

A preliminary study on Optimal Input Design for Nonlinear Systems

Alexander De Cock, Michel Gevers and Johan Schoukens

Abstract—In this paper, an initial study is made on the optimal input design for Wiener systems that consists of a FIR filter, followed by a polynomial nonlinearity. A design method, based on the dispersion function, is introduced, in order to find an optimal set of elementary designs. For the considered class of Wiener systems, it is shown that these elementary designs are couples of successive input values. However, concatenation of these elementary designs is not straight forward. By imposing symmetry conditions on the total design, a solution is obtained that can be realized as a time sequence and that is optimal in the subspace of symmetric designs.

I. INTRODUCTION

Optimal input design for linear dynamic systems or static nonlinear systems is well understood and well covered in the literature [1], [2]. However, for nonlinear dynamic systems, no general solution to the problem is known. Therefore, different attempts have been made to extend the scope of the linear methods to subclasses of nonlinear systems: see e.g. [3]. In this paper a similar approach is followed for the extension of the dispersion based input design for linear dynamic systems [4].

The class of systems will be restricted to Wiener systems consisting of a FIR filter followed by a polynomial nonlinearity (Fig.1). Such systems are already covered in [5], where the input design is formulated as a convex optimization problem, and [6], where the optimal input in the class of random Gaussian signals was obtained.

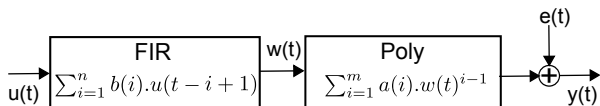


Fig. 1. FIR filter followed by a polynomial nonlinearity

While the problem statement and solution of this paper show strong similarities with [5], important differences exist. First of all, the problem formulation in this paper is obtained through generalization of a linear design method. This stands in contrast with the approach followed in [5] where the problem statement is obtained by restricting the general nonlinear case. Second, an optimization scheme based on the dispersion function is presented. This scheme offers an easy

This work was supported in part by the Fund for Scientific Research (FWO-Vlaanderen), by the Flemish Government (Methusalem), by the Belgian Government through the Inter university Poles of Attraction (IAP VII) Program, and the ERC Advanced Grant SNLSID.

Michel Gevers is with ICTEAM, Louvain University, B1348 Louvain la Neuve, Belgium, email: Michel.Gevers@uclouvain.be

Alexander De Cock and Johan Schoukens are with the Dept. of ELEC, Vrije Universiteit Brussel, email: adecock@vub.ac.be

to interpret alternative for the convex optimization routines suggested in [5].

To keep the explanation as simple as possible, all concepts in this paper will be illustrated on a simple example consisting of a 2-tap FIR filter followed by a cubic nonlinearity. Extending the method to systems with a longer FIR filter and/or a higher order polynomial can be done without conceptual problems. However numerical difficulties may occur due to the exponential scaling of the search space with respect to the filter length.

The remainder of this paper is structured as follows. Section II introduces basic concepts of dispersion based input design for linear dynamic systems. Section III presents how this concept can be extended to the considered class of nonlinear dynamic systems. In Section IV the extended method is illustrated on a numerical example. To conclude, Section V will summarize the obtained results.

II. INPUT DESIGN FOR LINEAR DYNAMIC SYSTEMS

This section describes the basic concepts of frequency domain input design for linear dynamic systems using the dispersion function. Details can be found in [2] and [4].

A. Problem statement

The simple setting of a SISO system, subjected to i.i.d Gaussian output noise, will be assumed.

Assumption 1. In this section the considered system is a linear dynamic SISO (Single Input Single Output) system. The relationship between the noiseless input $u_0(t)$ and the noiseless output $y_0(t)$ of the system can be written as:

$$\begin{aligned} y_0(t) &= g(t, \theta) * u_0(t) \\ Y_0(\omega) &= G(\omega, \theta) \cdot U_0(\omega) \end{aligned}$$

where U_0 and Y_0 are, respectively, the complex spectra of the input and output, θ are the parameters of the model, g is the impulse response of the system, G is the transfer function of the system, and $*$ denotes the convolution.

Assumption 2. There is only additive noise at the output, which is identically independently and normally distributed with known standard deviation σ . It is also independent of the input signal.

$$\begin{aligned} u(t) &= u_0(t) \\ y(t) &= y_0(t) + e(t) \\ e(t) &\sim N(0, \sigma^2) \end{aligned}$$

Assumption 3. The estimator used to identify the model parameters is unbiased and efficient.

Assumption 4. The class of input signals is restricted to inputs which have a discrete spectral representation and a total power of one.

$$\xi_u(k) = |U(\omega_k)|^2 \quad \text{and} \quad \sum_{k=1}^{N_\omega} \xi_u(k) = 1$$

where ξ_u represents the discrete power spectrum of the input u , and ω_k are the discrete frequencies.

Assumption 5. The D-optimality measure is used, which means that the optimal input sequence u_{opt} corresponds to the sequence for which the determinant of the information matrix Fi is maximal.

$$u_{opt} = \arg \max_u (\det(Fi))$$

Given the noise assumption, Fi can be computed based on the frequency domain data as [7]:

$$Fi = \frac{2}{\sigma^2} \cdot \text{Re} \left\{ \left(\frac{\partial Y_0}{\partial \theta} \right)^H \left(\frac{\partial Y_0}{\partial \theta} \right) \right\}$$

where $\frac{\partial Y_0}{\partial \theta}$ is a $N \times n_\theta$ matrix containing the partial derivatives of Y_0 , H stands for the Hermitian conjugate of a matrix and Re indicates the real part of a complex number.

B. Elementary design

For a linear dynamic system the output at frequency ω only depends on the input at that same frequency. If the input signal is decomposed as a sum of sines (which is possible due to Assumption 4), the contributions of each sine to the output can be identified.

Definition 1. For a linear dynamic system, an elementary design is a sine at frequency ω_k with unit power.

C. Information matrix

Under the assumptions stated above, it is shown in [4] that the information matrix Fi can be written as a weighted sum of elementary information matrices Fi_k .

$$Fi = \sum_{k=1}^{N_\omega} \xi_u(k) \cdot Fi_k$$

The weights $\xi_u(k)$ in the sum correspond to the discrete power spectrum of the input signal.

Definition 2. An elementary information matrix Fi_k is the information matrix computed for the k^{th} elementary design. In the linear case Fi_k can be computed as:

$$Fi_k(i, j) = \frac{2}{\sigma^2} \frac{\partial \overline{G(\omega_k)}}{\partial \theta_i} \frac{\partial G(\omega_k)}{\partial \theta_j} \quad (1)$$

D. Dispersion

In order to find the power spectrum that maximizes the determinant of the information matrix, an indirect approach is used based on the dispersion function [2].

Definition 3. Under Assumptions 1 to 4, and for an input with discrete spectrum ξ_u , the dispersion function $v(\cdot, \cdot)$ at frequency ω_k is defined as¹:

$$v(\xi_u, k) = \text{trace}(Fi^{-1} \cdot Fi_k) \quad (2)$$

where Fi is the information matrix computed for the given power spectrum ξ_u and Fi_k is the information matrix corresponding to a sine input at frequency ω_k .

Some useful properties of the dispersion function are [2]:

- The maximal value of the dispersion can never be lower than the number of free parameters in the model.
- The inner product between the dispersion $v(\xi_u, k)$ and the corresponding spectrum $\xi_u(k)$ equals the number of free parameters.
- The dispersion can also be interpreted as a normalized variance on the estimated model.

In [2] it is shown that the power spectrum that minimizes the maximal value of the dispersion function also maximizes the determinant of the information matrix.

Theorem 1. The following characterizations of an optimal power constrained design are equivalent:

- 1) $\xi_{u_{opt}}$ maximizes $\det(Fi)$
- 2) $\xi_{u_{opt}}$ minimizes $\max_k v(\xi_u, k)$
- 3) $\max_k v(\xi_{u_{opt}}, k) = n_\theta$

where n_θ is the number of free parameters in the model.

Proof: see [2] Chapter 6 page 147.

E. Optimization algorithm

In [4] a simple and robust, monotonically converging algorithm is defined, which finds a discrete spectrum that minimizes the dispersion on a predefined frequency grid. This algorithm can be summarized in four steps:

- 1) Initialize the power spectrum uniformly

$$\xi_u(k) = 1/N_\omega$$

with N_ω the total number of frequencies.

- 2) Compute the dispersion function $v(\xi_u, k)$ for the current input using (2)
- 3) Update the power spectrum of the input in correspondence with the dispersion function as follows

$$\xi_{new}(k) = \frac{v(\xi_u, k)}{N_\theta} \cdot \xi_{old}(k)$$

- 4) Stopping criterion: if $(\max_k v(\xi_{new}, k) - N_\theta)$ is smaller than a predefined threshold, the optimal solution is assumed to be found; else go to step 2.

The stopping criterion is based on the third expression of Theorem 1. The monotonic convergence of the algorithm is proven in [8].

¹The dispersion function is also called response dispersion.

F. Signal generation

The algorithm above provides an optimal weighted linear combination of elementary designs, of which each term consists of a single sine at frequency ω_k with specified power $\xi(k)$. All these weighted elementary designs could be applied separately, one after another. However, this would be inefficient with respect to measurement time. Instead, the elementary designs can be applied all at once with the signal u_{opt} , which is the weighted linear combination of the elementary designs.

$$u_{opt}(t) = \sum_{k=1}^{N_\omega} \sqrt{\xi(k)} \cdot \cos(\omega_k \cdot t + \phi_k)$$

The phases ϕ_k of the sines can be chosen freely. Notice that no additional constraints are needed in order to perform the optimal experiment with a single time sequence.

III. NONLINEAR DYNAMIC SYSTEMS

In the previous section, an optimal input design based on the dispersion function was introduced for the linear dynamic case. This design method will now be extended to a class of nonlinear dynamic systems. In order to keep the explanation of the method as simple as possible, all concepts will be explained for a 2-tap FIR filter followed by a cubic nonlinearity.

A. Problem statement

Assumptions 2, 3 and 5 from the linear dynamic case are kept, while Assumption 1 and 4 are replaced with the assumptions below.

Assumption 6. The considered system consists of a 2-tap FIR filter, followed by a cubic nonlinearity.

$$\begin{aligned} w(t) &= b_1 u(t) + b_2 u(t-1) \\ y_0(t) &= w(t)^3, \quad y(t) = y_0(t) + e(t) \end{aligned}$$

where $u(t)$ is the noiseless input, $y_0(t)$ is the noiseless output and b are the parameters of the FIR filter.

Assumption 7. The class of inputs will be restricted to deterministic time sequences of given length $N + 1$. The amplitude can only take values from a finite, predefined set of A values.

$$\forall t: u(t) \in \{u_1, \dots, u_A\} \quad (3)$$

Unlike the linear case, no direct constraints are imposed on the total power of the input. Of course, limiting the maximal amplitude value and fixing the signal length restricts the total power indirectly.

B. Elementary designs

Considering the system model, it is clear that $y(t)$ only depends on the values $u(t)$ and $u(t-1)$. Therefore each couple $(u(t-1), u(t))$ is considered to be an elementary design for the system. Because the possible amplitude values of the input are discretized, the number of distinct couples is finite.

Definition 4. An ordered set of 2 values drawn out of the predefined set $\{u_1, u_2, \dots, u_A\}$ is called a couple. In total A^2 different couples can be defined. All possible couples will be stored in a multidimensional cell array called $C \in \{\mathbb{R}^2\}^{A \times A}$ where each cell contains a couple.

$$\begin{aligned} C(i_1, i_2) &= (u_{i_1}, u_{i_2}) \\ \forall i_1, i_2 &\in \{1, 2, \dots, A\} \end{aligned} \quad (4)$$

Definition 5. For a system consisting of a 2-tap FIR filter followed by a cubic nonlinearity, an elementary design consists of a single couple.

Notation 1. When a $A \times A$ matrix X is rewritten as a $A^2 \times 1$ vector x , the following mapping between the linear index k and subindices i_1 and i_2 will be used:

$$\begin{aligned} x(k) &= \text{Vec}(X(i_1, i_2)) \\ k &= i_1 + A(i_2 - 1) \end{aligned} \quad (5)$$

C. Information matrix

Under Assumption 2 and given a specific input sequence $u(t)$ of length $N + 1$, the information matrix can be written as a sum over the time samples:

$$Fi(i, j) = \frac{1}{\sigma^2} \sum_{t=1}^N f_{i,j}(u(t-1), u(t), \theta) \quad (6)$$

$$f_{i,j}(u(t-1), u(t), \theta) = \frac{\partial y_0(t, \theta)}{\partial b_i} \frac{\partial y_0(t, \theta)}{\partial b_j} \quad (7)$$

where $\theta = [b_1 \ b_2]$. It is assumed that $y(0)$ is omitted during the estimation because the value of $u(-1)$ is unknown. Because the derivatives only depend on two successive input values, the number of different terms equals the number of different couples in the array C . By grouping the equal terms, the sum over time can be rewritten as a sum over the couples:

$$Fi(i, j) = \frac{1}{\sigma^2} \sum_{k=1}^{A^2} \xi_N(k) \cdot f_{i,j}(c(k), \theta) \quad (8)$$

where c is the vectorized version of the matrix C (see Notation 1). The weights $\xi_N(k)$ indicate how often each couple occurs in the sequence $u(t)$. An optimal design will consist of optimizing over the $\xi_N(k)$.

Definition 6. The vector containing the number of times a couple occurs in a signal $u(t)$ is called the couple frequency vector ξ_N of that signal. The couple frequency vector divided by the total number of couples N in the signal is called the normalized couple frequency vector ξ of the signal $u(t)$:

$$\xi = \frac{1}{N} \cdot \xi_N \quad (9)$$

Remark 1. The normalized couple frequencies have similar properties as a unit power spectrum (see Assumption 4). The sum of a couple frequency vector always equals one, and each entry lies between zero and one.

$$\forall k: \xi(k) \in [0, 1] \text{ and } \sum_{k=1}^{A^2} \xi(k) = 1 \quad (10)$$

Coming up with an optimal experiment consists of finding the vector ξ_N that optimizes the determinant of the information matrix, given that $N + 1$ samples can be measured. This problem is very similar to finding the optimal input spectrum for a linear system, given a total power constraint.

In order to recover the optimization algorithm based on the dispersion of the linear problem, the information matrix needs to be written as a weighed sum of matrices. Moreover, the weights all have to be positive and should sum to one. Such formulation can be obtained if the problem is normalized with respect to the total number of couples N in the signal:

$$\frac{Fi(i, j)}{N} = \sum_{k=1}^{A^2} \xi(k) \cdot Fi_k \quad (11)$$

Definition 7. For a FIR filter followed by a cubic nonlinearity the elementary information matrix Fi_k can be computed as:

$$Fi_k(i, j) = \frac{1}{\sigma^2} f_{i,j}(c(k), \theta) \quad (12)$$

The optimization algorithm, described in subsection II-E for linear systems, can now be used unchanged to find the normalized couple frequency vector ξ_{opt} that maximizes the normalized information matrix for the nonlinear dynamic systems of this section.

D. Signal generation

The optimal couple frequency vector $\xi_{N,opt}$ can be interpreted as an experiment of N measurements, where each measurement consists of applying a single couple to the system and measuring the corresponding output sample.

A naive way to perform this experiment is to place all the couples in succession, and only measure the output samples which correspond to these couples. This means that a signal with $2N$ samples is used at the input, but only N samples are used at the output. Clearly this is not an efficient approach, since half of the output samples is omitted.

It would be better to generate an input sequence of $N + 1$ samples containing the N couples needed for the experiment as defined by the optimal couple frequency vector. A possible generation method could be based on Markov chains as was already mentioned in [5]. However, not every couple vector has a corresponding input sequence of length $N + 1$, because the second input of an optimal couple may not correspond to the first input of another optimal couple.

The conditions needed for the existence of at least one time sequence of length $N + 1$ containing the couples of the optimal frequency vector are:

$$\forall i \in \{1, \dots, A\} : \sum_{j=1}^A \Xi_{N,opt}(i, j) = \sum_{j=1}^A \Xi_{N,opt}(j, i) \quad (13)$$

where $\Xi_{N,opt}$ is the matrix representation of $\xi_{N,opt}$ (see Notation 1). These conditions guarantee that for each value $u_i \in \{u_1, \dots, u_A\}$ the number of couples starting with u_i are equal to the number of couples ending with u_i , such that all couples can be placed in succession without creating unwanted transition couples.

E. Constrained optimization

To impose the constraints (13) during the optimization, such that the obtained solution can be realized as a time signal, the frequency vector will be rewritten as a linear combination of N_b base vectors $\xi_{b_i} \in \mathbb{R}^{A^2}$. These base vectors span the space of vectors that are in accordance with the constraints (13).

$$\forall k \in \{1, 2, \dots, A^2\} : \xi(k) = \sum_{i=1}^{N_b} \gamma_i \cdot \xi_{b_i}(k) \quad (14)$$

The optimization will now be performed with respect to the coefficients γ_i .

If the optimization algorithm is now used to find the optimal values γ_i , these values will all be positive and sum to one; however it is not guaranteed that the couple frequency vector $\xi(k)$ defined by (14) has the required properties (10) and (13). Therefore additional constraints are imposed on the basis.

Using a non-overlapping symmetric basis is a sufficient (but not necessary) condition to guarantee that the couple frequencies will have the required properties. This allows the re-use of the optimization algorithm, with coefficients in the role of the couple frequencies.

Definition 8. A set of vectors $[\xi_{b_1}, \xi_{b_2}, \dots, \xi_{b_{N_b}}]$ in \mathbb{R}^{A^2} is called non-overlapping symmetric if they have the following four properties:

- 1) *positivity constraint:*
 $\forall k, \forall i : \xi_{b_i}(k) \geq 0$
- 2) *non-overlapping constraint:*
 $\forall k, \exists! i : \xi_{b_i}(k) > 0$
- 3) *unity sum constraint:*
 $\forall i : \sum_{k=1}^{A^2} \xi_{b_i}(k) = 1$
- 4) *symmetry constraint:*
 $\forall i_1, i_2 : \Xi_{b_i}(i_1, i_2) = \Xi_{b_i}(i_2, i_1)$

where Notation 1 is used to rewrite the vectors as matrices.

Remark 2. The first two properties guarantee that the couple frequencies are positive, the third property makes sure that the sum of the couple frequencies is one, while the fourth property imposes symmetry on the couple frequency matrix such that the constraints of (13) are fulfilled.

Remark 3. The number of non-overlapping symmetric base vectors N_b equals $\frac{A(A+1)}{2}$. This corresponds with the number of degrees of freedom in a symmetric matrix of size $A \times A$

Example 1. To illustrate the definition, consider that only two different amplitude values are possible say u_1 and u_2 ; the non-overlapping symmetric base is then:

$$\Xi_{b_1} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \Xi_{b_2} = \begin{bmatrix} 0 & 0.5 \\ 0.5 & 0 \end{bmatrix}, \Xi_{b_3} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

In order to optimize the dispersion function over the coefficients γ_j , the expression for the normalized information matrix (11) will be rewritten:

$$\frac{Fi}{N} = \sum_{j=1}^{N_b} \gamma_j \sum_{k=1}^{A^2} \xi_{b_j}(k) \cdot Fi_k = \sum_{j=1}^{N_b} \gamma_j \cdot Fi_{\gamma_j} \quad (15)$$

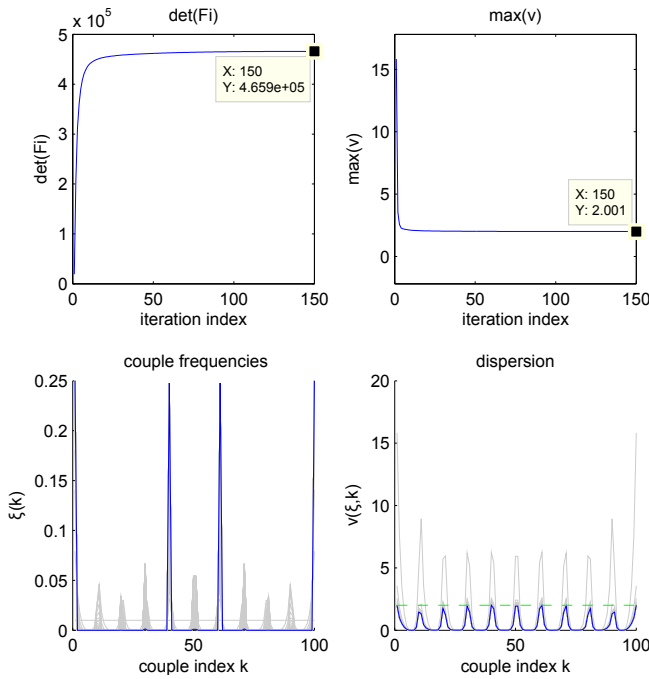


Fig. 2. Results without constraints. Top, left: determinant of the normalized information matrix. Top, right: maximal dispersion. Bottom, left: couple frequencies. Bottom, right: dispersion. Final value in blue, others in gray. The green, dotted line is the theoretical minimum.

where Fi_{γ_j} are the new elementary information matrices. During the constrained optimization the dispersion will be computed based on these Fi_{γ_j} .

$$v_{\gamma}(\xi, j) = \text{trace}(Fi^{-1} \cdot Fi_{\gamma_j}) \quad (16)$$

Notice that the values and dimension of the dispersion functions $v_{\gamma}(\xi, j)$ and $v(\xi, k)$ are different, while the information matrix remains invariant.

IV. NUMERICAL EXAMPLE

The above results will now be illustrated on the following numerical example:

$$\begin{aligned} y(t) &= [b_1 u(t) + b_2 u(t-1)]^3 + e(t) \\ u(t) &\in [-1 : 2/9 : 1] \quad \text{and} \quad e(t) \sim N(0, 1) \end{aligned}$$

with $b = (1, 3)$. Notice that the amplitude set contains 10 uniformly spaced values between -1 and 1. This means that 100 different couples should be considered.

A. Unconstrained optimization

First, the method without the constraints was applied. In total 150 iterations steps were performed. In each step the dispersion, couple frequency and determinant of the normalized information matrix was computed and stored. The evolution of these values can be seen in the plots of Fig. 2.

Considering the top plots, it becomes clear that the method successfully lowers the maximal value of the dispersion and at the same time increases the determinant of the information matrix. Important to note is that in the end the maximal

TABLE I

TABLE CONTAINING THE COUPLE INDEX, COUPLES AND COUPLE FREQUENCIES OF THE UNCONSTRAINED SOLUTION

couple index	couple	frequency
1	[1,1]	0.25
40	[-1,1/3]	0.25
61	[1,-1/3]	0.25
100	[-1,-1]	0.25

dispersion reaches the value of 2, which corresponds to the number of free parameters. This indicates that the obtained solution is optimal (Theorem 1).

In the bottom right plot, the evolution of the couple frequencies is shown. Initially, all couples have the same frequency, while at the end only a limited set of couples has a nonzero frequency.

Table I contains the couple indices, couples and couple frequencies of the final solution after truncation of the frequencies (see Remark 4). From this table it is clear that there can never be an input sequence that corresponds to these frequencies. For example the couple $[-1, -1/3]$ can never be placed without making a transition couple that is not a part of the optimal design.

Remark 4. The results of Table I, and also Tables II and III below, are truncated: all values of the couple frequency vector that are two orders of magnitude smaller than the maximal value are put to zero. Afterwards, the vector is rescaled so that the sum of its elements equals one.

B. Constrained optimization

Next, the optimization was performed with constraints as explained in Section III-E. This leads to the results of Fig. 3. They are very similar to the unconstrained case. Again the determinant is systematically increased, while the maximal value of the dispersion v_{γ} is driven to its minimal value.

The dispersion of the constrained and unconstrained solution can not be compared directly, because different elementary designs are used. In order to make a comparison possible, the dispersion and determinant of the constrained solution is recomputed for the elementary designs Fi_k .

The results are plotted in Fig. 4. From the right plot it is clear that v has higher values than v_{γ} . This indicates that the constrained solution is not optimal if the full solution space is considered. The left plot shows that the determinant is independent of the considered elementary designs, as was expected from (15).

The fact that the dispersion v_{γ} reaches the value 2 in Fig 3 indicates that the constrained solution is the best possible solution, spanned by the non-overlapping symmetric base vectors. Of course this does not exclude the possible existence of a non-symmetric frequency vector that still complies with the constraint (13) and that has a higher determinant than the symmetric solution.

By looking at Table II, which contains the couple indices, couples and frequencies of the constrained solution, it becomes clear that imposing constraints leads to a couple

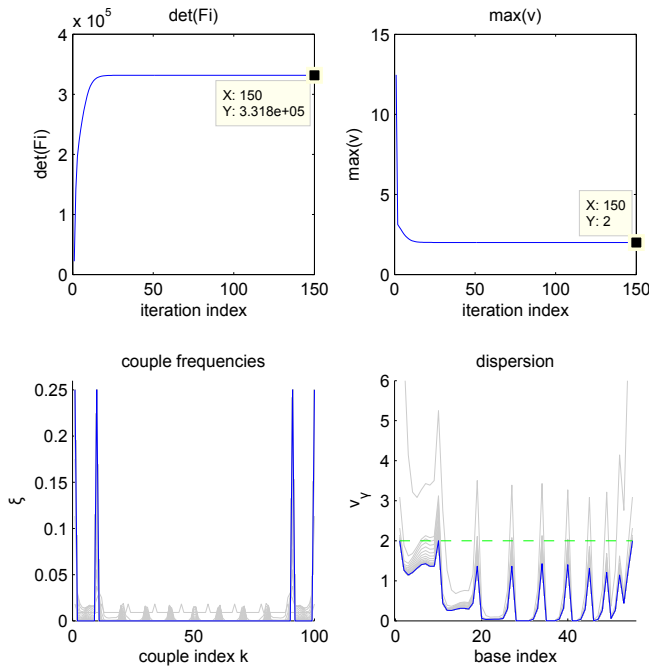


Fig. 3. Results with constraints. Top, left: determinant of the normalized information matrix. Top, right: maximal dispersion. Bottom, left: couple frequencies. Bottom, right: dispersion. Final value in blue, others in gray. The green, dotted line is the theoretical minimum.

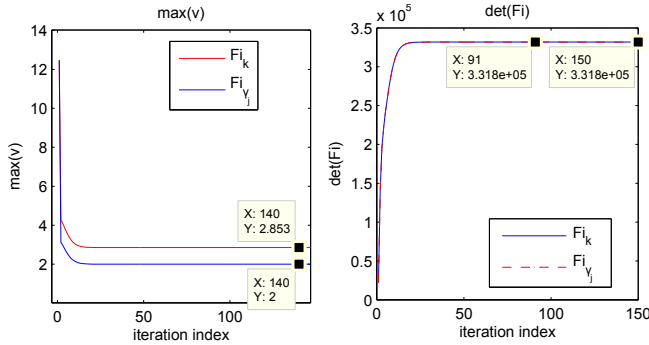


Fig. 4. Maximal dispersion and determinant of the normalized information matrix in the constrained case, computed for different elementary designs. Full space corresponds to F_{i_k} . Subspace corresponds to $F_{i_{\gamma_j}}$

frequency vector which corresponds to a realizable time sequence.

C. Grid search

To confirm that the algorithm with constraints does not get stuck in a local minimum, an exhaustive grid search was performed for the same example, but only considering 2 different amplitude values -1 and 1. The frequency values were discretized in 20 levels between 0 and 1. All possible frequency vectors are generated for which the sum of their elements equals one. For each vector the dispersion and determinant of the information matrix were computed.

The results of this grid search are presented in Table III and compared with the constrained and unconstrained solutions obtained with the iterative algorithm.

The best solution found during the grid search has the

TABLE II
TABLE CONTAINING THE COUPLE INDEX, COUPLES AND COUPLE FREQUENCIES OF THE CONSTRAINED SOLUTION

couple index	couple	frequency
1	[1,1]	0.25
40	[-1,1]	0.25
61	[1,-1]	0.25
100	[-1,-1]	0.25

TABLE III
TABLE CONTAINING THE RESULTS OF THE 2 DIMENSIONAL GRID SEARCH

solution type	det(Fi)	max(v)
starting value	8.04e+04	3.32e+00
unconstrained	1.94e+05	2.00e+00
constrained	1.42e+05	2.85e+00
best unconstrained grid search	1.94e+05	2.00e+00
best constrained grid search	1.42e+05	2.85e+00

same determinant as the unconstrained iterative solution. However, if only the solutions satisfying the constraints (13) are considered, no better solution than the constrained one was found. Similar results were found in the case where three different amplitudes are considered.

V. CONCLUSION

In this work, an optimization scheme, based on the dispersion was adapted for systems consisting of a 2-tap FIR filter followed by a cubic nonlinearity. The optimization was done with respect to the couple frequency vector, instead of the time sequence. Without constraints, a solution was found that did not correspond to a realizable time sequence. By imposing that the solution should lie in the subspace described by a symmetric and non-overlapping base, a realizable solution was obtained that is optimal in its subspace.

REFERENCES

- [1] V.V.Federov, *Theory of Optimal Experiments*. Academic Press, 1972.
- [2] G.C.Goodwin and R.L.Payne, *Dynamic System Identification: Experiment Design and Data Analysis*, vol. 136 of *Mathematics in Science and Engineering*. Academic Press, 1977.
- [3] H.Hjalmarsson, J.Martensson, and B.Ninness, "Optimal input design for identification of non-linear systems: Learning from the linear case," in *American Control Conference*, 2007.
- [4] J.Schoukens and R.Pintelon, *Identification of Linear Systems: A Practical Guideline to Accurate Modeling*. Pergamon Press, 1991.
- [5] C.A.Larsson, H.Hjalmarsson, and C.R.Rojas, "On optimal input design for nonlinear FIR-type systems," in *49th IEEE Conference on Decision and Control*, 2010.
- [6] M. Gevers, M. Caenepeel, and J. Schoukens, "Experiment design for the identification of a simple wiener system," in *IEEE Conference on Decision and Control*, 2012.
- [7] R. Pintelon and J. Schoukens, *System Identification: A Frequency Domain Approach*. John Wiley & Sons, 2004.
- [8] Y. Yu, "Monotonic convergence of a general algorithm for computing optimal designs," *The Annals of Statistics*, vol. 38, pp. 1593–1606, 2010.