

Approximate Spectral Factorization for Design of Efficient Sub-Filter Sequences

Björn Norell*, Oleg Burdakov, Mats Andersson and Hans Knutsson,

Abstract—A well-known approach to the design of computationally efficient filters is to use spectral factorization, i.e. a decomposition of a filter into a sequence of sub-filters. Due to the sparsity of the sub-filters, the typical processing speedup factor is within the range 1–10 in 2D, and for 3D it achieves 10–100. The design of such decompositions consists in choosing the proper number of sub-filters, their individual types and sparsity. We focus here on finding optimal values of coefficients for given sequences of sparse sub-filters. It is a non-convex large scale optimization problem. The existing approaches are characterized by a lack of robustness - a very slow convergence with no guarantee of success. They are typically based on generating random initial points for further refinement with the use of local search methods. To deal with the multi-extremal nature of the original problem, we introduce a new constrained optimization problem. Its solution is then used as an initial point in the original problem for further refinement. Our approach is applicable to designing multi-dimensional filters. Its efficiency and robustness is illustrated by designing sub-filter sequences for 2D low-pass, band-pass and high-pass filters of approximately the same quality as with the use of a standard approach, but with the overall design speedup factor of several hundred.

I. INTRODUCTION

Noise reduction, segmentation, motion analysis and registration are central in medical image analysis. Algorithms for solving such problems typically rely on features extracted using linear filters [1], [2], [3]. Many features, despite of that they are well-performing and widely used in research, are often ruled out by clinical demands on computation time. This happens for the lack of efficient implementations, especially when the signal dimensionality d is over two.

When computation time is an issue, common practice is to choose features based on Gaussian functions, which are Cartesian separable and can be efficiently implemented. Another approach is to use decomposition techniques like in [4] to approximate an arbitrary filter by a sequence of d

1-dimensional sub-filters, one along each signal dimension. The latter approach is then slightly more flexible. However, for the approximation to be accurate this approach imposes severe limitations on the filter which must be nearly Cartesian separable.

A more general approach is to decompose the desired filter into a sequence of sparse sub-filters instead of 1-dimensional sub-filters. The sparsity is here the key property to achieve efficiency. The freedom to choose the locations of the non-zero coefficients of each individual sub-filter allows accurate approximations of a much larger class of filters compared to the former two approaches. Decomposing a filter into a sequence of sparse sub-filters is far from a new idea, but so far most efforts have been focused on 1D. An overview of early results on this topic is given in [5]. For multi-dimensional signals such decompositions are indirectly employed in e.g. filter banks, wavelets and image pyramids [6], [7]. These decompositions are in general multi-rate¹ tree-structured systems with several output nodes. The design criteria used are often different from the framework presented here and typically requires perfect reconstruction. The purpose of these systems is however rarely efficient convolution and feature extraction, which are the primary goals for the sub-filter sequences presented here. Three examples of features that can be facilitated by the use of sub-filter sequences are local structure tensors, anisotropy and signal phase illustrated in Fig. 1.

A. Applications

Firstly, we see an example of a local structure tensor field estimated from a T1-weighted magnetic resonance image. The feature known as local structure is central for the edge-preserving denoising techniques such as anisotropic diffusion and adaptive anisotropic filtering (see e.g. [8], [9]). The latter is an explicit method, which for 3D requires the computation of 16 linear filter responses. This is prohibitively time-consuming and the use of filter networks for efficient filtering was an important improvement of this method [10].

The second example is microscopy, where the image shown is a time frame from a sequence used for analysis of stem cell movement patterns. The motion is obtained from the signal phase, which is the feature used in this example. Extracting motion from signal phase in 2D requires computation of 5 filter responses. The use of efficient filters in video applications is highly motivated, since video processing in close to real-time is in general desired. Another option is to treat the image sequence as a volume and extract motion using 3D filtering,

This work was supported by VINNOVA, the Swedish Governmental Agency for Innovation Systems, and the Linköping University Center for Industrial Information Technology (CENIT).

* Corresponding Author: B. Norell, currently at Philips Digital Mammography Sweden AB, was with the Department of Biomedical Engineering, Medical Informatics and Center for Medical Image Science and Visualization (CMIV) at Linköping University. (e-mail: bjosv@imt.liu.se)

M. Andersson and H. Knutsson are with the Department of Biomedical Engineering, Medical Informatics and Center for Medical Image Science and Visualization (CMIV) at Linköping University. (postal address: Linköping Universitet, Institutionen för Medicinsk Teknik, SE - 581 85 Linköping, Sweden)(e-mail: matsa@imt.liu.se; knutte@imt.liu.se) (phone: +46 13 22 00 00) (fax: +46 13 10 19 02)

O. Burdakov is with the Department of Mathematics, at Linköping University. (postal address: Linköping Universitet, Matematiska Institutionen, SE - 581 83 Linköping, Sweden) (e-mail: Oleg.Burdakov@liu.se) (phone: +46 13 28 14 73) (fax: +46 13 10 07 46)

¹A multi-rate system makes use of sub-sampling.

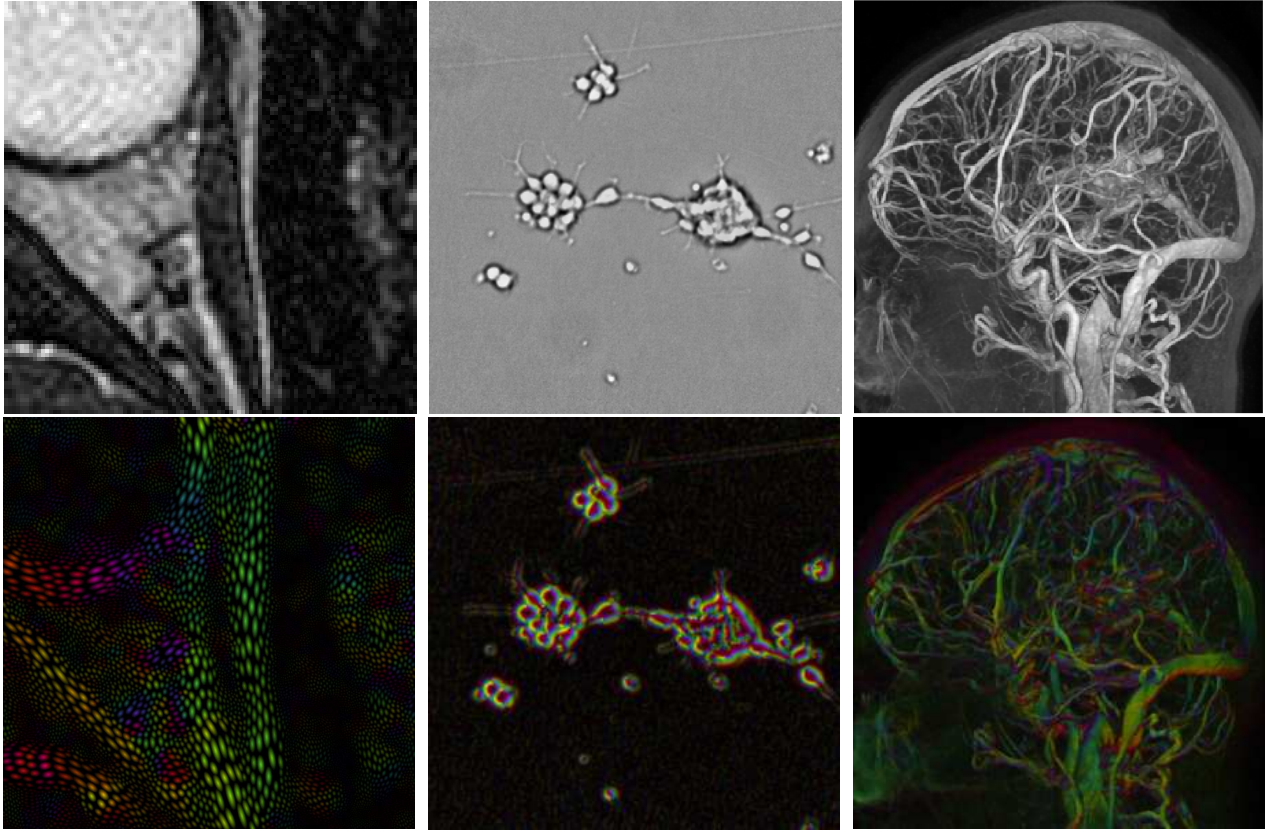


Fig. 1. The local structure tensor field describes how the signal energy of the MRI image varies locally. Tensors, which describe low energy, are displayed with low intensity and the colors of the tensors show how they are oriented. The signal phase is colored to visualize the stem cell movement between the frame displayed and the adjacent frames. The opacity of structures in this MR angiography is mapped by a measure of anisotropy, which delineate the blood vessels from the background. The color represents the vessel orientation in the coronal plane, which is orthogonal to the sagittal projection shown.

which would improve the accuracy of the motion analysis. On the other hand, this would require 9 filters which would dramatically increase the processing time.

The last example is a high resolution magnetic resonance angiography, where the features anisotropy and orientation are used to describe and delineate the blood vessels. Since the blood vessels significantly vary in size analysis in several scales would improve the accuracy. Typically two or three scales are used and the number of filters required would then be 9 for each scale. Again this is an example of where sub-filter sequences can reduce the processing time.

B. Sub-filter Sequences and Properties of their Design Models

It is convenient to express the desired filter in the Fourier domain as a frequency response. In fact many filters do not have a closed form expression in the spatio-temporal domain. If such an expression exists it is rarely of finite size and can not be implemented efficiently. By enforcing a finite spatial size we implicitly impose a smoothness constraint on the frequency response, which means that we can only hope for an approximation of the desired filter. The frequency response of a filter sequence is the product of all sub-filter frequency responses. Subsequently, a sequence of sub-filters with finite spatial size and sparsity constraints is then an approximate factorization of the desired frequency response.

Designing filters in this fashion is performed in two steps. First, certain type and sparsity are chosen individually for each sub-filter. This step is discussed in [11]. The intensively developing areas of sparse optimization and compressed sensing [12] offer new efficient techniques that can be used for choosing the sparsity.

In this paper, we focus on the second step at which values of the sparse sub-filter coefficients are to be chosen to ensure the best approximation of the desired filter. The step requires solving a multi-linear least squares problem. The most challenging of such problems originate from large filter sets or dimensionalities above 1. They are characterized not only by a very large number of variables, but also by a large number of local minimizers. Other research areas where similar multi-linear least squares problems occur are for instance factor analysis, chemometrics and spectroscopy [13], [14], [15]. The multi-extremal nature of these problems makes them difficult, and in some cases even practically impossible to solve with conventional methods. Typically, a combination of local optimization methods with randomly generated initial points is applied with the hope to obtain desired solutions. One of the most efficient local search methods is alternating least squares (ALS) [16], [17]. The randomly generated initial points often result in convergence to the same local solutions. Moreover the convergence of local search methods is very slow because the problem is highly ill-conditioned.

The sub-filter sparsity optimization requires finding sub-filter coefficients for each of, potentially numerous, candidate sparsities. Therefore, it is of critical importance to develop computationally efficient methods for the step two which is the main aim of this paper.

C. Solution Strategy

We present an approach for efficiently solving a multi-linear least squares problem originating from design of filters implemented as a sequence of sparse sub-filters. Since the problem is directly intractable, we propose a two-stage process. Given an arbitrary frequency response we search for a solution in the space of all *exact* factorizations. The solution we seek is the exact spectral factorization with the closest fit to the *approximate* factorization that satisfies the smoothness constraints imposed by the sub-filters. The solution found constitutes a good initial point for further refinement of the solution using conventional local methods.

In our numerical experiments, ALS started from randomly generated points seldom produced better results compared to the presented approach. Moreover, the differences between the best ALS solutions and the solutions obtained with the presented approach were vanishing. In this sense our approach is more robust and provides better local minimizers. It also reduces the computation time, because in order to get comparable results one must run ALS a sufficiently large number of times.

II. LEAST SQUARES FILTER DESIGN

To simplify our presentation of the design problem for a sequence of sparse sub-filters, we give in this section the least squares formulation of the FIR filter design problem [18]. The latter problem is to choose values of the complex non-zero coefficients $c \in C^n$ in the discrete impulse response $\tilde{f}(\xi)$ so that they provide the closest fit to a number of desired functions. Without loss of generality we can associate each filter coefficient, i.e. each element c_k of c , with a discrete spatio-temporal position $\xi_k \in Z^d$ such that $\tilde{f}(\xi_k) \neq 0$.

The number of non-zero filter coefficients n define the number of multiplications and additions per signal sample in the filtering process. For a single non-sparse filter, n grows exponentially with the signal dimensionality d . Since the number of computations is proportional to n , it is of utmost importance to keep n as low as possible for multi-dimensional FIR filters. Given discrete spatio-temporal positions $\{\xi_k\}_{k=1}^n$ and coefficients $c = [c_1, c_2, \dots, c_n]^T \in C^n$, the corresponding frequency response $\tilde{F}(u)$ is defined as the Fourier transform $\tilde{F}(u) = \mathcal{F}\{\tilde{f}(\xi)\}$ by the formula:

$$\tilde{F}(u) = \sum_{Z^d} \tilde{f}(\xi) \exp(-i u^T \xi) = \sum_{k=1}^n c_k \exp(-i u^T \xi_k) \quad (1)$$

which means that $\tilde{F}(u)$ depends linearly on c .

It is natural to choose the values of c so that it makes the frequency response $\tilde{F}(u)$ as close to the desired frequency response $F(u)$ as possible. This leads to minimizing the approximation error $\varepsilon(u) = F(u) - \tilde{F}(u)$. Common practice

in design of multi-dimensional filters is to choose c so that it solves the following weighted linear least squares problem

$$\min_c \|\varepsilon(u)\|^2 = \int_U |W(u) \varepsilon(u)|^2 du, \quad (2)$$

$$U = \{u \in R^d : |u_i| \leq \pi\},$$

where $\|\cdot\|$ is a weighted Euclidean norm. The weight function $W(u)$ is used to emphasize the importance of certain frequencies. Due to the 2π -periodicity of the Fourier domain, it is sufficient to study $\varepsilon(u)$ over the set U only. To treat the design problem using conventional numerical analysis, the set U is traditionally sampled by a proper choice of frequencies $\bar{u}_1, \dots, \bar{u}_m \in U$. The sampling defines the residual vector $r(x) = Ax - b \in C^m$, where

$$A = \begin{bmatrix} \exp(-i \bar{u}_1^T \xi_1) & \dots & \exp(-i \bar{u}_1^T \xi_n) \\ \vdots & \ddots & \vdots \\ \exp(-i \bar{u}_m^T \xi_1) & \dots & \exp(-i \bar{u}_m^T \xi_n) \end{bmatrix},$$

$$b = \begin{bmatrix} F(\bar{u}_1) \\ \vdots \\ F(\bar{u}_m) \end{bmatrix}, \quad x = c. \quad (3)$$

Note that the filter design is determined not only by the filter coefficients denoted by x , but also by the spatio-temporal positions incorporated in A . In this paper, A is predetermined, and therefore, we focus on finding optimal value of x , which we call design variable. Furthermore, we assume without loss of generality that A , x and b are real-valued. This can be accomplished by splitting each variable into its real and imaginary components. Another option is to impose symmetry constraints on the filter to be designed.

Uniform sampling of the frequency domain results in the Riemann sum approximation

$$\|\varepsilon(u)\|^2 \approx \frac{(2\pi)^d}{m} \|r(x)\|^2. \quad (4)$$

Under the standard conditions, the right-hand side in (4) tends to the left-hand side as m tends to infinity. To assure that this approximation is accurate the set U must be sampled densely enough. The number of samples needed depends on how smooth $F(u)$ is. Practical experience has shown that a good rule of thumb is to at least have $m > 2^d n$, i.e. twice the number of samples per dimensionality [17].

The relation (4) reduces (2), and hence the FIR filter design, to the linear least squares problem:

$$\min_x \|b - Ax\|^2. \quad (5)$$

Here and henceforth $\|\cdot\|$ denotes a weighted Euclidean vector norm in R^m .

III. MULTI-LINEAR LEAST SQUARES DESIGN

Consider a sequence of l sub-filters (see Fig. 2). Let $x_i \in R^{n_i}$ and $A_i \in R^{m \times n_i}$ be individual characteristics of sub-filter $i = 1, \dots, l$, like those introduced in (3) for a single filter. By the convolution theorem, the frequency response of the sub-filter sequence is $(A_1 x_1) \circ (A_2 x_2) \circ \dots \circ (A_l x_l)$, where $u \circ v$ denotes the component-wise product of vectors u and

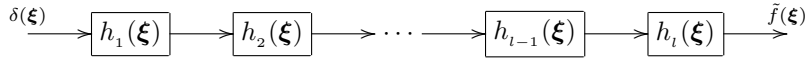


Fig. 2. The computational load of filtering, i.e. the number of nonzero coefficients, can be reduced by designing and implementing the impulse response $\tilde{f}(\xi)$ using a sequence of sub-filters $h_i(\xi)$, $i = 1, 2, \dots, l$.

v. With A_1, \dots, A_l predetermined for the sub-filter sequence, the best fit for the ideal frequency response $b \in R^m$ is ensured by solving the multi-linear least squares (MLLS) problem

$$\min_{x_1, \dots, x_l} \|b - (A_1 x_1) \circ (A_2 x_2) \circ \dots \circ (A_l x_l)\|^2. \quad (6)$$

It is a generalization of linear least squares problem (5) in which case $l = 1$.

In design of d -dimensional filters, the number of frequency samples m is large, typically 10^d - 40^d . In contrast to dense filters, the number of non-zero coefficients $n = n_1 + \dots + n_l$ grows in general linearly with d , and it can range within $5d$ - $60d$, depending on the number of sub-filters l . The value of n can be decreased by increasing l . However, n is determined to a large extent by the desired filter response, and we believe that 2-10 sub-filters cover most of the practical applications. MLLS is a non-convex large-scale optimization problem. It has a large number of local minima. One of the reasons is that if x_1, \dots, x_l is a local minimizer, then any change of signs for x_1, \dots, x_{l-1} compensated by the proper choice of sign for x_l does not change the objective function value in (6). Note that the number of such combinations is equal to 2^{l-1} . This is not the only source for the large number of local minima, which is illustrated by an example considered in the next sub-section. Moreover, each local minimizer is singular and non-isolated. Indeed, if x_1, \dots, x_l is a local minimizer, then $t_1 x_1, \dots, t_l x_l$ is also a local minimizer for any combination of scalars t_1, \dots, t_l such that $t_1 \cdot \dots \cdot t_l = 1$. This means that each local minimizer belongs to a surface of local minimizers. All these characteristics of MLLS problem explains why the conventional unconstrained minimization methods and nonlinear least squares methods are inefficient in finding local minimizers of this problem.

Optimization of filter sequences and also sub-filters of a more general network structure can be performed using ALS. It is aimed at finding a local minimizer of MLLS problem (6) by sequentially solving linear least squares sub-problems obtained from (6) by fixing all but one vector x_i , where the index i is alternating. Although each of the sub-problems is easy to solve, ALS converges very slowly. This motivates the necessity of finding more efficient ways of solving the MLLS problem.

A. Example

Consider the following filter design problem. Given a 2D lognormal band-pass filter whose frequency response is shown on Fig. 3, provide its best approximation by a sequence of two sparse 2D sub-filters of the sizes 3×3 and 7×7 . The kernels of these sub-filters are required to be radially symmetric. Then for the first spatial sub-filter, its sparse 3×3 kernel with 5 non-zero coefficients (the origin and its 4 closest neighbors) can be parameterized by 3 coefficients as shown on Fig. 4.

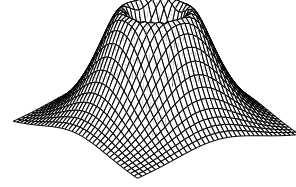


Fig. 3. The frequency response of the ideal lognormal band-pass filter.

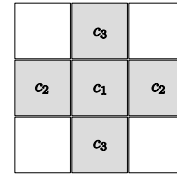


Fig. 4. Sparse pattern for the 3×3 spatial kernel of the first sub-filter, with 5 non-zero filter coefficients.

They compose the vector of design variables $x_1 \in R^3$. The sparse 7×7 kernel of the second sub-filter can be similarly parametrized by 25 coefficients composing the vector of design variables $x_2 \in R^{25}$. Fig. 5 presents the optimal filter design defined by a global minimizer in problem (6).

This simple example allows us to study the objective function of problem (6) in 2D. Note that, the set of all objective function values does not change if to add in (6) the extra constraint $\|x_1\| = 1$, because this vector normalization can be compensated by choosing the vector x_2 of a proper length. Note also that for any given x_1 , it is easy to find $x_2(x_1)$ that solves the linear least squares problem

$$\min_{x_2} \|b - D(x_1)A_2 x_2\|^2,$$

where $D(x_1) = \text{diag}(A_1 x_1)$. This allows us to present the objective function value in (6) as a function of x_1 only. Since $\|x_1\| = 1$, this function actually depends on two variables of the spherical coordinates. It is sufficient to restrict the spherical coordinates by the interval $[0, \pi]$, because any couple of anti-podal points x_1 and $-x_1$ on the sphere produces identical function values. All possible objective function values in (6) are presented in Fig. 5 by image intensity values in the spherical coordinates. One can recognize 5 local minima. The two open contours in the lower right and the upper left parts are produced by a couple of anti-podal points. Thus, the total number of local minima, including those produced by anti-podal couples, is 8. The aim of this sub-section was to show that even our toy example with very smooth filter functions can reveal such a large number of local minima. In general, it grows with l , m and n .

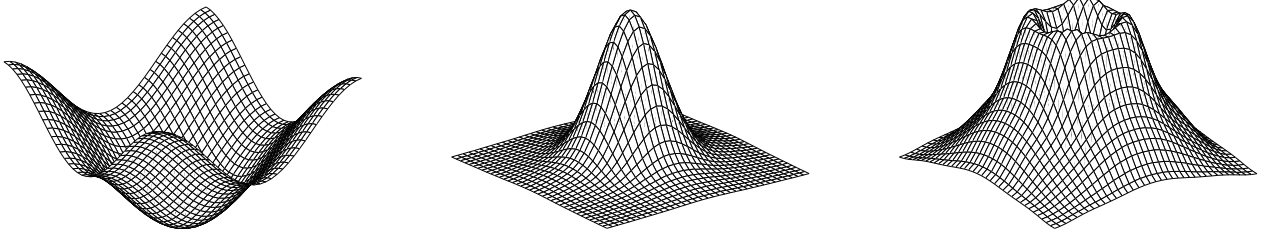


Fig. 5. The best approximation (**right**) to the lognormal band-pass filter as the product of the frequency responses of the first (**left**) and the second (**middle**) sub-filters.

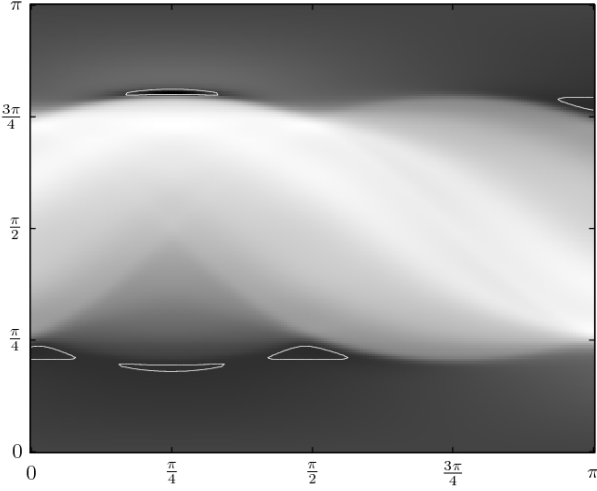


Fig. 6. Objective function values and local minima in the spherical coordinates for the sequence of two sub-filters.

IV. IDEAL SUB-FILTER FREQUENCY RESPONSES DESIGN

As a consequence of the multi-extremal nature of our problem, the success of local optimization methods much depends on their initialization. For robustness, it is desirable to find a good approximate solution to the MLLS problem which could then serve as an initial point for the local search. Our approach aims at providing such initial point.

We begin with reformulating problem (6) equivalently as

$$\begin{aligned} \min \quad & \|b - b_1 \circ \dots \circ b_l\|^2 \\ \text{s.t.} \quad & A_i x_i = b_i, \quad i = 1, \dots, l \\ & b_1, \dots, b_l \end{aligned}$$

where ‘s.t.’ stands for ‘subject to’. The characteristic features of this problem can be summarized in the following way:

$$b \approx b_1 \circ \dots \circ b_l, \quad A_i x_i = b_i, \quad i = 1, \dots, l.$$

For generating initial points, we consider a similar, conceptually close, problem in which $b = b_1 \circ \dots \circ b_l$, but $A_i x_i \approx b_i$ for all $i = 1, \dots, l$. We formulate it as

$$\begin{aligned} \min \quad & \sum_{i=1}^l \|b_i - A_i x_i\|^2 \\ \text{s.t.} \quad & b_1 \circ \dots \circ b_l = b \\ & x_1, \dots, x_l, \quad b_1, \dots, b_l \end{aligned} \quad (7)$$

It should be emphasized that the ideal frequency response b_i introduced for each sub-filter is a design variable. In this perspective, the problem formulation (7) can be interpreted as follows. The equality constraints guarantees an admissible factorization of the ideal frequency response b . As opposed to most other decompositions, our objective function takes into account the spatial extension of each sub-filter, which provides a strong bias in favor of factorizations that can be implemented on a spatial grid. The grid of each sub-filter is implicitly defined by the corresponding A_i .

Note that x_1, \dots, x_l are not present in the constraints. Therefore, problem (7) can easily be solved in these variables. For the sake of simplicity, we will consider the standard Euclidean norm in (7), although the subsequent reasoning holds also for the weighted Euclidean norm. The solution in x_1, \dots, x_l has the form

$$x_i = A_i^\dagger b_i, \quad i = 1, 2, \dots, l, \quad (8)$$

where † denotes the pseudo-inverse. This substitution reduces problem (7) to

$$\begin{aligned} \min \quad & \sum_{i=1}^l b_i^T P_i b_i \\ \text{s.t.} \quad & b_1 \circ \dots \circ b_l = b \end{aligned} \quad (9)$$

where $P_i = I_i - A_i A_i^\dagger$ is the matrix of orthogonal projection determined by A_i , and I_i is the unit matrix of the proper size.

In the numerical implementation, it may not be reasonable to compute the large dense matrix P_i explicitly, but instead, it can be treated as a linear operator defined by A_i in a standard way [19].

It is necessary to regularize problem (9), because it may admit trivial asymptotic solutions with the objective function value tending to zero. To show this, consider vectors $\beta_1, \beta_2, \dots, \beta_l$ of the same size as b_1, b_2, \dots, b_l . Let β_1 be such that $P_1 \beta_1 = 0$, and β_2, \dots, β_l have any values for which $\beta_1 \circ \dots \circ \beta_l = b$. Then the vectors

$$b_1(t) = \beta_1/t^{l-1}, \quad b_2(t) = t\beta_2, \quad \dots, \quad b_l(t) = t\beta_l$$

are feasible in problem (9) for any value of scalar $t \neq 0$, and $\sum_{i=1}^l b_i^T(t) P_i b_i(t) \rightarrow 0$ with $t \rightarrow 0$.

To regularize problem (9), we add to the objective function a quadratic penalty term as follows

$$\sum_{i=1}^l (b_i^T P_i b_i + \mu \|b_i\|^2),$$

where $\mu > 0$ is a small parameter. Denoting

$$\tilde{P}_i = P_i + \mu I_i, \quad (10)$$

we rewrite the regularized problem as

$$\min_{b_1, \dots, b_l} \sum_1^l b_i^T \tilde{P}_i b_i \quad \text{s.t.} \quad b_1 \circ \dots \circ b_l = b \quad (11)$$

Observe that the objective function in (11) is strictly convex with a unique minimizer in the origin. Unlike (6), this problem does not suffer from the ill property of having non-isolated minimizers. However, any simultaneous change of signs to the opposed for all components in b_i 's such that $\text{sign}(b_1 \circ \dots \circ b_l) = \text{sign}(b)$ still produces equivalent solutions.

An example of objective function and feasible set is presented by Fig. 7 (left), which for the sake of simplicity are drawn for a sub-filter sequence of length 2 and only for one component of b_1 and b_2 . Since the objective function depends also on all the other components of b_1 and b_2 , whose values are fixed in this example, the objective function minimizer is not located in the origin. One can also see that the feasible set of points is disjoint, and it is located in the upper right and the lower left quadrants. For comparison, the feasible set for a sub-filter sequence of length 3 is also shown (see Fig. 7, right), but only for the quadrant where $b_i > 0$.

Without loss of generality, we can assume that $b \geq 0$ in (6) and (11). Indeed, if any component of b is negative, the change of its sign to positive can be compensated by the change of sign in the corresponding row of A_1 . In sub-section VI-B we discuss how to treat the case of zero components. Assume, for the moment, that $b > 0$ and that the optimal values of b_1, \dots, b_l are also positive. The problem (11) is then equivalent to

$$\min_{b_1 > 0, \dots, b_l > 0} \sum_1^l b_i^T \tilde{P}_i b_i \quad \text{s.t.} \quad b_1 \circ \dots \circ b_l = b \quad (12)$$

Substituting $b_i = \exp(y_i)$, we reduce this problem, as illustrated in Fig. 7, to the following linear equality constrained problem:

$$\min_{y_1, \dots, y_l} \sum_1^l \exp(y_i)^T \tilde{P}_i \exp(y_i) \quad \text{s.t.} \quad \sum_1^l y_i = \ln b \quad (13)$$

where $\exp(\cdot)$, $\ln(\cdot)$ are component-wise operations. It looks nicely shaped in Fig. 7 and it can be efficiently solved by the conventional methods [20], which are able to take advantage of using the easily available derivatives of the objective function and the simple structure of the linear constraints in (13). In our calculations the computational time required for solving this problem was about half of the time of one run of the ALS algorithm on problem (6).

The considered problem with positive signs (12), and its equivalent problem (13), will serve as a subproblem in the next section, where the variables b_1, \dots, b_l are allowed to have any combination of signs.

V. BINARY PROBLEM

It is obvious that the optimal values of b_1, \dots, b_l in (11) are not necessarily positive. To release the posed sign restrictions, we divide the problem into an outer binary problem to deal

with the signs and an inner subproblem, similar to (13). We continue assuming that $b > 0$, which implies that the feasible vectors b_1, \dots, b_l in (11) have no zero components. Let us present problem (11) as a specially enumerated set of subproblems of the form (12). We will use the following notations:

$$s_i = \text{sign}(b_i), \quad \bar{b}_i = s_i \circ b_i, \quad \bar{P}_i = \text{diag}(s_i) \tilde{P}_i \text{diag}(s_i). \quad (14)$$

Let S denote the set of vectors in R^m whose elements are equal to ± 1 , which means that $s_i \in S$. Clearly, $\bar{b}_i > 0$ for all $i = 1, \dots, l$. Furthermore, for all feasible vectors b_1, \dots, b_l in (11) we have $s_1 \circ \dots \circ s_l = e$, where $e = (1, \dots, 1)^T \in R^m$. It is easy to verify that problem (11) is equivalent to

$$\min_{s_1 \in S, \dots, s_l \in S} \phi(s_1, \dots, s_l) \quad \text{s.t.} \quad s_1 \circ \dots \circ s_l = e \quad (15)$$

where the objective function

$$\phi(\cdot) = \min_{\bar{b}_1 > 0, \dots, \bar{b}_l > 0} \sum_1^l \bar{b}_i^T \bar{P}_i \bar{b}_i \quad \text{s.t.} \quad \bar{b}_1 \circ \dots \circ \bar{b}_l = b \quad (16)$$

Here \bar{P}_i depends on s_i in accordance with (14). Note that the substitution $s_l = s_1 \circ \dots \circ s_{l-1}$ is able to eliminate the equality constraint in outer problem (15), which is a binary problem with $2^{m(l-1)}$ feasible points. This number of feasible points defines the number of all inner problems (16).

Since (16) is of the same form as (12), we can apply, like previously, the substitution $\bar{b}_i = \exp(y_i)$ to reduce the inner problem to the following one

$$\phi(\cdot) = \min_{y_1, \dots, y_l} \sum_1^l \exp(y_i)^T \bar{P}_i \exp(y_i) \quad \text{s.t.} \quad \sum_1^l y_i = \ln(b) \quad (17)$$

The only difference between this problem and (13) is that

$$\bar{P}_i = \text{diag}(s_i) P_i \text{diag}(s_i) + \mu I_i, \quad (18)$$

obtained by substituting (10) in (14), is used here instead of \tilde{P}_i . Comparing (10) and (18), where $\text{diag}(s_i) P_i \text{diag}(s_i)$ and P_i are projection matrices, we conclude that problem (17) is as easy to solve as problem (13).

The important feature of problem (15) is that it performs a partitioning of the feasible set in (11) and reduces this problem to a finite number of easy-to-solve inner problems (16), or equivalently (17). This allows us to capture the nature of the local minimizers of problem (11) and to efficiently enumerate them by combining the signs. Any optimal, or close to optimal, solution to problem (15) can be used as an initial point in problem (6), to be further refined by local search algorithms, like ALS. The success of this approach is illustrated by the results of our numerical experiments presented in the next section.

It should be emphasized that binary problems are, in general, difficult to solve, but fortunately, in our case, the nature of signs in the sub-filter outputs are often well understood. Prior knowledge of the filter characteristics and its structure helps to facilitate substantially the solution process of the outer problem by focusing on a relatively small number of combinations of signs. We discuss this issue in sub-section VI-A.

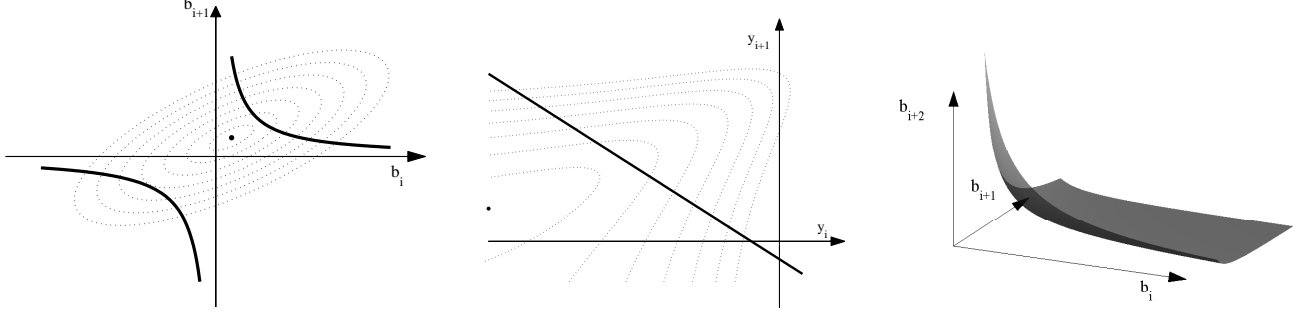


Fig. 7. **Left:** The objective function (dashed lines), the feasible set of points (black lines) and the unconstrained minimizer (black dot) for a single frequency component of a sub-filter sequence of length 2. **Mid:** The first quadrant of the left figure for the new variables. **Right:** The feasible set of points for one quadrant for a sub-filter sequence of length 3.

We developed a global search strategy for moving from one combination of signs to another. The value of $\phi(s_1, \dots, s_l)$ is decreasing in this process. Details of this strategy will be presented in a separate paper. For the filters considered in the next section, we compared two approximate solutions to problem (15). One was based on the combination of signs described in sub-section VI-A, and the other one was the improvement of this combination produced by our global search strategy. The improvement in terms of $\phi(s_1, \dots, s_l)$ was moderate, and it did not result in any noticeable improvement in solving problem (6). This favors our choice of signs based on prior knowledge of filter characteristics and its structure.

VI. IMPLEMENTATION AND EXPERIMENTS

For the experiments, we use the set of filters considered in [21]. These filters are spherically separable, which is a basic requirement for the applications discussed in Section I. In the spherical coordinates $u = \rho \hat{u}$, the filters can be written as the product:

$$F(u) = R(\rho)D(\hat{u}), \quad (19)$$

where the frequency selective part $R(\rho)$ is a radially symmetric function and the orientation selective part $D(\hat{u})$ is an angular function. The radial function is a band-pass filter sensitive to frequencies $\rho \in [\rho_l, \rho_u]$ and is given by

$$R(\rho) = \frac{1}{2}(\text{erf}(s \ln(\frac{\rho}{\rho_l})) - \text{erf}(s \ln(\frac{\rho}{\rho_u}))), \quad (20)$$

where the *erf*-function is defined as follows:

$$\text{erf}(t) = \frac{2}{\sqrt{\pi}} \int_0^t \exp(-\tau^2) d\tau. \quad (21)$$

By letting ρ_l tend to zero we obtain a low-pass filter. A high-pass filter is in a similar way obtained by letting ρ_u tend to infinity. The angular function is the inner product

$$D(\hat{u}) = \langle \hat{u}, \hat{v} \rangle^d \quad (22)$$

and is said to be of order d with main direction v . Observe that filters of odd order have both positive and negative frequency components and that the filter origin $R(0) = 0$ for all but the low-pass filters.

A. Prior Knowledge for Use in the Binary Problem

As pointed out in section V, prior knowledge of the filters characteristics can reduce substantially the number of inner problems (16) to be solved, which is actually the number of combinations of signs to be considered in outer problem (15). In general, filters that benefit from sub-filter decomposition have smooth frequency responses which limit the number of possible sign changes. The smoothness is assured by filters of spatially small size, a requirement brought on by the non-stationary nature of multidimensional signals. In addition, a prerequisite for convolution to be efficient is to have relatively small spatial filter. Even though our approach is presented here for a specific class of filters, it can be easily extended to a much larger class of filters with smooth frequency responses.

We will demonstrate how prior knowledge can be used in the case of spherically separable filters. For this purpose, we rewrite (19) equivalently as

$$F(u) = R(\rho) \langle \hat{u}, \hat{v} \rangle^d = \tilde{R}(\rho) \langle u, \hat{v} \rangle^d. \quad (23)$$

This reformulation presents a pre-filtered directional derivative with the pre-filter $\tilde{R}(\rho) = \rho^{-d} R(\rho)$ which, by definition, is non-negative. The angular function $\langle u, \hat{v} \rangle^d$ specifying the directional derivative can be efficiently approximated by central differences. The zero crossings of the central difference approximation are well defined and in this case coincide with the zero-crossings of the desired frequency response. Therefore, it is natural to choose b_j for one sub-filter so that it is a central difference approximation which implies that $\text{sign}(b_j) = \text{sign}(b)$. Then since the product of the remaining sub-filters must be positive, there is little or no reason to believe that any sub-filter except for b_j contain negative components. Thus, it is sufficient to solve inner problems (16) for only one combination of signs, in which $b_i > 0$ for all sub-filters except for $i = j$.

B. Zero Components of Ideal Frequency Response

Our approach presented in the previous sections is based on the assumption that b has no zero components. It allows us to apply the transformation of variables $b_i = \exp(y_i)$. In this

sub-section, we present two ways of reducing zero-component cases to non-zero cases.

A straightforward approach is to replace any zero components of b with a small value. The sign of this small value is not critical since the approximation error is prone to dominate over the small value. Then our solution to problem (11) for the slightly modified b should not differ too much from the solution corresponding to the original value of b . Once the perturbed solution has produced an initial point for solving problem (6), its further refinement with the use of ALS has in most interesting cases a negligible difference with the refinement obtained by ALS for the non-perturbed initial point.

An exception is, however, when it is important to produce a sub-filter sequence with some components exactly equal to zero. Due to the smoothness requirements implicitly imposed by the individual properties of each sub-filter this can only be accomplished for a very small number of components at the expense of a less close fit for the other components. Band-pass and high-pass filters are such exceptions, where it is important that the filters are insensitive to the average intensity of the signal. Such filters are characterized by a zero component of the frequency response b which corresponds to $R(0)$. To make the filter output as close as possible to zero in this component, it is necessary to choose a very large weight value corresponding to the same component, but this impairs significantly the efficiency of the refinement process. Our alternative suggestion is to make one of the sub-filters, say the k -th one, to produce the zero output in the required component. If to assume that the j -th component of b is equal to zero, then any vector x_k orthogonal to the j -th row of A_k ensures the required zero output. This can be implemented by modifying the matrix A_k as follows

$$A_k \left(I - \frac{a_k^T a_k}{\|a_k\|^2} \right),$$

where the row vector a_k denotes the j -th row of A_k . Note that the j -th row of the modified A_k is zero. Then we remove the j -th row in each A_i including the modified A_k . Similarly we remove the j -th element in b that was equal to zero. This allows us to apply the approach proposed in Sections IV and V, because the reduced b has no zero components. Then the produced initial point can be further refined by a local search algorithm, like ALS, to be applied to the reduced problem (6). Since for the x_k in the obtained solution it is not guaranteed that the j -th component of $A_k x_k$ is equal to zero for the original A_k , it is important to use

$$\left(I - \frac{a_k^T a_k}{\|a_k\|^2} \right) x_k = x_k - \frac{(a_k x_k)}{\|a_k\|^2} a_k^T$$

instead of x_k as the vector of coefficients for sub-filter k .

Inasmuch as there is a freedom in choosing one of the sub-filters for producing the zero output in the required component, it makes sense to try one-by-one with a few sub-filters or even all, if their number is not too large, and then to choose the best of the obtained results. The approach presented in this sub-section can be obviously extended to the case of a few zero components in b .

C. Numerical Results

The aim of this sub-section is to demonstrate the efficiency and robustness of our approach based on solving problem (15). Its solution is transformed by formula (8) into a point in problem (6) used for initializing the ALS algorithm.

In our numerical experiments, ALS is based on alternating the increasing sequence $1, 2, \dots, n$ followed by the decreasing one $n, n-1, \dots, 1$ and so forth. We use samples from the class of filters presented by (19). In Fig. 8, three 2D filters with different characteristics are shown. Each of them were implemented as a sequence of 10 sub-filters. The first 9 sub-filters comprise in total 49 nonzero filter coefficients (sparsely scattered). The last of these sub-filters serves for compensating zero and negative components of b as described in the previous sub-sections. This sub-filter contains 9, 2 or 3 non-zero entries respectively for zero, first and second order filters. Thus, the total number of nonzero filter coefficients ranges within the interval 51 – 58.

For comparison, a single dense filter of equal quality in terms of approximation error was designed. Its spatial size was required to be 13^2 , i.e. three times more multiplications are to be performed compared to the chosen sub-filter sequence. In 3D this would correspond to 13^3 , while the sub-filter sequence would have approximately 90 non-zero coefficients. Thus, the number of floating point operations can be reduced in this case by a factor of 24.

The main focus of this paper is however not on how efficiently the filter features can be implemented, but rather on how our approach can improve the design process related to the choice of the optimal values of the sub-filter coefficients.

In our numerical experiments, we used

$$\begin{aligned} b_i &= (\text{sign}(b) \circ b)^{\frac{1}{2}}, \quad i = 1, \dots, 9 \\ b_i &= \text{sign}(b) \circ (\text{sign}(b) \circ b)^{\frac{1}{2}}, \quad i = 10 \end{aligned} \quad (24)$$

as an initial point for problem (15). Here the signs are chosen in accordance with Sub-section VI-A. Then following our approach, we applied formula (8) to the obtained solution in order to produce an initial point for the ALS algorithm. As an alternative initial point we used the one calculated directly from (24) by formula (8) without solving problem (15). We call it *deterministic* initial point to underline the difference with the other 500 randomly generated initial points.

In Fig. 9 the progress of ALS is shown as a function of the iteration number. One can see that the results much depend on the initial points. Those presented for the randomly generated initial points clearly indicate that the number of local minimizers is large. Important is that for most of the randomly generated initial points, the ALS algorithm was not able to produce satisfactory approximation, while the initial point obtained on the base of our approach was always among the most successful. This means that our approach allows us to decrease substantially the number of ALS runs. Moreover, it is much more robust in comparison with the standard approach, according to which initial points are generated randomly without any guarantee of success.

Our experiments were made for the considered class of filters on a computer equipped with the processor 2.4GHz

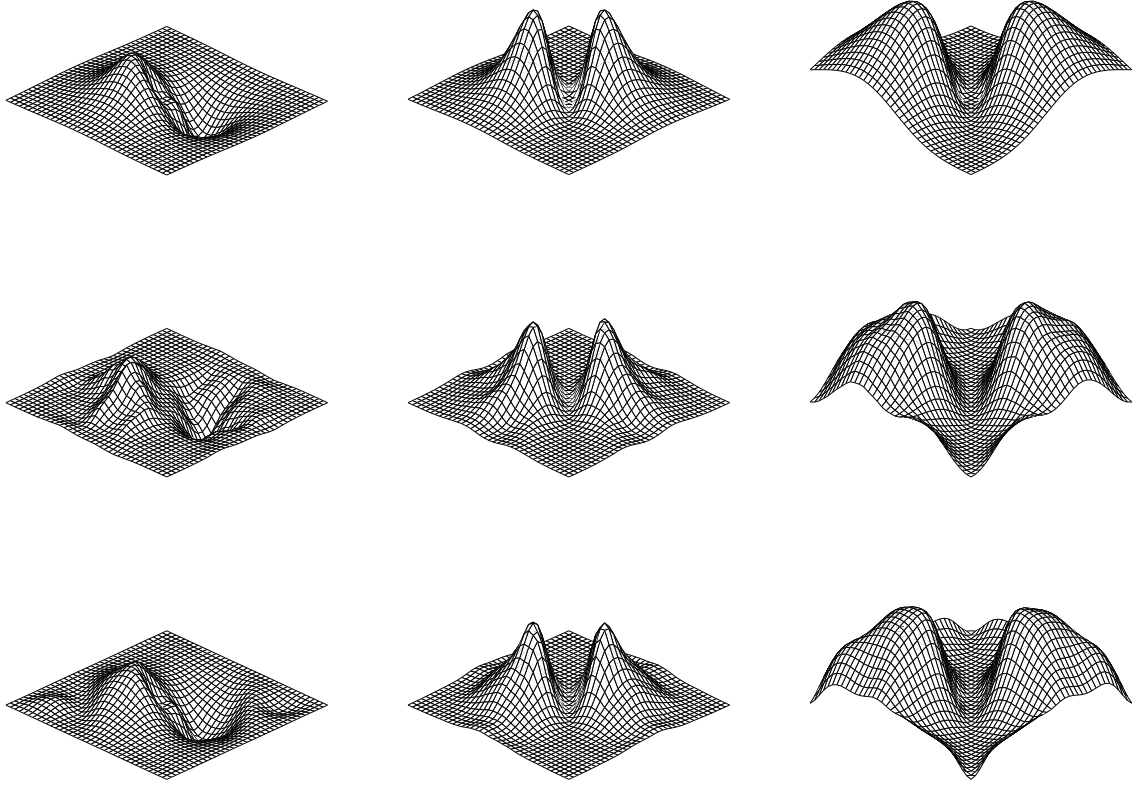


Fig. 8. The results for a first order band-pass filter, a second order band-pass filter and a second order high-pass filter are here shown, all of them with an angular selectivity with $v = \frac{1}{\sqrt{2}}[1, 1]^T$. (**Top row:**) Desired frequency responses. (**Mid row:**) Results from ALS after 150 iterations initiated with a deterministic initial point. (**Bottom row:**) Results from ALS after 150 iterations initiated by the solution to (15).

AMD opteron 250. The Matlab routine *fmincon* was used for solving (17). The results are summarized in Table I, where t_0 is the CPU time required to solve (17) and to generate an initial point. Note that one run (150 iterations) of the ALS algorithm takes, on the average, 6.2s which is approximately twice of the time t_0 . The error ε_0 is the result of running ALS from the initial point generated by our approach. We compare it with the errors ε_j obtained by ALS for 500 random initial points.

TABLE I
CPU TIME, SQUARED ERROR AND PERCENTAGE OF SUCCESSFUL
SOLUTIONS FOR 500 RANDOM INITIAL POINTS

Filter	t_0	ε_0^2	$\min(\varepsilon_j^2)$	$\frac{\text{card}(\varepsilon_j \leq \varepsilon_0)}{\text{card}(\varepsilon_j)}$
Low-Pass	3.3 s	1.28e-4	3.18e-5	0.6 %
Band-Pass (BP)	3.0 s	7.36e-6	5.65e-6	23.3 %
High-Pass (HP)	2.3 s	3.08e-4	2.95e-4	0.4 %
BP, 1:st order	3.1 s	4.37e-6	1.19e-6	5.6 %
HP, 1:st order	3.6 s	6.02e-4	3.91e-4	6.3 %
BP, 2:nd order	2.9 s	8.68e-6	5.94e-6	3.2 %
HP, 2:nd order	3.3 s	3.06e-4	2.77e-4	3.2 %

Observe that there is no significant difference between the smallest error $\min(\varepsilon_j)$ and ε_0 , and that only a small fraction of

these 500 solutions are slightly better than the solution found by our approach.

The band-pass filter is the only exception in which this fraction is not small. But for the considered sub-filter sequence, the band-pass filter is an easy design problem with a relatively small number of local minima. One can see that for the best of the found solutions and those 25% which are close to it, their errors differ just a little from the one obtained by our approach. Therefore, even in this case, where the choice of initial point for the ALS is not a very critical parameter, it is reasonable to incorporate our approach, because it takes just a little computational time and it substantially improves the robustness.

The most important observation is that, using our approach, it was possible to design sub-filter sequences for low-pass, band-pass and high-pass filters of approximately the same quality as with the use of the standard approach based on 500 runs of ALS, but more than 300 times faster.

VII. DISCUSSION

In this paper, we introduced an approach to efficiently solving the multi-linear least squares problem originating from the design of multi-dimensional sub-filter sequences. At

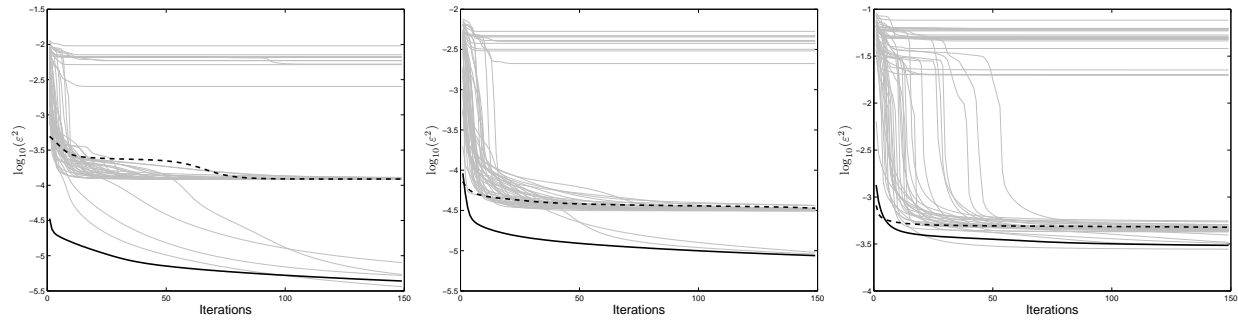


Fig. 9. Progress in approximation error as a function of iteration number of the ALS algorithm for various initial points. Solid black line: the initial point provided by our approach. Gray lines: 50 randomly generated initial points. Dashed line: the deterministic initial point. **Left:** A band-pass filter of order 1. **Mid:** A band-pass filter of order 2. **Right:** A high-pass filter of order 2.

present, this approach is directly applicable to filters with all frequency components of the same sign or close to zero. This class of filters includes, for instance, low-pass, band-pass and high-pass filters. It was also discussed how to incorporate prior knowledge into the design of sub-filter sequences for the filters, whose frequency components are allowed to be of different signs. The experiments reveal the multi-extremal nature of the MLLS problem which motivates the necessity of applying a robust initialization. They demonstrate also the robustness of our approach in generating initial points for the considered class of filters, as well as its extremely high speedup factor.

The master binary problem (15) is difficult for other type of filters as well and the requirements on the problem to be solved efficiently still apply. One should take into account that in practice most filters have symmetries that can be exploited by analogy with what was proposed in this paper. It looks reasonable to develop a methodology tailored properly to each specific class of filters to be designed.

Further extension of this approach to more general structure of sub-filter networks would be an attractive perspective. Generalization of the presented approach to tree-structured networks is straightforward since we simply can consider the tree as a finite number of sequences. Further generalization to directed acyclic graphs is however non-trivial.

REFERENCES

- [1] M. Hemmendorff, "Motion estimation and compensation in medical imaging," Ph.D. dissertation, Linköping University, Sweden, Linköping, Sweden, 2001.
- [2] R. San Jose Estepar, "Local structure tensor for multidimensional signal processing. applications to medical image analysis," Ph.D. dissertation, University of Valladolid, Spain, 2005.
- [3] C.-F. Westin, L. Wigström, T. Looch, L. Sjöqvist, R. Kikinis, and H. Knutsson, "Three-dimensional adaptive filtering in magnetic resonance angiography," *Journal of Magnetic Resonance Imaging (JMRI)*, vol. 14, pp. 63–71, 2001.
- [4] T. Zhang and G. Golub, "Rank-one approximation to high order tensors," *Matrix Analysis and Applications*, vol. 23, pp. 534–550, 2001.
- [5] T. Saramäki and A. Fam, "Subfilter approach for designing efficient FIR filters," in *Proceedings IEEE International Symposium on Circuits and Systems*, June 1988, pp. 2903–2915.
- [6] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Wellesley-Cambridge Press, 1996.
- [7] S. Mallat, *A Wavelet Tour of Signal Processing*. Academic Press, 1999.
- [8] J. Weickert, *Anisotropic Diffusion in Image Processing*. Teubner-Verlag, 1998.
- [9] L. Haglund, "Adaptive multidimensional filtering," Ph.D. dissertation, Linköping University, Sweden, SE-581 83 Linköping, Sweden, October 1992, dissertation No 284, ISBN 91-7870-988-1.
- [10] B. Svensson, M. Andersson, Ö. Smedby, and H. Knutsson, "Efficient 3-d adaptive filtering for medical image enhancement," in *Proceedings of the IEEE International Symposium on Biomedical Imaging*. Arlington, USA: IEEE, April 2006.
- [11] B. Svensson, "A multidimensional filtering framework with applications to local structure analysis and image enhancement," Ph.D. dissertation, Linköping University, Sweden, April 2008.
- [12] J. Tropp and S. Wright, "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE*, vol. 98, no. 6, pp. 948–958, 2010.
- [13] P. Paatero and U. Tapper, "Least squares formulation of robust nonnegative factor analysis," *Chemomet. Intell. Lab. Systems*, vol. 37, pp. 23–35, 1997.
- [14] S. Leurgans and R. T. Ross, "Multilinear models: Applications in spectroscopy (with discussion)," *Statistical Science*, vol. 7, pp. 289–319, 1992.
- [15] J.-H. Wang, P. Hopke, T. Hanciewicz, and S. Zhang, "Application of modified alternating least squares regression to spectroscopic image analysis," *Analytica Chim. Acta*, vol. 476, pp. 93–109, 2003.
- [16] M. Andersson, J. Wiklund, and H. Knutsson, "Filter networks," in *Proceedings of Signal and Image Processing (SIP'99)*. Nassau, Bahamas: IASTED, October 1999.
- [17] H. Knutsson, M. Andersson, and J. Wiklund, "Advanced filter design," in *Proceedings of the 11th Scandinavian Conference on Image Analysis*. Greenland: SCIA, June 1999.
- [18] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Digital Signal Processing*. Prentice Hall Professional Technical Reference, 1990.
- [19] Å. Björck, *Numerical Methods for Least Squares Problems*. SIAM, Society for Industrial and Applied Mathematics, 1996.
- [20] J. Nocedal and S. Wright, *Numerical Optimization*. Springer-Verlag, 1999.
- [21] H. Knutsson and M. Andersson, "Implications of invariance and uncertainty for local structure analysis filter sets," *Signal Processing: Image Communications*, vol. 20, no. 6, pp. 569–581, July 2005.