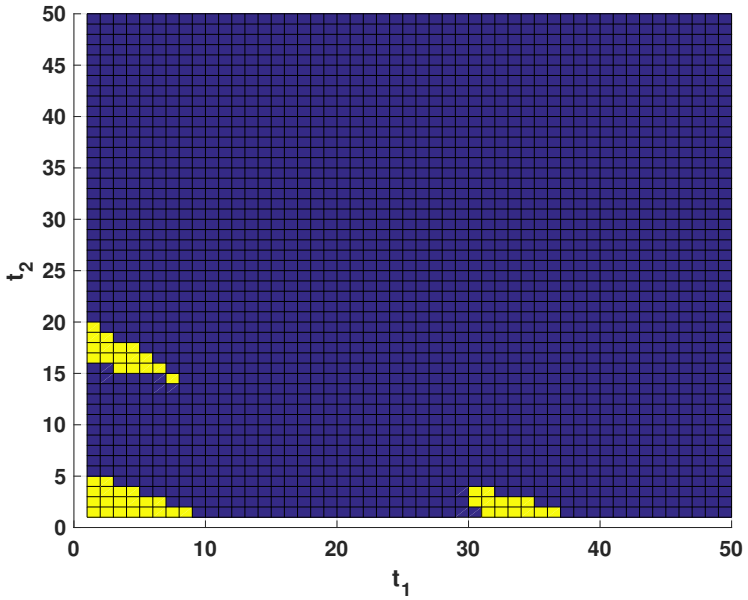

Figure 3: The resulting sampling scheme consisting of 50 selected samples for a signal consisting of a 2-D damped sinusoid as found when solving (11) with all $\psi_p = 1$.

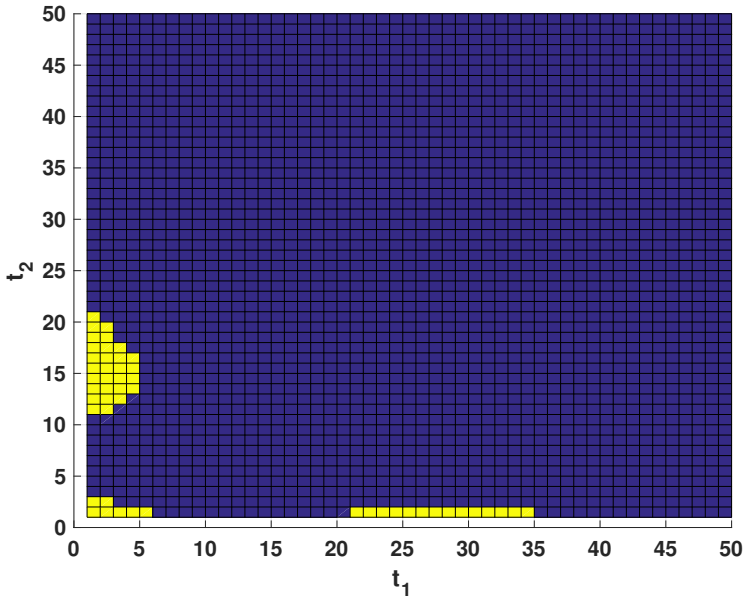

Figure 4: The resulting sampling scheme consisting of 50 selected samples for a signal consisting of a 2-D damped sinusoid as found when solving (11) with all $\psi_p = 1$ except for the amplitude and phase parameters, for which $\psi_p = 0$.

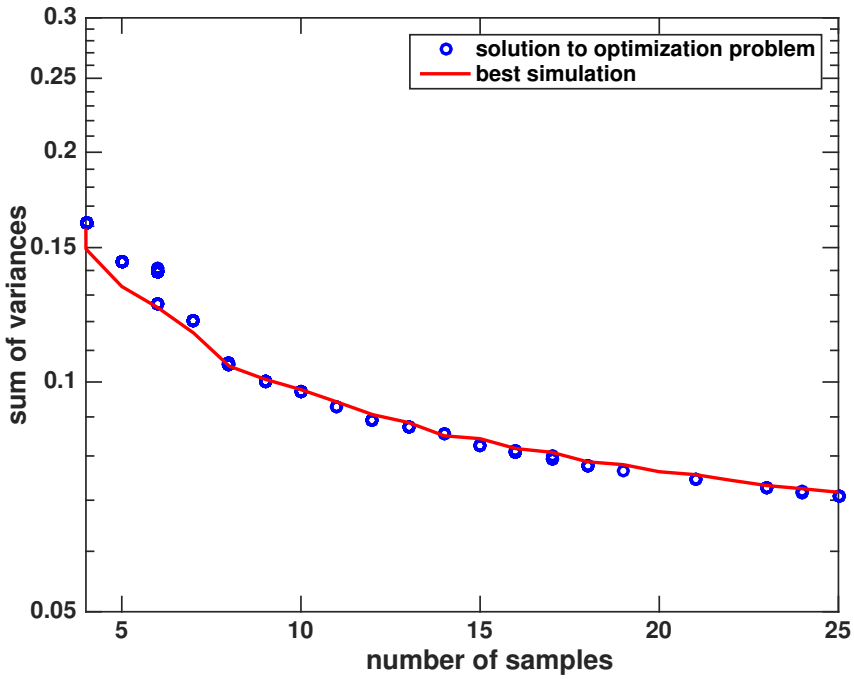

Figure 5: Sum of CRLBs for the parameters, i.e., $\mathrm{tr}\left(\mathcal{I}(\hat{\mathbf{w}};\boldsymbol{\theta})^{-1}\right)$, for the sampling patterns given by the optimization problem and the best simulation, respectively, for different number of sampling points.

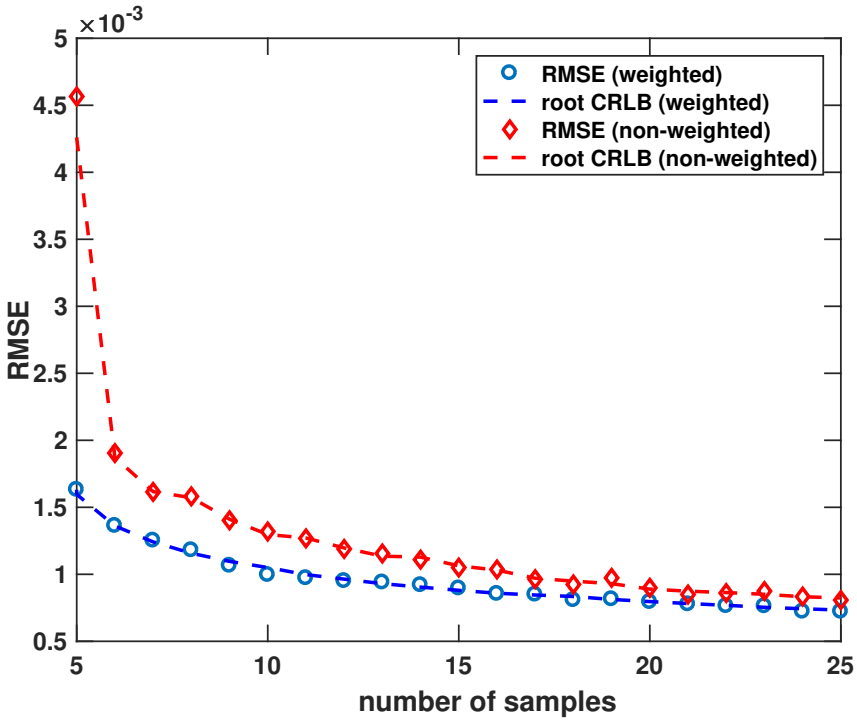

Figure 6: Obtained RMSE for the frequencies, when using the sampling patterns for the weighted and non-weighted cases, respectively.

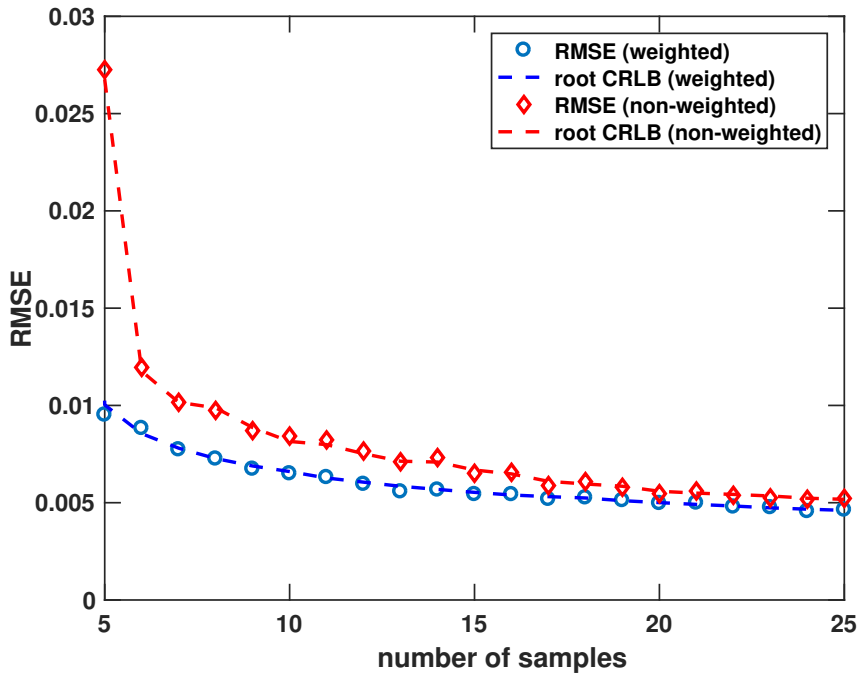

Figure 7: Obtained RMSE for the damping, when using the sampling patterns for the weighted and non-weighted cases, respectively.

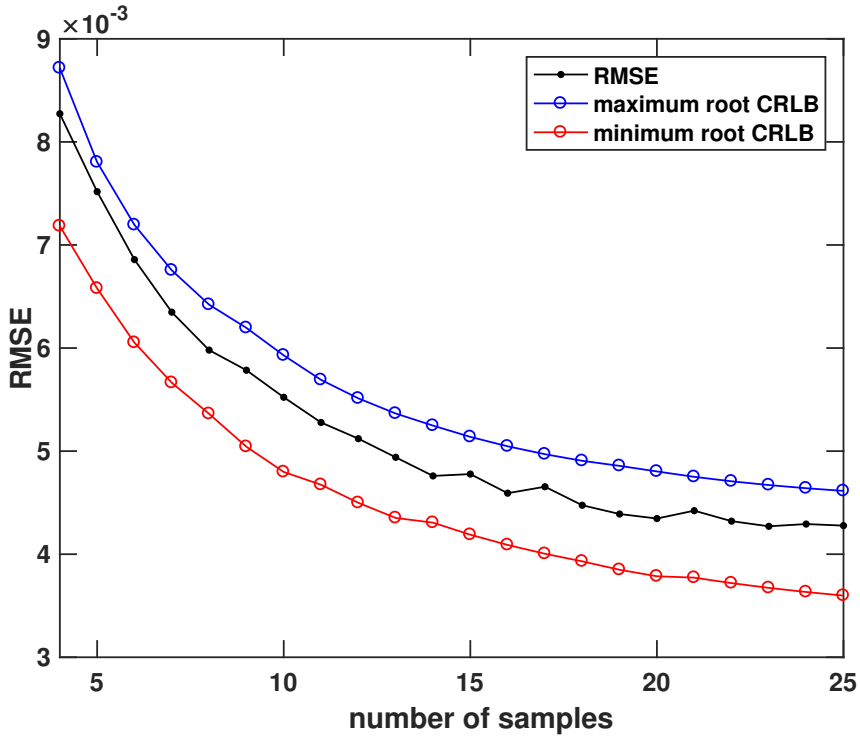

Figure 8: Obtained RMSE for the frequency f, when estimating $\boldsymbol{\theta}$ for the sampling pattern obtained for a grid of damping parameters β.

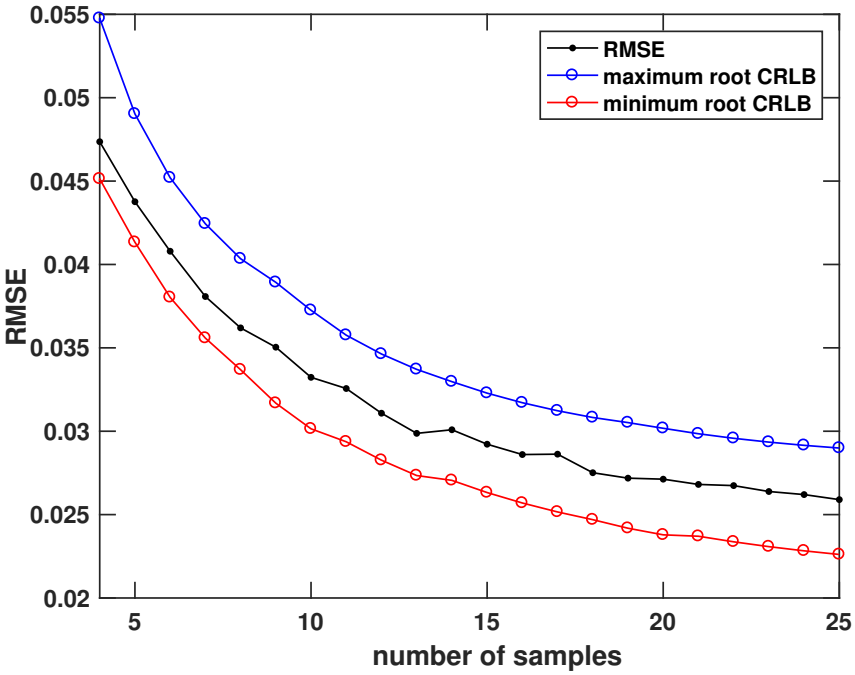

Figure 9: Obtained RMSE for the damping β, when estimating $\boldsymbol{\theta}$ for the sampling pattern obtained for a grid of damping parameters β.

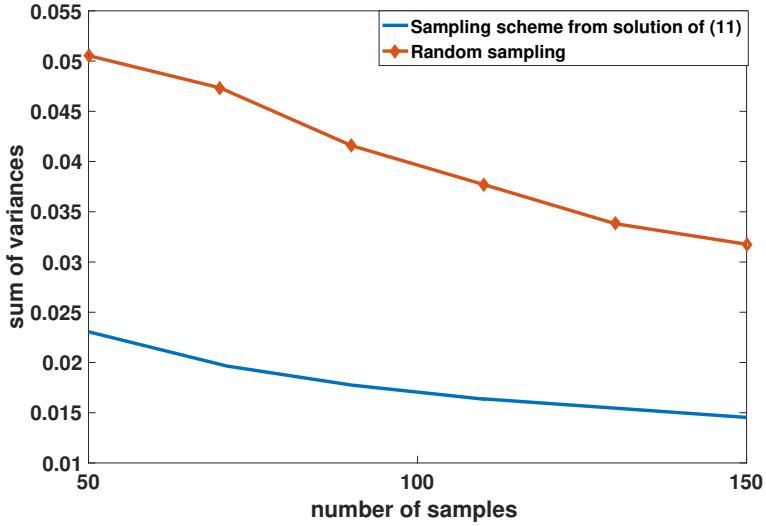

Figure 10: The sum of variances of the parameters of interest as a function of the number of selected samples.

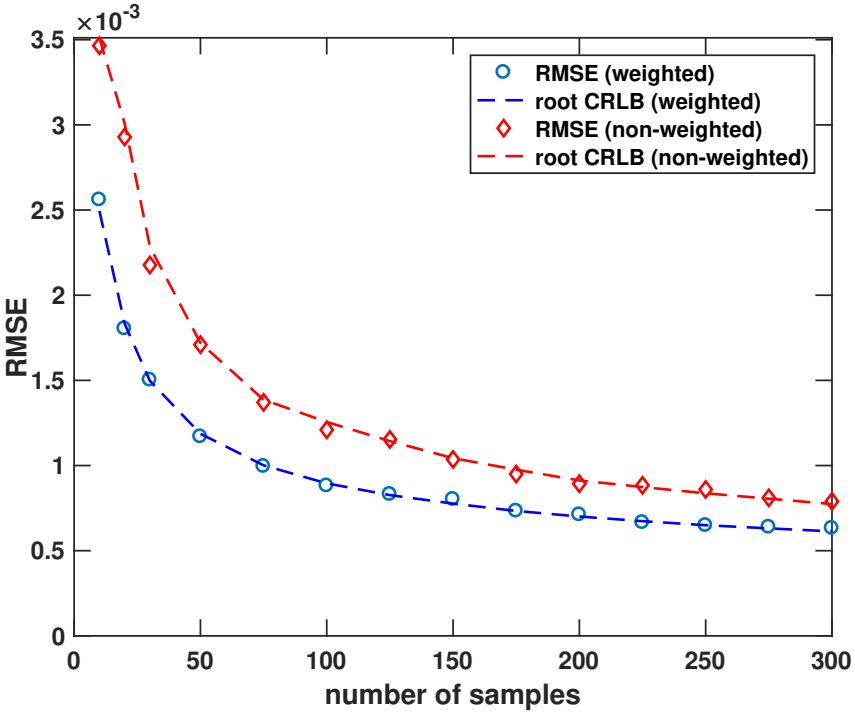

Figure 11: Obtained RMSE for the frequencies in the first dimension, when using the sampling patterns for the weighted and non-weighted cases, respectively.

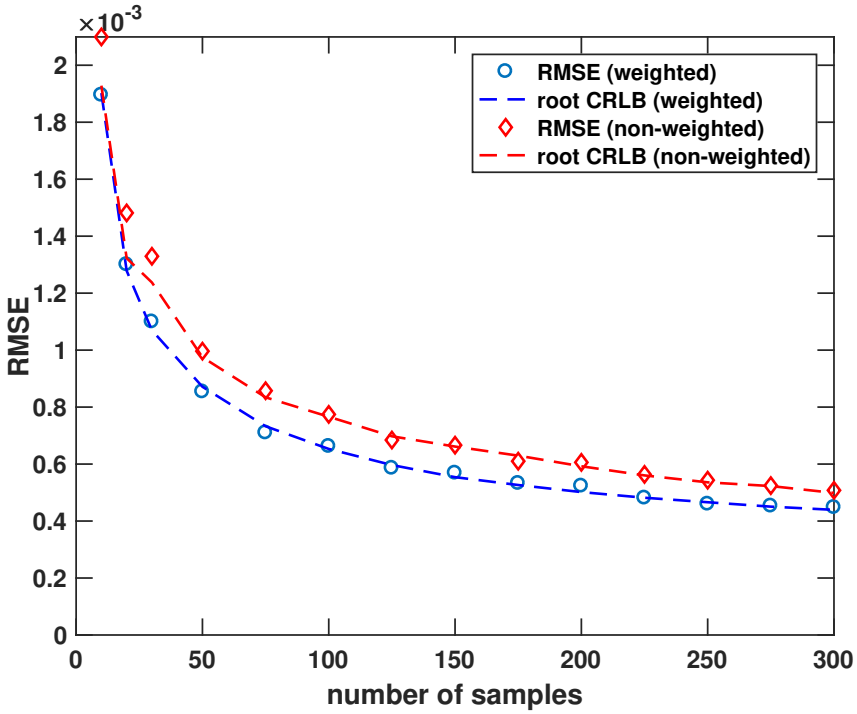

Figure 12: Obtained RMSE for the frequencies in the second dimension, when using the sampling patterns for the weighted and non-weighted cases, respectively.

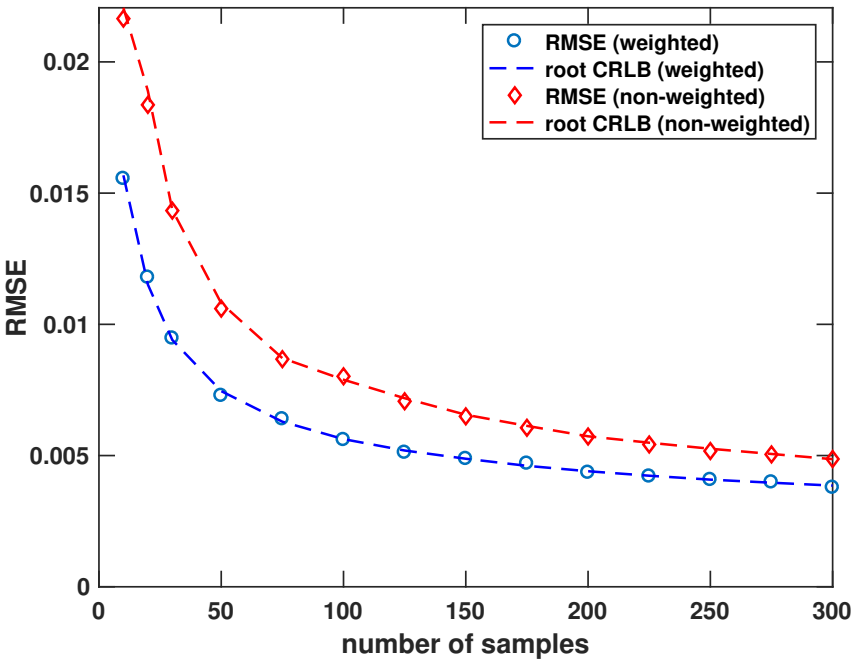

Figure 13: Obtained RMSE for the dampings in the first dimension, when using the sampling patterns for the weighted and non-weighted cases, respectively.

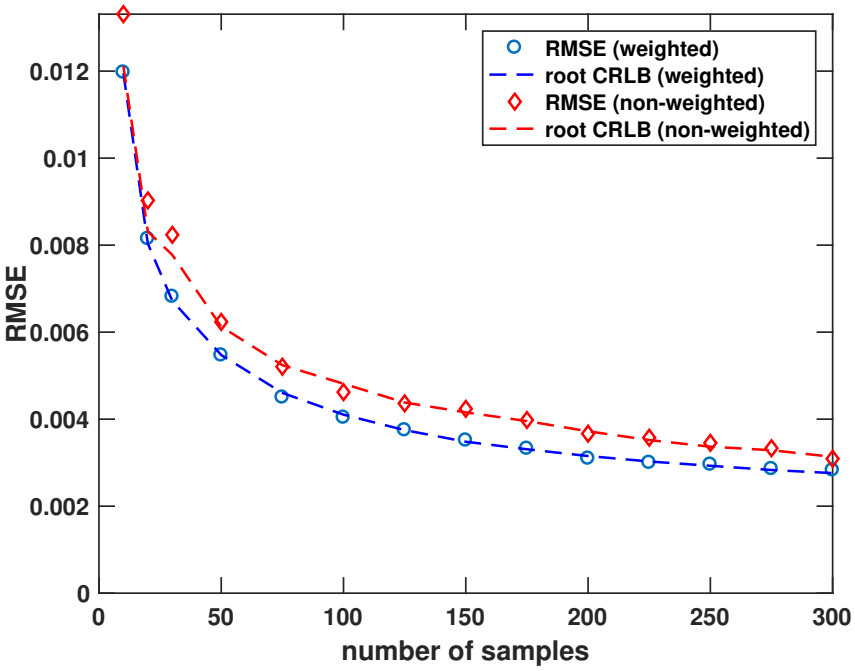

Figure 14: Obtained RMSE for the dampings in the second dimension, when using the sampling patterns for the weighted and non-weighted cases, respectively.

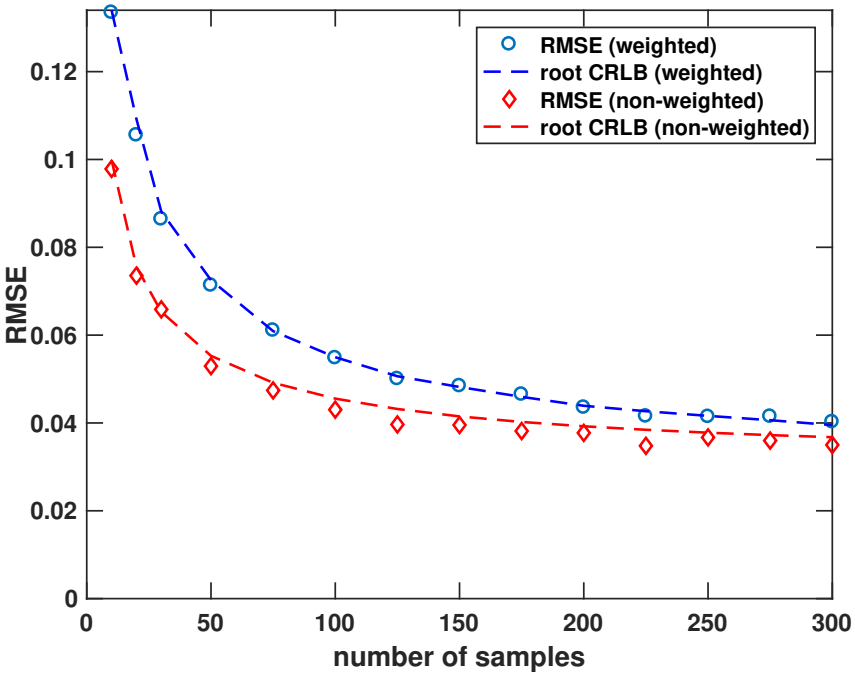

Figure 15: Obtained RMSE for the amplitudes, when using the sampling patterns for the weighted and non-weighted cases, respectively.

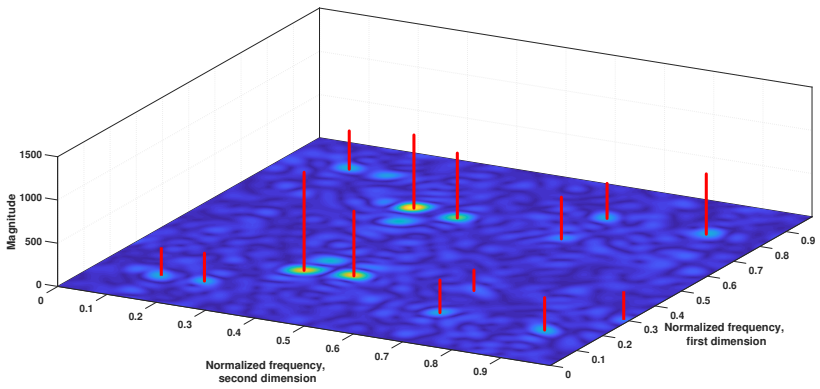

Figure 16: Estimates obtained using SEMA when applied to a real NMR signal using all available samples, superimposed on the corresponding periodogram estimate.

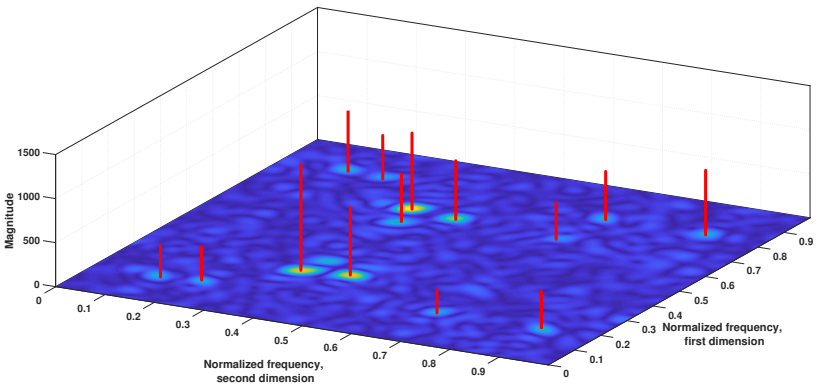

Figure 17: Estimates obtained using SEMA when applied to a real NMR signal 36% of the available samples as selected by (11), superimposed on the periodogram estimate using all available samples.

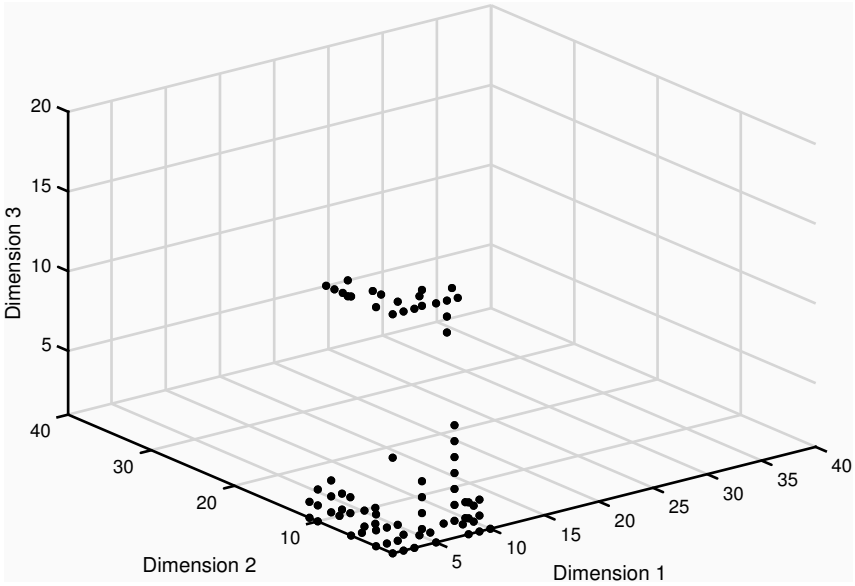

Figure 18: The resulting sampling scheme for a 3-D laser spectroscopic signal containing 4 damped sinusoids.