

Orthogonal Matched Wavelets with Vanishing Moments: A Sparsity Design Approach

Joël Karel¹  · Ralf Peeters¹ 

Received: 3 November 2016 / Revised: 7 November 2017 / Accepted: 8 November 2017 /
Published online: 18 November 2017
© The Author(s) 2017. This article is an open access publication

Abstract This paper presents a novel approach to design orthogonal wavelets matched to a signal with compact support and vanishing moments. It provides a systematic and versatile framework for matching an orthogonal wavelet to a specific signal or application. The central idea is to select a wavelet by optimizing a criterion which promotes sparsity of the wavelet representation of a prototype signal. Optimization is performed over the space of orthogonal wavelet functions with compact support, coming from filter banks with a given order. Parametrizations of this space are accompanied by explicit conditions for extra vanishing moments, to be used for constrained optimization. The approach is developed first for critically sampled wavelet transforms, then for undecimated wavelet transforms. Two different optimization criteria are presented: to achieve sparsity by L^1 -minimization or by L^4 -norm maximization. Masking can be employed by introducing weighting factors. Examples are given to demonstrate and evaluate the approach.

Keywords Orthogonal wavelets · Sparsity · Wavelet design · Matched wavelet · Polyphase

✉ Joël Karel
joel.karel@maastrichtuniversity.nl
Ralf Peeters
ralf.peeters@maastrichtuniversity.nl

¹ Department of Data Science and Knowledge Engineering, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands

1 Introduction

Wavelets have now been around for several decades and have well proven their usefulness in many areas of signal processing [20,30]. Unlike the Fourier transform, which exclusively employs harmonic functions, the wavelet framework allows for a *choice of wave form*. In practical applications, it is often not so clear how to make this choice [28]. Sometimes a choice is driven by theoretical considerations of a general nature, such as the wavelet being orthogonal and having as many vanishing moments as possible [7]. Other aspects may include the linear phase property of filters or symmetry of the wave form [23]. In practice, researchers may also first test a variety of wavelets and then proceed to use the few that initially provide the most promising results.

How to move away from the standard wavelet families and toolboxes and to take greater advantage of the great freedom there is in choosing or constructing a wavelet *for an application at hand*, is far from trivial, but some work has been done. In [6], the amplitude and phase spectrum of a wavelet are separately matched to that of a reference signal in the Fourier domain. In [27], a parameterized Klauder wavelet, resembling a QRS complex in a cardiac signal, is tuned to a measured signal. In [12], a design procedure for bi- and semi-orthogonal wavelets is discussed where the optimization criterion aims to maximize the energy in the approximation coefficients. In [21], a parameterization for orthogonal wavelets is used to design wavelets of which either the wavelet or the scaling function resembles a prototype signal. It has the resulting effect that wavelet analysis parallels template matching. How to enforce additional vanishing moments, and thus smoothness, in such approaches, is an issue having received little attention so far. Another approach is the lifting scheme [31]. This works in polyphase like the current paper, but the parameterization is different, split into an updating and a prediction step, where designing the latter is most challenging. In [2], biorthogonal wavelets in the lifting scheme are matched to a signal where attention is given to linear phase, but not to vanishing moments.

In the current paper, an approach is discussed to design orthogonal wavelets, matched to a prototype signal from a given application. We work with polyphase filters, and the orthogonal wavelets are required to have compact support and a user-specified number of built-in vanishing moments. Two design criteria are developed which promote sparsity and allow the wavelet to be matched to a prototype signal for the application at hand: one by minimizing an L^1 -norm, the other by maximizing an L^4 -norm. Sparsity is naturally motivated by compression and detection applications. The two criteria achieve sparsity by a shared underlying principle of maximization of variance. The method is first developed for the critically sampled wavelet transform, but then extended to work with the undecimated wavelet transform too. Additionally, masking can be employed by introducing weighting factors. To build in orthogonality, compact support, and a few vanishing moments, a suitable parameterization is required. Here we provide a parameterization which has orthogonality and compact support built in, and for which the vanishing moment conditions take a fairly easy form as long as the filter order is not too high. This allows one to handle them conveniently with standard constrained optimization techniques.

The paper builds on preliminary research previously reported in [15,16] and is organized as follows. First, in Sect. 2, we introduce wavelets from filter banks, the polyphase

representation, orthogonal wavelets, and their connection to lossless polyphase filters. Two (closely related) parameterizations of FIR lossless filters are at the core of the wavelet design procedure and presented in Sect. 3. In Sect. 4, we express the vanishing moment conditions in terms of these parameterizations. Suitable criteria for measuring how well a certain orthogonal wavelet matches an application at hand are developed in Sect. 5, followed by a discussion of practical optimization considerations in Sect. 6. Several wavelet matching examples are presented in Sect. 7. Conclusions regarding the matching design procedure are discussed in Sect. 8. All proofs are collected in “Appendix.”

2 Filter Banks, Polyphase Representation, and Orthogonal Wavelets

2.1 Filter Banks for Signal Processing

An important approach to wavelets employs filter banks for digital signal processing. See [30], on which much of the material in this section is based. A filter bank consists of two discrete-time filters: a low-pass filter $C(z) = \sum_{k \in \mathbb{Z}} c_k z^{-k}$ and a high-pass filter $D(z) = \sum_{k \in \mathbb{Z}} d_k z^{-k}$, where z is the complex variable of the z -transform [30, p.5]. (This relates to the Fourier transform through $z = e^{i\omega}$, for a frequency ω .) They are used to process a given discrete-time signal $s = \{s_k\}_{k \in \mathbb{Z}}$ by first passing it through both filters separately, and then applying dyadic downsampling $\downarrow 2$ (i.e., the odd-indexed values are discarded and the remaining subsequences are relabeled). This produces a level-1 approximation signal $a = \{a_k\}_{k \in \mathbb{Z}}$ and a level-1 detail signal $b = \{b_k\}_{k \in \mathbb{Z}}$, given by

$$a_k = \sum_{\ell \in \mathbb{Z}} c_\ell s_{2k-\ell}, \quad b_k = \sum_{\ell \in \mathbb{Z}} d_\ell s_{2k-\ell}. \quad (1)$$

The same procedure is then reapplied to the approximation signal a to obtain level-2 approximation and detail signals. Repeating the process, one obtains approximation and detail signals up to a user-selected level j_{\max} . The approximations and details become coarser with each new iteration step. Note that in practical applications with a signal s of finite length, the approximation and detail signals a and b resulting from one iteration step are roughly only half as long as s , which induces a natural upper bound on j_{\max} .

2.2 Polyphase Filtering

It is well known (see, e.g., [30, Chapter 4]) that the same level-1 signals a and b can alternatively be obtained: by first splitting the given signal s into its two phases $s_{\text{even}} = \{s_{2k}\}_{k \in \mathbb{Z}}$ and $s_{\text{odd}} = \{s_{2k+1}\}_{k \in \mathbb{Z}}$, and then jointly passing s_{even} and the delayed odd phase $\{s_{2k-1}\}_{k \in \mathbb{Z}}$ through a 2×2 polyphase filter $H(z)$. It is defined by

$$H(z) = \begin{pmatrix} C_{\text{even}}(z) & C_{\text{odd}}(z) \\ D_{\text{even}}(z) & D_{\text{odd}}(z) \end{pmatrix}, \quad (2)$$

with

$$\begin{aligned} C_{\text{even}}(z) &= \sum_{k \in \mathbb{Z}} c_{2k} z^{-k}, \\ C_{\text{odd}}(z) &= \sum_{k \in \mathbb{Z}} c_{2k+1} z^{-k}, \\ D_{\text{even}}(z) &= \sum_{k \in \mathbb{Z}} d_{2k} z^{-k}, \\ D_{\text{odd}}(z) &= \sum_{k \in \mathbb{Z}} d_{2k+1} z^{-k}. \end{aligned}$$

To be explicit, it then holds that

$$\begin{pmatrix} a_k \\ b_k \end{pmatrix} = \sum_{\ell \in \mathbb{Z}} \begin{pmatrix} c_{2\ell} & c_{2\ell+1} \\ d_{2\ell} & d_{2\ell+1} \end{pmatrix} \begin{pmatrix} s_{2k-2\ell} \\ s_{2k-2\ell-1} \end{pmatrix}. \quad (3)$$

For the z -transforms, we equivalently have

$$\begin{pmatrix} A(z) \\ B(z) \end{pmatrix} = H(z) \begin{pmatrix} S_{\text{even}}(z) \\ z^{-1} S_{\text{odd}}(z) \end{pmatrix}, \quad (4)$$

where $A(z)$, $B(z)$, $S_{\text{even}}(z)$, and $S_{\text{odd}}(z)$ denote the z -transforms of a , b , s_{even} , and s_{odd} , respectively. Note also that the filters $C(z)$ and $D(z)$ satisfy $C(z) = C_{\text{even}}(z^2) + z^{-1} C_{\text{odd}}(z^2)$ and $D(z) = D_{\text{even}}(z^2) + z^{-1} D_{\text{odd}}(z^2)$ and are obtained by:

$$\begin{pmatrix} C(z) \\ D(z) \end{pmatrix} = H(z^2) \begin{pmatrix} 1 \\ z^{-1} \end{pmatrix}, \quad (5)$$

see [30]. This will be used to explore vanishing moment conditions in terms of polyphase filters in Sect. 4.

2.3 Perfect Reconstruction, Power Complementarity and Orthogonal Wavelets

A polyphase filter $H(z)$ allows for perfect reconstruction of every s if and only if $H(z)^{-1}$ exists. When the filters $C(z)$ and $D(z)$ are power complementary, see [33] and Eq. (8), then this holds and the inverse is particularly easy to compute. Such power complementary filters are associated with orthogonal decompositions generated by the filter bank and with orthogonal wavelet representations of the signals [20, 30].

The dilation equation and the wavelet equation serve to explore this connection. The *dilation equation* is a two-scale functional equation (a refinement equation), given

here by¹

$$\phi(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} c_k \phi(2t + k), \quad (6)$$

for which we assume a non-trivial solution $\phi(t)$ to exist which is square integrable, which is normalized with a unit L^2 -norm, and which has a nonzero integral.² The function $\phi(t)$ is called the scaling function, and it is used to build signal approximations. The wavelet function $\psi(t)$ is used to represent signal details on different scales and is defined by the *wavelet equation*:

$$\psi(t) = \sqrt{2} \sum_{k \in \mathbb{Z}} d_k \phi(2t + k). \quad (7)$$

When the filters $C(z)$ and $D(z)$ are power complementary, that is:

$$C(z)C(z^{-1}) + D(z)D(z^{-1}) = 2, \quad (8)$$

it follows that the set $\{\phi(t - m) \mid m \in \mathbb{Z}\}$ forms an orthonormal basis for the function space V_0 it spans [30]. Likewise, the set $\{\psi(t - m) \mid m \in \mathbb{Z}\}$ gives an orthonormal basis for its span W_0 , while V_0 is orthogonal to W_0 . The space $V_{-1} = V_0 \oplus W_0$ has the induced orthonormal basis $\{\sqrt{2}\phi(2t - m) \mid m \in \mathbb{Z}\}$ as a consequence of the dilation and wavelet equations (6) and (7). Repeating the procedure, a useful multi-resolution structure $\dots \supset V_{-1} \supset V_0 \supset V_1 \supset \dots$ of nested subspaces of $L^2(\mathbb{R})$ is generated; see again [20, 30].

To characterize orthogonality, it is helpful to observe that the filters $C(z)$ and $D(z)$ are power complementary if and only if the polyphase filter $H(z)$ is an all-pass filter, i.e., if $H(z)$ is unitary for all z on the complex unit circle. In digital signal processing applications, filters are often causal, stable and of finite order. Causal stable all-pass filters are known as lossless filters. They have been studied extensively in the literature, see, e.g., [33] and can be conveniently parameterized, see Sect. 3. This gives a framework to carry out orthogonal wavelet design by optimizing a criterion function over a parameterized class of lossless polyphase filters. Incorporating orthogonality of wavelets in this way saves us from imposing explicit nonlinear algebraic constraints—which otherwise tend to significantly deteriorate the performance of numerical optimization routines.

2.4 Causality, FIR Filters, Compact Support, and Wavelet Admissibility

A linear time-invariant filter $F(z)$ is causal if it can be written as $F(z) = \sum_{k \in \mathbb{Z}} f_k z^{-k}$ with $f_k = 0$ for all $k < 0$. When processing a signal, such a filter only uses values from the present and past to compute the current filter output. To capture local features of a

¹ Often the dilation equation expresses $\phi(t)$ in terms of functions $\phi(2t - k)$ instead. The current setup minimizes notation and avoids the use of another filter with reversed coefficients.

² In the literature necessary and sufficient conditions for this are given in terms of the joint spectral radius. See, e.g., [8, 14].

signal, wavelets are required to be well located in both time and frequency, meaning that they should effectively have compact support and a bounded bandwidth. Here the choice is made for true compact support, for which the discrete-time filters must have a finite impulse response (FIR) and are moving average (MA) systems instead of the more generic ARMA models. In this case, the filters $C(z)$ and $D(z)$ have all their poles at the origin; they take the form

$$C(z) = c_0 + c_1 z^{-1} + \cdots + c_N z^{-N}, \quad (9)$$

$$D(z) = d_0 + d_1 z^{-1} + \cdots + d_N z^{-N}, \quad (10)$$

where the maximum lag N denotes the filter order. For such an orthogonal FIR filter bank, it is well known that the orthogonality conditions imply that $D(z) = \pm(-z)^{-N}C(-z^{-1})$, see [33]. Since the choice of sign for $D(z)$ only corresponds to a sign choice for $\psi(t)$ and is of no further consequence, we fix it to be positive. Then the coefficients of $D(z)$ are related to those of $C(z)$ by the ‘alternating flip’ construction [20]:

$$d_k = (-1)^k c_{N-k}, \quad (k = 0, 1, \dots, N). \quad (11)$$

Also, since zero coefficients at the start or end of the sequence $\{c_k\}$, $k = 0, \dots, N$, can either be compensated for by shifting the signal or reducing the filter order, it is natural to impose both $c_0 \neq 0$ and $c_N \neq 0$. Then orthonormality of the basis $\{\phi(t - m) \mid m \in \mathbb{Z}\}$ implies that N is odd: $N = 2n - 1$. The polyphase filter $H(z)$ is also FIR. It has the order $n - 1$. The scaling function $\phi(t)$ and the wavelet $\psi(t)$ both live on the interval $[-N, 0]$ in this setup.³

Regarding orthonormality and losslessness of $H(z)$, the filters $C(z)$ and $D(z)$ play a symmetric role. This symmetry is broken by their connection to the dilation and the wavelet equation. It implies that $C(z)$ is low pass and generates an approximation signal, while $D(z)$ is high pass and generates a detail signal. And also that the *wavelet admissibility condition* is imposed which requires that $\psi(t)$ has integral $\int_{\mathbb{R}} \psi(t) dt = 0$. The wavelet $\psi(t)$ and the high-pass filter $D(z)$ are both said to have a first vanishing moment (which is of order zero) [20].

3 Parameterization of Wavelets as Lossless FIR Polyphase Filters

Various parameterizations of lossless FIR polyphase filters are available in the literature. General parameterizations for $p \times p$ lossless systems of a given order, for instance based on interpolation theory and the tangential Schur algorithm such as given in [1] and [13], can be specialized to the 2×2 FIR case by placing all the poles at the origin. Similarly, one may employ the Potapov factorization [26]. In [30, 33] and references

³ The interval $[-N, 0]$ emerges because of the use of $\phi(2t + k)$ in the dilation equation. The coefficient s_k of a signal $s(t) = \sum_{k \in \mathbb{Z}} s_k \phi(t - k)$ can then be associated with the right end point of the interval $[-N + k, k]$ on which $\phi(t - k)$ lives (similar observations apply to the approximation and detail signals). This is convenient for online signal processing in view of causality. For detection applications, one must take into account that the location of the mode of the wavelet is important, which is biased to the left and scale dependent.

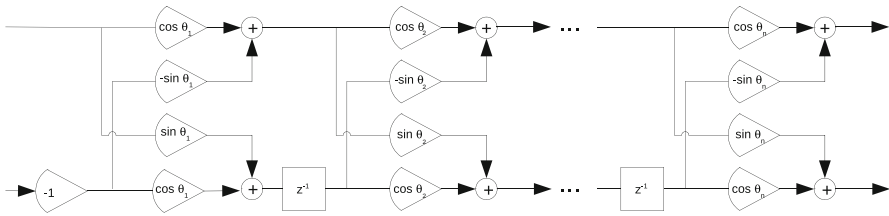


Fig. 1 Lattice filter implementation of a 2×2 lossless FIR polyphase filter $H(z)$

given there, related parameterizations for the 2×2 FIR case are given in which $H(z)$, of order $n - 1$, is factored into a product of rotations and elementary single-channel delays:

$$H(z) = R(\theta_n)\Lambda(z)R(\theta_{n-1})\Lambda(z) \cdots \Lambda(z)R(\theta_2)\Lambda(z)R(\theta_1)\Lambda(-1) \quad (12)$$

where

$$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \quad \Lambda(z) = \begin{pmatrix} 1 & 0 \\ 0 & z^{-1} \end{pmatrix} \quad (13)$$

This parameterization involves n angular parameters $\theta_1, \theta_2, \dots, \theta_n$, associated with the rotation matrices $R(\theta_k)$. The matrix $\Lambda(-1) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$ appears as the last factor of $H(z)$ to comply with the sign convention of the alternating flip construction (11). Note that the factorization (12) makes it easy to compute the partial derivatives of $H(z)$ with respect to the parameters, which can be an advantage for implementing optimization routines.

Another nice feature of this parameterization is that it admits a lattice filter implementation [30,32], in which the initial section represents the constant transfer matrix $R(\theta_1)\Lambda(-1)$, while the other $n - 1$ lattices represent the first-order lossless transfer matrices $R(\theta_k)\Lambda(z)$, ($k = 2, \dots, n$); it is displayed in Fig. 1.

The factorization (12) also allows to recursively compute the filter coefficients $c_0, c_1, \dots, c_{2n-1}$ from a given set of parameters $\theta_1, \dots, \theta_n$. If the coefficients of a lossless FIR polyphase filter of order $n - 2$ with parameters $\theta_1, \dots, \theta_{n-1}$ are denoted by $\hat{c}_0, \hat{c}_1, \dots, \hat{c}_{2n-3}$, then:

$$\begin{pmatrix} c_0 & c_2 & \dots & c_{2n-4} & c_{2n-2} \\ c_{2n-1} & c_{2n-3} & \dots & c_3 & c_1 \end{pmatrix} = R(\theta_n) \begin{pmatrix} \hat{c}_0 & \hat{c}_2 & \dots & \hat{c}_{2n-4} & 0 \\ 0 & \hat{c}_{2n-3} & \dots & \hat{c}_3 & \hat{c}_1 \end{pmatrix}. \quad (14)$$

Conversely, to compute θ_n from a given set of coefficients of a lossless FIR polyphase filter, note that θ_n must be chosen such that zeros appear on the specified locations in the coefficient matrix of the right-hand side expression of (14). To parameterize all the lossless FIR polyphase filters of order $n - 1$, it suffices to take $\theta_1 \in [-\pi, \pi)$ (any half-open interval of length 2π will do) and $\theta_2, \dots, \theta_n \in [-\frac{\pi}{2}, \frac{\pi}{2})$ (any half-open interval of length π will do). A set of filter coefficients then corresponds to a unique set of angular parameters (from the chosen intervals). Filters of order less than $2n - 1$ are embedded (with additional delays) in multiple ways into the parameterization.

As indicated, a first vanishing moment (of order zero) is needed for the wavelet interpretation of $\psi(t)$ and it is implied by the dilation and wavelet equations. The (normalized) scaling function $\phi(t)$ has integral 1 and the wavelet $\psi(t)$ has integral zero. By integrating (6) and (7), it follows that the low-pass filter coefficients add up to $c_0 + c_1 + \dots + c_{2n-1} = \sqrt{2}$ and the high-pass filter coefficients to $d_0 + d_1 + \dots + d_{2n-1} = 0$. For the polyphase filter $H(z)$, this implies the following result.

Proposition 1 (Cf. [30, Theorem 4.6]) *Consider a lossless FIR polyphase matrix $H(z)$, given in terms of the angular parameters $\theta_1, \dots, \theta_n$ by Eq. (12).*

(a) *Then the wavelet admissibility condition holds (i.e., the wavelet has a vanishing moment of order zero), if and only if*

$$H(1) = \frac{1}{2}\sqrt{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{15}$$

(b) *The condition under (a) is satisfied in terms of the angular parameters if and only if*

$$\theta_1 + \dots + \theta_n = \frac{\pi}{4} \pmod{2\pi}. \tag{16}$$

For an admissible lossless FIR polyphase filter of order $n - 1$, the condition of Proposition 1(b) leaves $n - 1$ degrees of freedom, which we normally assign to $\theta_2, \dots, \theta_n$ by fixing θ_1 as $\theta_1 = \frac{\pi}{4} - (\theta_2 + \dots + \theta_n) \pmod{2\pi}$.

For some computations in this paper, it will be convenient to fix θ_n instead, and to assign the $n - 1$ degrees of freedom to the parameters $\theta_1, \dots, \theta_{n-1}$. We then propose to re-parameterize the polyphase filters using a Potapov decomposition [26] in the following way. First, note that in Proposition 1 the matrix $H(1) = H(1)^{-1}$ is symmetric and orthogonal. Let $\tilde{H}(z)$ be defined by

$$\tilde{H}(z) := H(z)H(1)^{-1}. \tag{17}$$

This is also a lossless FIR polyphase filter of order $n - 1$, now satisfying $\tilde{H}(1) = I_2$. Therefore, $\tilde{H}(z)$ admits a Potapov decomposition, which is a product of ‘elementary’ lossless filters of order 1:

$$\tilde{H}(z) = B_{\phi_{n-1}}(z)B_{\phi_{n-2}}(z) \dots B_{\phi_1}(z), \tag{18}$$

with

$$B_{\phi_k}(z) = I_2 + \frac{1}{2}(z^{-1} - 1) \begin{pmatrix} I_2 - \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} R(2\phi_k) \end{pmatrix}. \tag{19}$$

It can be worked out directly that the parameters ϕ_k in this new setup are given in terms of $\theta_1, \dots, \theta_{n-1}$ by:

$$\phi_k = \theta_1 + \dots + \theta_k, \quad (k = 1, \dots, n - 1). \tag{20}$$

Note that because each parameter θ_k can be freely chosen from an arbitrary half-open interval of length π , the same holds for each parameter ϕ_k . In addition, observe that

$B_{\phi_k}(1) = I_2$, which makes it easy to increase or decrease the order of the polyphase filter by merely adding or deleting parameters ϕ_k .

4 Conditions for Vanishing Moments

Vanishing moments are useful wavelet properties for several reasons. In this section, we aim to explain some of this background and also to express conditions for them in terms of the polyphase matrices $H(z)$ and $\tilde{H}(z)$, and in terms of the two parameterizations discussed in Sect. 3. A wavelet is said to have p vanishing moments (of orders $0, 1, \dots, p-1$) if it holds that

$$\int_{\mathbb{R}} t^m \psi(t) dt = 0, \quad m = 0, 1, \dots, p-1. \quad (21)$$

This implies that all polynomial time functions of degree $\leq p-1$ are orthogonal to $\psi(t)$ and to its shifts and dilations. Because the scaling function and the wavelet have compact support, this interpretation holds locally too: Local polynomial approximations up to degree $p-1$ belong to the approximation spaces.

Also, it is well known that smoothness of wavelets requires vanishing moments too. Orthogonal wavelets need p vanishing moments to be $p-1$ times continuously differentiable [18].

The vanishing moments property of *wavelets* $\psi(t)$ can also be expressed as vanishing moment conditions for the high-pass *filter* $D(z)$. It is well known that Eq. (21) holds if and only if

$$\sum_{k=0}^{2n-1} k^m d_k = 0, \quad m = 0, 1, \dots, p-1. \quad (22)$$

Equivalently,

$$D^{(m)}(1) = 0, \quad m = 0, 1, \dots, p-1, \quad (23)$$

where $D^{(m)}(z)$ denotes the m th-order derivative of $D(z)$. Because of the alternating flip construction, this is equivalent to related conditions on the low-pass filter:

$$\sum_{k=0}^{2n-1} (-1)^k k^m c_k = 0, \quad m = 0, 1, \dots, p-1, \quad (24)$$

and

$$C^{(m)}(-1) = 0, \quad m = 0, 1, \dots, p-1. \quad (25)$$

This shows that the filters $C(z)$ and $D(z)$ have flatness at $z = -1$ and $z = 1$, respectively, up to order $p-1$. See also, e.g., [7, 30].

For the two parameterizations at hand, the question arises which conditions on the angular parameters $\theta_2, \dots, \theta_n$ or $\phi_1, \dots, \phi_{n-1}$, respectively, enforce $p-1$ extra vanishing moments (in addition to the wavelet admissibility condition). For a second vanishing moment (i.e., of order 1), we have the following result.

Proposition 2 Let $H(z)$ be a 2×2 lossless FIR polyphase matrix of order $n - 1$, satisfying the admissibility condition (15), ensuring a vanishing moment of order 0. Let $H(z)$ be parameterized in terms of the angular parameters $\theta_1, \dots, \theta_n$ according to Eq. (12), satisfying the condition (16). Let $\tilde{H}(z)$ be the associated lossless filter given by Eq. (17), parameterized in terms of the angular parameters $\phi_1, \dots, \phi_{n-1}$, defined by Eq. (20). Then $H(z)$ has a vanishing moment of order 1 if and only if one of the following equivalent conditions holds:

$$(a) \quad H'_{11}(1) - H'_{12}(1) = -\frac{1}{4}\sqrt{2},$$

$$(b) \quad H'_{21}(1) + H'_{22}(1) = -\frac{1}{4}\sqrt{2},$$

$$(c) \quad \sum_{k=1}^{n-1} \sin(2(\theta_{k+1} + \dots + \theta_n)) = -\frac{1}{2},$$

$$(d) \quad \tilde{H}'_{12}(1) = -\frac{1}{4},$$

$$(e) \quad \tilde{H}'_{21}(1) = -\frac{1}{4},$$

$$(f) \quad \sum_{k=1}^{n-1} \cos(2\phi_k) = -\frac{1}{2}.$$

Though these conditions follow directly from well-known Eqs. (7) and (13), the conditions of Proposition 2 are not trivially built into a parameterization. While before we had that condition (16) was easily handled by eliminating the parameter θ_1 , condition (c) of Proposition 2 cannot always be rewritten to express θ_2 in terms of the other parameters. This happens because values for $\theta_3, \dots, \theta_n$ can be chosen such that no feasible value for θ_2 exists which satisfies the condition. For $n = 3$ this is illustrated by Fig. 2a. In this case, because of the low order, the manifold of such filters is topologically equivalent to a circle, and one may in fact identify an interval of feasible values for θ_2 ; it then is possible to (generically) build in this property. The same problem exists for the parameterization in terms of ϕ_k , as is illustrated in Fig. 2b, now relating to condition (f) of Proposition 2.

A third vanishing moment (of second order) can be characterized as follows.

Proposition 3 Let $H(z)$ be a 2×2 lossless FIR polyphase matrix of order $n - 1$, satisfying the admissibility condition (15) and the conditions of Proposition 2, ensuring vanishing moments of orders 0 and 1. Let $H(z)$ be parameterized by the angular parameters $\theta_1, \dots, \theta_n$ according to Eq. (12). Let $\tilde{H}(z)$ be the associated lossless filter given by Eq. (17), parameterized by the angular parameters $\phi_1, \dots, \phi_{n-1}$ given by Eq. (20). Then $H(z)$ has a vanishing moment of order 2 if and only if one of the following equivalent conditions holds:

$$(a) \quad H''_{11}(1) - H''_{12}(1) + \frac{1}{2}H'_{11}(1) + \frac{1}{2}H'_{12}(1) = \frac{1}{4}\sqrt{2},$$

$$(b) \quad H''_{21}(1) + H''_{22}(1) + \frac{1}{2}H'_{21}(1) - \frac{1}{2}H'_{22}(1) = \frac{1}{4}\sqrt{2},$$

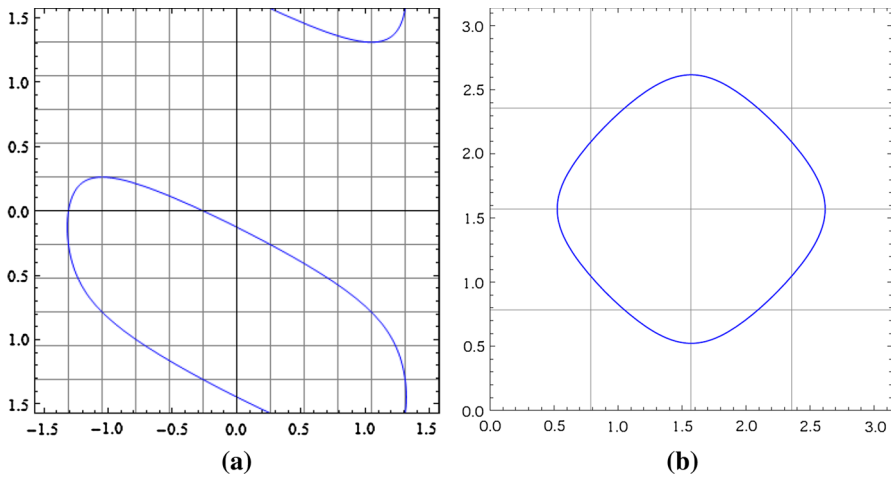


Fig. 2 Curves in the parameter space for $n = 3$, corresponding to orthogonal wavelets with two vanishing moments. The lossless polyphase filter satisfying the wavelet admissibility structure has $n - 1 = 2$ free parameters: θ_2 (abscissa) and θ_3 (ordinate), or ϕ_1 and ϕ_2 , respectively, depending on the choice of parameterization. Clearly there are choices of θ_2 for which no θ_3 can be found to satisfy condition (c) of Proposition 2. Likewise, this holds for ϕ_1 and ϕ_2 with respect to condition (f). The parameter spaces are periodic: opposite sides of the square parameter areas are to be identified. **a** Parameterization with θ_2, θ_3 , **b** Parameterization with ϕ_1, ϕ_2

$$(c) \quad \sum_{1 \leq k < \ell \leq n-1} \sin(2(\theta_{k+1} + \dots + \theta_\ell)) + \frac{1}{2} \sum_{k=1}^{n-1} \cos(2(\theta_{k+1} + \dots + \theta_n)) = 0,$$

$$(d) \quad \tilde{H}''_{12}(1) + \frac{1}{2} \tilde{H}'_{11}(1) = \frac{1}{4},$$

$$(e) \quad \tilde{H}''_{21}(1) + \frac{1}{2} \tilde{H}'_{22}(1) = \frac{1}{4},$$

$$(f) \quad \sum_{1 \leq k < \ell \leq n-1} \sin(2(\phi_\ell - \phi_k)) + \frac{1}{2} \sum_{k=1}^{n-1} \sin(2\phi_k) = 0.$$

For $n = 4$, the parameter θ_1 can be eliminated as before. But to handle θ_2, θ_3 and θ_4 conveniently is far more complicated. In Fig. 3a the manifolds of filters having 2 or 3 vanishing moments are displayed. For 3 vanishing moments it consists of two connected components (the green curves). Switching to the parameters ϕ_1, ϕ_2, ϕ_3 as presented in Fig. 3b is more convenient, but likewise shows the complexity involved. Because periodicity holds (opposite sides are to be identified) the blue surfaces for 2 vanishing moments are closed and have genus 3.

In [25] a procedure was developed to build up to 4 vanishing moments into a different parameterization of 2×2 lossless systems of arbitrary (sufficiently large) order, using the tangential Schur algorithm with interpolation conditions at the stability boundary. However, that parameterization was not restricted to FIR filters. A parameterization which imposes both the FIR property and has several vanishing moments is still lacking. So far, the parameterizations given here, in terms of $\theta_1, \dots, \theta_n$ or in terms of

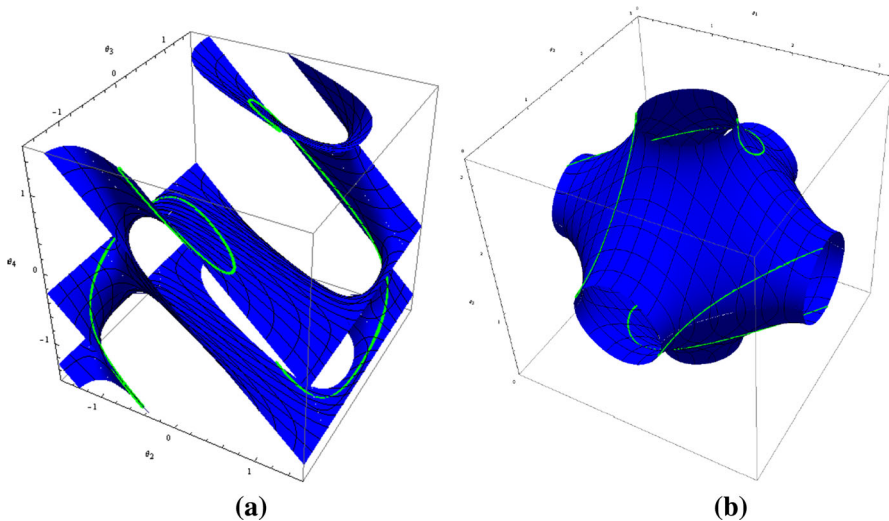


Fig. 3 Surfaces and curves in the parameter space for $n = 4$, corresponding to orthogonal wavelets with two vanishing moments (blue surfaces) and three vanishing moments (green curves). The lossless polyphase filter satisfying the wavelet admissibility structure has $n - 1 = 3$ free parameters: θ_2 , θ_3 and θ_4 , or ϕ_1 , ϕ_2 and ϕ_3 , respectively, depending on the choice of parameterization. Clearly there are choices of θ_2 and θ_3 for which no θ_4 can be found to satisfy condition (c) of Proposition 2. Likewise, this holds for ϕ_1 , ϕ_2 and ϕ_3 with respect to condition (f). The same holds for the green curves for conditions (c) and (f) of Proposition 3. The parameter spaces are periodic: opposite faces of the cubic parameter areas are to be identified. **a** Parameterization with θ_2 , θ_3 , θ_4 , **b** Parameterization with ϕ_1 , ϕ_2 , ϕ_3 (Color figure online)

$\phi_1, \dots, \phi_{n-1}$, with the accompanying constraints for vanishing moments up to order 2, seem to be the most practical currently available for numerical optimization.

5 Measuring the Quality of a Wavelet Representation

As previously argued in [15, 16], choosing a good wavelet is not trivial. Some vanishing moments and an optimal ratio of the Heisenberg uncertainty rectangle in time–frequency, see [22], are usually considered desirable properties for wavelets to have, for general reasons. One can also attempt to measure the quality of a wavelet in terms of the representation of a prototype signal in the wavelet domain. For compression purposes it then is intuitive that it is beneficial if a prototype signal gets represented in the wavelet domain *sparsely*: It will give a high compression rate. Similarly, sparsity can be useful for feature detection purposes, interpreting a sparse feature representation as a ‘fingerprint’. Note that sparsity encompasses template matching, but is quite a bit more flexible: It allows a feature to be built from several copies of the wavelet at different positions and scales, while values of the wavelet coefficients may still show some variation.

Here we advocate to design a matched wavelet for a given application by *maximizing sparsity of the wavelet representation of a prototype signal*. To represent a signal in the wavelet domain there are several options. First, one may use the ‘critically sampled’

discrete wavelet transform where dyadic downsampling takes place [19]. Second, one can employ the ‘redundant’ undecimated wavelet transform [24], to have time invariance and full resolution at every dyadic scale. Third, one could use the wavelet and the dilation equation to first find the wavelet and scaling function [30] and then employ the continuous wavelet transform [20]. We will first focus on a matching approach for the critically sampled wavelet transform and then, in the next section, indicate how this can be extended to accommodate the undecimated wavelet transform too.

To obtain a critically sampled signal decomposition, the wavelet filter bank (with downsampling) is repeatedly applied to the low-pass filter outputs [19]. At each level j we shall denote the detail coefficients by $b_k^{(j)}$ and the approximation coefficients by $a_k^{(j)}$. Since the approximation signals get processed to obtain the next level, only the approximation coefficients $a_k^{(j_{\max})}$ at the coarsest level j_{\max} show up in the end result of the decomposition.

For orthonormal wavelets, a wavelet analogue of Parseval’s identity holds. It states [20] that the energy of an input signal s of finite length m , which for convenience is assumed to be a power of two, is preserved in the approximation and detail signals [15,30]:

$$\sum_{k=1}^m s_k^2 = \sum_{j=1}^{j_{\max}} \sum_{k=1}^{m/2^j} (b_k^{(j)})^2 + \sum_{k=1}^{m/2^{j_{\max}}} (a_k^{(j_{\max})})^2. \tag{26}$$

If a vector w is introduced to contain all the detail and approximation coefficients of a signal s , as in [15]:

$$w = (b_1^{(1)}, \dots, b_{m/2}^{(1)} | b_1^{(2)}, \dots, b_{m/4}^{(2)} | \dots \dots | b_1^{(j_{\max})}, \dots, b_{m/2^{j_{\max}}}^{(j_{\max})} | a_1^{(j_{\max})}, \dots, a_{m/2^{j_{\max}}}^{(j_{\max})}) \tag{27}$$

then the L^2 -norms of s and w are equal: $\|s\|_2 = \|w\|_2$.

When performing wavelet design over the class of orthonormal wavelets with a fixed choice of the filter order $N = 2n - 1$, it is important to note that $\|w\|_2$ is the same for every choice of the parameters. If the goal is to promote sparsity subject to this constraint on the L^2 -norm of w , a conceptual approach is to maximize the variance. To make this concept operational, we propose two approaches, see also [15, 16]:

1. maximize the variance of the absolute values of the wavelet coefficients;
2. maximize the variance of the squared wavelet coefficients.

The following theorem [15,16] shows that the former approach corresponds to minimization of an L^1 -norm, which is a well-known criterion to achieve sparsity in various other contexts, see for example [4,5,9,10]. The latter approach maximizes the variance of the energy distribution over the approximation and detail coefficients at the various scales, and corresponds to maximization of an L^4 -norm.

Theorem 1 *Let w be the vector of the wavelet and the approximation coefficients as in (27), resulting from the processing of a signal s of length m by means of an orthogonal filter bank, depending on a set of free parameters to be optimized. Then:*

- (1) *Maximization of the variance of the sequence of absolute values $|w_k|$ is equivalent to minimization of the L^1 -norm $V_1 = \sum_{k=1}^m |w_k|$.*
- (2) *Maximization of the variance of the sequence of energies $|w_k|^2$ is equivalent to maximization of the L^4 -norm $V_4 = (\sum_{k=1}^m |w_k|^4)^{1/4}$.*

We have used both optimization criteria and explored their merits for wavelet matching. Their advantages and drawbacks are discussed, together with several examples, in the following two sections.

6 Practical Considerations for Wavelet Matching

The parameterization with angular parameters $\theta_1, \dots, \theta_n$ developed first in Sect. 3 allows for a convenient design of orthogonal wavelets, by means of a local search over the search space $[-\frac{\pi}{2}, \frac{\pi}{2}]^{n-1}$, where either θ_1 or θ_n is fixed to enforce the wavelet admissibility condition. (Alternatively, the parameters $\phi_1, \dots, \phi_{n-1}$ can be used in a similar manner.) For additional vanishing moments, some other parameters can be fixed, depending on the values of the remaining free parameters θ_k . During iterative search, if fixing these parameters for a current set of free parameter values is not possible, then the optimization criterion V_1 or V_4 is augmented with a large penalty term (with an appropriate sign) to force the iterations back into the region where the constraints can be satisfied. Caution must be exercised due to the existence of local optima as discussed in [15]. As an example, five local optima encountered for a case with two free parameters are shown in Fig. 4. Here the solutions labeled 1 and 3 are very similar and they are located in two different valleys of the search space. Between solutions 1 and 3, there is an apparent time delay in effect, which by the alternating flip construction also produces a sign change in the wavelet function (Fig. 5).

A second point to consider when choosing between V_1 and V_4 is *masking*. When only a certain time span of the prototype signal (e.g., carrying a distinctive feature) or a certain frequency band corresponding to approximate wavelet scales is of interest, then one can use a time–frequency mask on the wavelet domain representation of the prototype signal and evaluate the criteria on that. But when a *minimization* criterion is employed, the energy will simply tend to be forced outside the mask. This might easily result in an undesirable delay filter, and this phenomenon is not straightforward to counteract. Hence when masking is used, it is recommendable to use the *maximization* criterion V_4 for which this drawback does not occur. As an example, in Fig. 12 in Sect. 7 masking is used to force energy into the fine scales corresponding to the time location of two QRS complexes in an ECG signal.

Third, an extension that might be beneficial, e.g., for detection purposes, is to employ an *overcomplete representation*. Retaining dyadic scales, but using uniform time sampling as in the *undecimated* [29] or the *stationary wavelet transform* [24], ensures that the representation becomes translation invariant. Since the coefficients of the critically sampled wavelet transform for different time shifts are all subsets of the undecimated coefficients, ‘conservation of energy’ still applies for any such selection of subsets. The overall energy therefore is a weighted sum of the energies per scale, where the weights decrease by a factor two per scale:

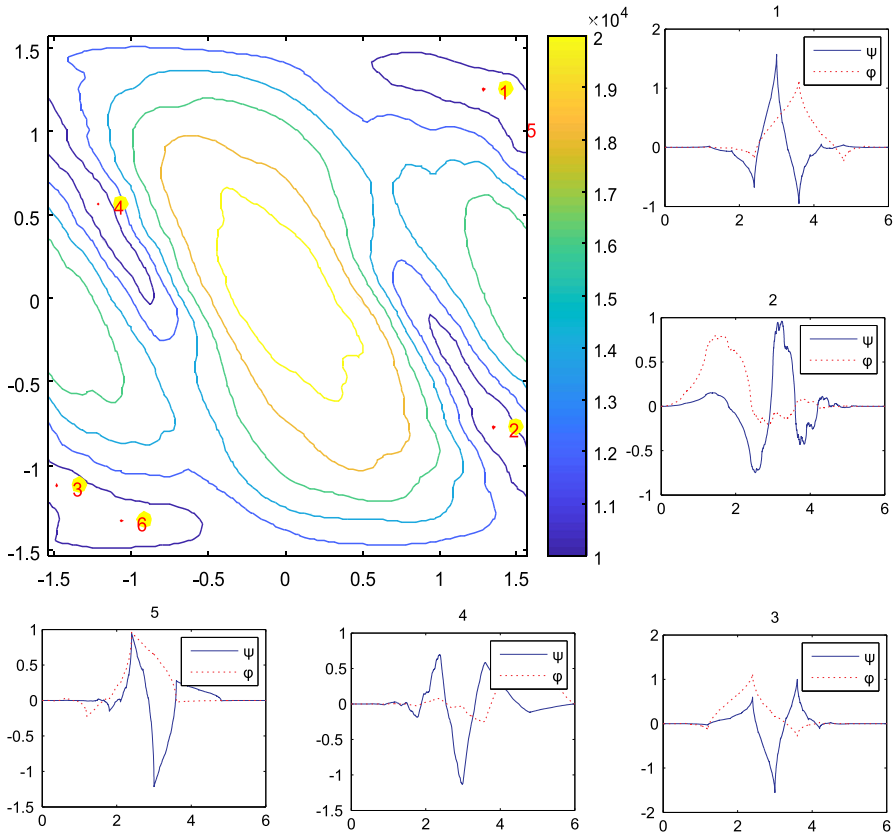


Fig. 4 Minimization of criterion V_1 for a case with $n = 3$ and with one vanishing moment built in. There are two free parameters: θ_2 and θ_3 . Top-left: Contour plot of V_1 . Five labeled dots mark the computed local minima, though there are more local minima present. Small graphs, clockwise The five wavelets $\psi(t)$ (blue) and scaling functions $\phi(t)$ (red) (Color figure online)

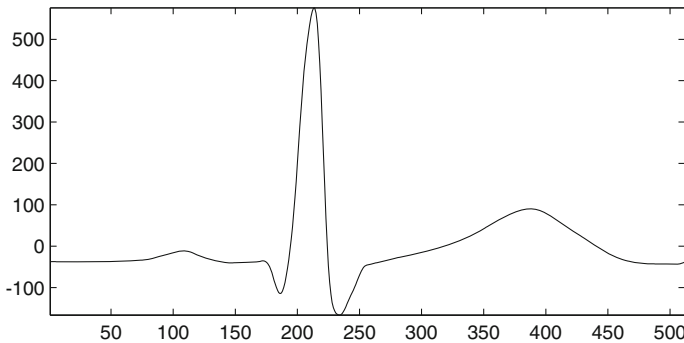


Fig. 5 Prototype signal for the example in Fig. 4

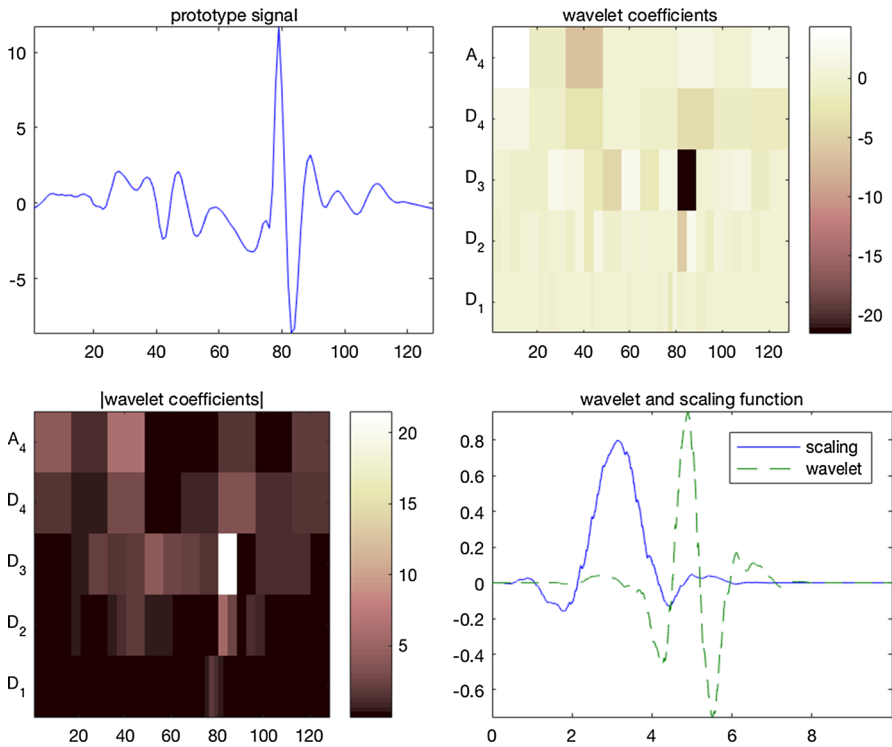


Fig. 6 Example 1, matched wavelets for an acceleration signal (anterior/posterior) from normal gait. Optimization of the V_1 criterion, using 4 free parameters and one vanishing moment. Top-left: original signal. Top-right: wavelet decomposition of the signal. In the top row, the approximation coefficients are displayed and in the rows thereafter the detail coefficients from coarse to fine. The intensities of the blocks correspond to the coefficient values. Bottom-left: absolute values of wavelet coefficients. Bottom-right: wavelet and scaling function of the optimized wavelet

$$\sum_{k=1}^m s_k^2 = \sum_{j=1}^{j_{\max}} \left(2^{-j} \sum_{k=1}^m (b_k^{(j)})^2 \right) + 2^{-j_{\max}} \sum_{k=1}^m (a_k^{(j_{\max})})^2 \tag{28}$$

Here s is the signal of length m , j_{\max} the number of scales (levels) of the wavelet decomposition, and all the scales of the wavelet decomposition have m detail coefficients $b_k^{(j)}$ and approximation coefficients $a_k^{(j)}$ (i.e., with $k = 1, \dots, m$). To avoid finite length effects, we have applied ‘wrap around’ to induce periodicity and to obtain wavelet coefficient sequences of length m at every scale. As a result, likewise adjusted criteria V_1 and V_4 can be used with the undecimated wavelet transforms. We will refer to these adjusted criteria as the *undecimated* V_1 and V_4 . The translation invariance of the undecimated criteria helps to avoid undesired phenomena and ensures that sparsity is more robustly built in.

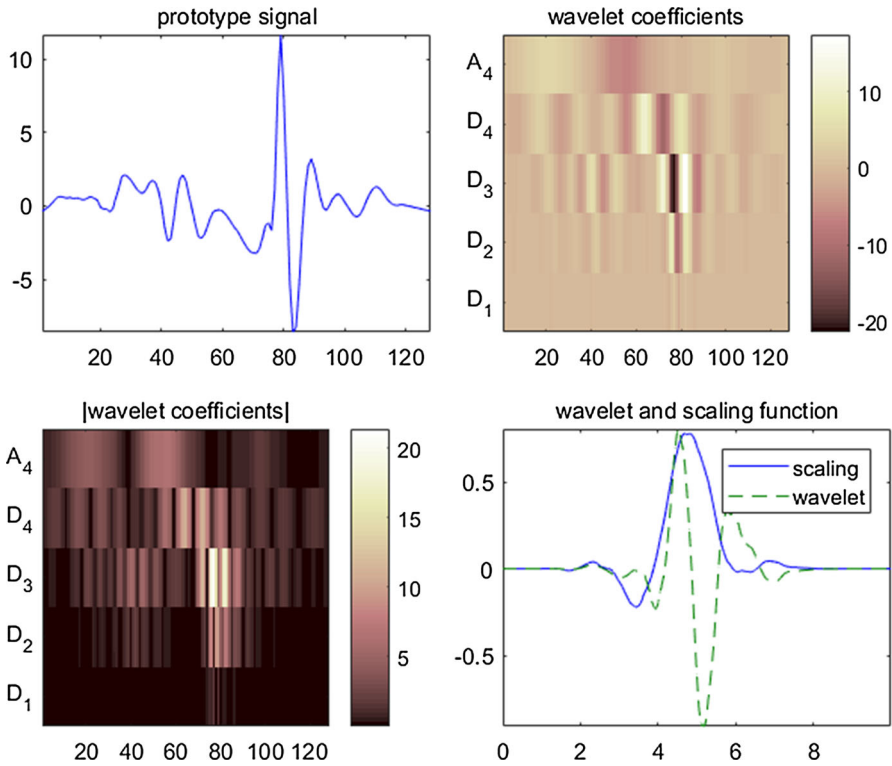


Fig. 7 Example 1, as in Fig. 6, now optimizing the undecimated V_1 criterion with 3 free parameters and two vanishing moments

7 Wavelet Matching Examples

Example 1 To compare wavelet matching for the critically sampled wavelet transform and the undecimated wavelet transform, we took an acceleration signal for a human patient which was measured during a step in normal gait. It is displayed in Figs. 6 and 7.

A first matched orthogonal wavelet was designed by minimizing the criterion V_1 , using the critically sampled wavelet transform. The polyphase filter order was chosen as $n - 1 = 4$, and one vanishing moment was enforced, leaving 4 free parameters for optimization. The low-pass filter coefficients are given in Table 1 and the results shown in Fig. 6. Sparsity of the resulting wavelet representation is confirmed by the presence of a single large wavelet component on the detail scale D_3 . The morphology of the matched wavelet function on this scale indeed closely mimics the morphology of the dominant component of the prototype signal, as displayed in Fig. 8.

In Fig. 7, the results are shown for minimization of the undecimated V_1 criterion using the same filter order with 3 free parameters and two vanishing moments. Again a sparse representation is obtained, and the new orthogonal wavelet (green) turns out to be similar to the one computed earlier, while the new scaling function (blue) shows

Table 1 Low-pass filter coefficients for the matched wavelets of Examples 1–3

Figures	c_0	c_1	c_2	c_3	c_4	c_5	c_6	c_7	c_8	c_9	c_{10}	c_{11}
5a	-0.1193	0.0165	0.5583	0.7681	0.2620	-0.1164	0.0061	0.0388	0.0000	0.0001		
5b	0.0211	-0.0795	-0.1535	0.2900	0.7820	0.5203	0.0168	-0.0346	0.0407	0.0108		
7a	-0.0085	0.0319	0.0147	-0.1532	-0.1528	0.3782	0.7991	0.4095	0.0228	0.0322	0.0319	0.0085
7b	0.1966	0.2001	0.1371	0.4599	0.6244	0.2509	-0.2690	-0.3230	-0.0690	0.2047	0.0870	-0.0855
8a	0.2161	0.6733	0.6653	0.0316	-0.2294	0.0200	0.0552	-0.0177				
8b	0.0263	-0.1061	0.1623	0.8078	0.5545	0.0143	-0.0361	-0.0090				

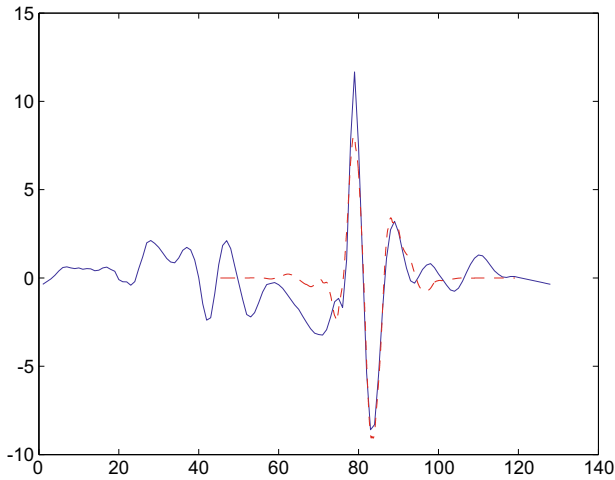


Fig. 8 The wavelet function (red, dashed) and the acceleration signal (blue, solid) after alignment (Color figure online)

Table 2 L^1 -norm of critically sampled wavelet decomposition coefficients for the prototype signal of example 1, for different choices of wavelets

Wavelet	L^1 -norm	Percentage
Designed wavelet example 1, 1 vanishing moment	96.5470	100
Designed wavelet example 2, 2 vanishing moments	112.4878	117
Daubechies, 10 filter coefficients	122.7822	127
Symlet, 10 filter coefficients	115.7073	120

a time shift with respect to the previous one. The undecimated wavelet transform has the same high resolution on all scales and admits a more precise detection of the locations of the peaks in the signal. The low-pass filter coefficients are again given in Table 1. One can investigate the differences in sparsity obtained by the wavelet design by considering the L^1 -norm of the wavelet coefficients. This is illustrated in Table 2 for the two designed wavelets in example 1 and the Daubechies and Symlet wavelets all with the same prototype signal, 10 filter coefficients and 4 scales. From these measurements, it is clear that the design of a wavelet with only one vanishing moment gives a considerable increase in sparsity (lower L^1 -norm), which can be also witnessed in Fig. 6, where a lot of the coefficient values are close to zero. Adding an additional vanishing moment comes at the cost of freedom, and in this case the consequence is that there is a lower increase in sparsity.

The sparse representation of the normal gait in the wavelet domain makes it possible to isolate and remove the gait pattern from other measured acceleration signals of the same subject. This approach, with the matched wavelet presented here, was successfully applied in [17], to identify stumbles in acceleration signals.

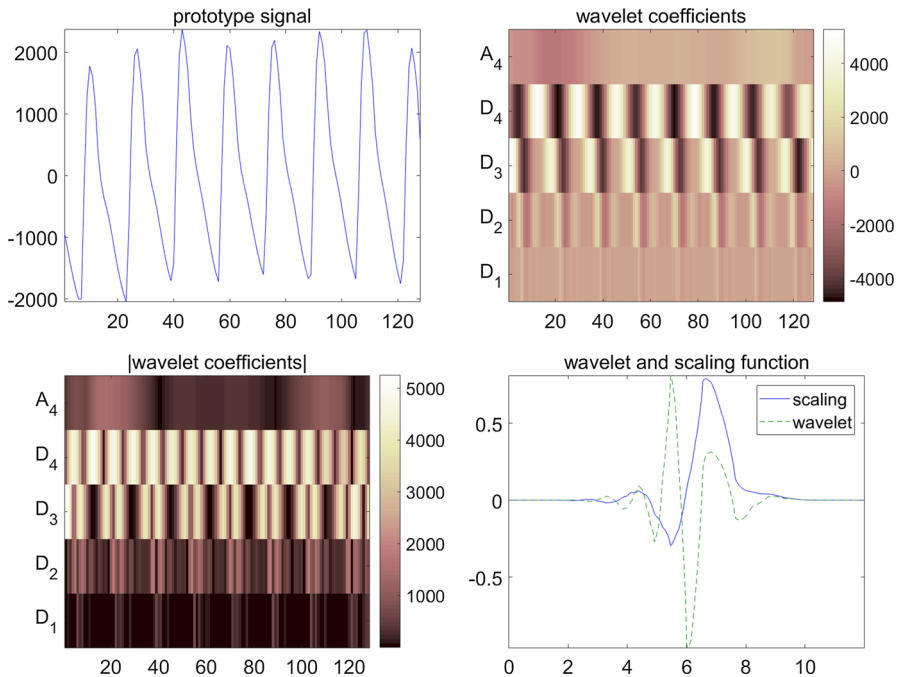


Fig. 9 Example 2, matched wavelet and scaling function for a blood pressure signal, minimizing the undecimated criterion V_1 with 4 free parameters and 2 vanishing moments

Example 2 To compare the undecimated criteria V_1 and V_4 , we took a quasi-periodic blood pressure signal from the ‘time course data for blood pressure in Dahl SS and SSBN13 rats’, see [3]. We designed two orthogonal matched wavelets, in both cases using a polyphase filter order $n - 1 = 5$ and enforcing two vanishing moments, leaving 4 free parameters for optimization. See Table 1 for their low-pass filter coefficients. The first matched wavelet was designed by minimizing the undecimated criterion V_1 , see Fig. 9. The second by maximizing the undecimated criterion V_4 , see Fig. 10. The first wavelet appears to achieve a better localization of the signal in the wavelet domain, as the signal peaks are slightly more pronounced on the detail scales D_3 and D_4 . The second wavelet shows more oscillations and is less smooth than the first one.

Example 3 To study the effects of masking, we took an ECG signal from the St. Petersburg INCART database, see [11]. It is displayed in Figs. 11 and 12. We chose a polyphase filter order $n - 1 = 3$ and enforced two vanishing moments, leaving 2 free parameters for optimization. First we designed a matched orthogonal wavelet by maximizing the undecimated criterion V_4 , see Fig. 11. The optimization procedure, employing multiple random starting points, terminated with the coefficients reported in Table 1. The Euclidean distance of the coefficient vector to that of the Daubechies 4 filter coefficients is only 0.0955.

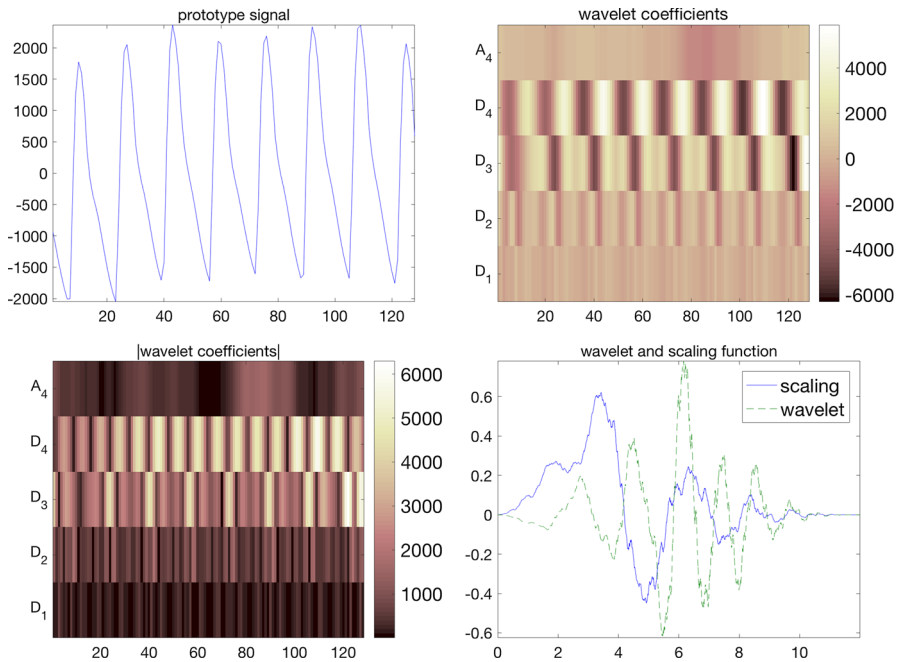


Fig. 10 Example 2, matched wavelet and scaling function for a blood pressure signal, maximizing the undecimated criterion V_4 with 4 free parameters and 2 vanishing moments

A second wavelet was designed using the undecimated criterion V_4 with masking. Two masks were introduced to emphasize two selected QRS complexes in the fine scales, as illustrated in Fig. 12. This resulted in a wavelet that is close to the Daubechies 2 wavelet. The Euclidean distance of the coefficient vector to that of the Daubechies 2 filter coefficients is 0.1122. For this second wavelet, one can observe that the other QRS complexes not used for masking become better pronounced in the fine scales too.

8 Conclusions and Discussion

From the examples presented in the previous section, we conclude that the proposed approach for designing matched orthogonal wavelets is a feasible one, with a clear potential for applications. The availability of parameterizations, which build in orthogonality and allow for explicit conditions to get extra vanishing moments, makes it relatively easy to compute matched wavelets through constrained optimization. But still some user choices are to be made, which will be application dependent, and so the method cannot fully be automated.

First, the user has to decide on the filter order $n - 1$ and the number of extra vanishing moments $p - 1$. The higher the value of n , the more complex the optimization problem becomes. A higher value of p does reduce the degrees of freedom, but increases the

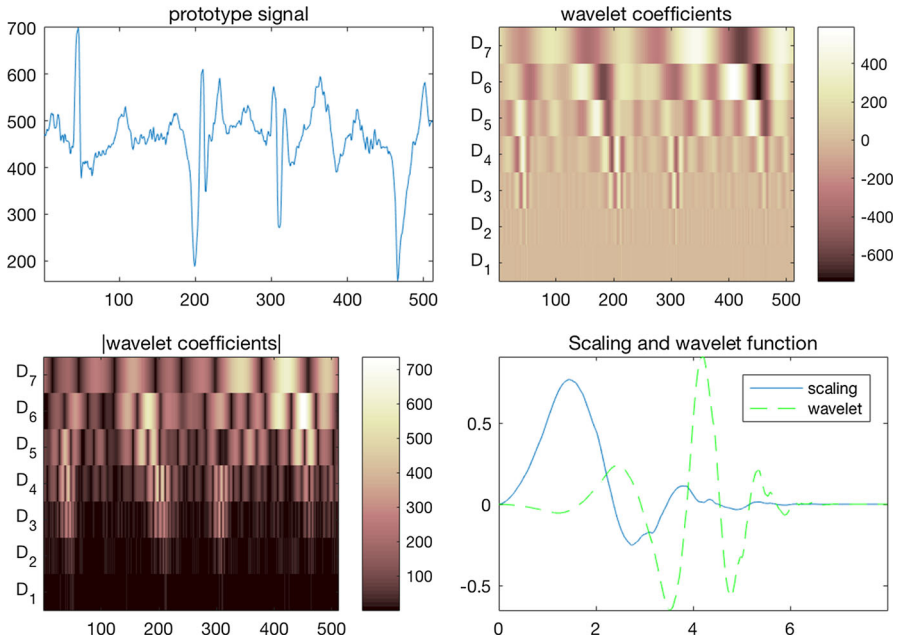


Fig. 11 Example 3, design of a matched wavelet with and without masking for an ECG signal. Shown is the optimization of the undecimated V_4 criterion without masking, with 2 free parameters and two vanishing moments

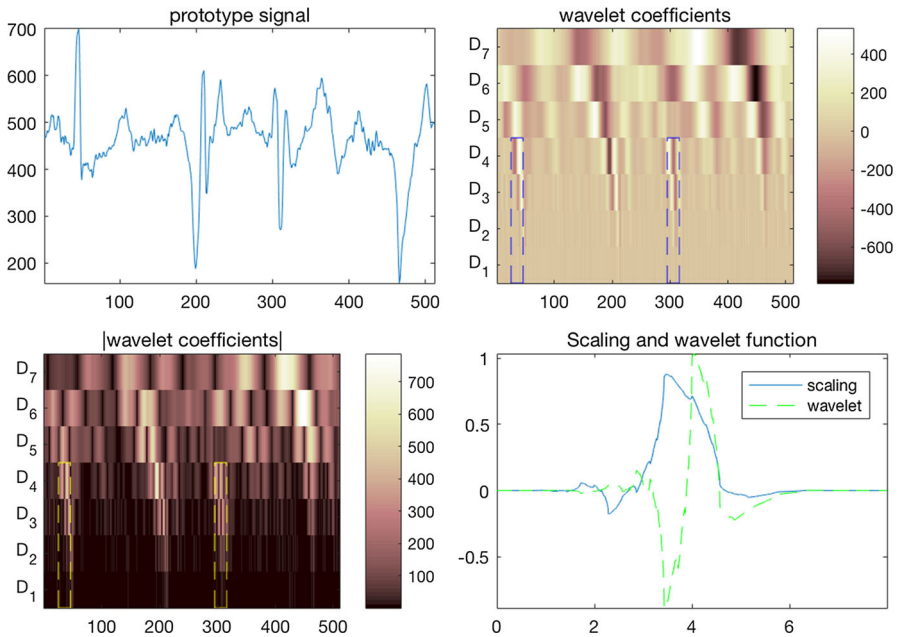


Fig. 12 Example 3, as in Fig. 11 but with masking. The mask is indicated by the dashed boxes. It pronounces the QRS complex in the fine scales

complexity in a different way, as more complicated constraints are then to be satisfied. Here we considered $p \leq 3$, though conditions can be derived for higher values too. When $n - p$ (the degrees of freedom) does not exceed 3, optimization is still easily supported by visualization of the criterion. This may help to detect and avoid local optima. Otherwise one may have to use multiple starting points, for which one needs to decide on how to choose them.

Second, the user needs a prototype signal to match the wavelets too. For some applications, this may be more straightforward than for others. Obviously, the choice of a prototype signal may have an impact on the matched wavelet that is obtained, so it could be advisable to try a few different ones and compare performance of the resulting matched wavelets.

Third, in our setup of orthogonal wavelets the user needs to decide whether to minimize V_1 or to maximize V_4 , to use the critically sampled or the undecimated transforms, and also if using weights (or masking) seems useful. Though some of these choices might be clear from the application one has in mind, it may again require some trial and error before satisfactory results are obtained.

The approach evolves around building orthonormal bases and as such the proposed design criteria assume conservation of energy. Therefore, to apply similar ideas to the biorthogonal wavelet case, one would have to come up with an optimization criterion which takes this non-conservation into account and somehow compensates for it, for instance through appropriate weighting.

The approach also has some advantages that we would like to highlight. Most importantly, the approach has more flexibility than if one restricts to using the standard classes available through software toolboxes and it provides a method to explore this in a systematic way. Next, sparsity appears to be a natural criterion to aim for in compression and detection applications, and this framework aims to directly work with it. In the examples, we presented, the sparsity concept seems to lead to useful matched wavelets. Finally, because several well-known classes of orthogonal wavelets, such as Daubechies wavelets and Coiflets, are naturally included in this framework, it becomes better possible to compare the performance of an optimized wavelet to any one from such classes, than merely by properties such as vanishing moments and how they affect polynomial signals. One may then for instance still decide to use a wavelet from a standard class when it is shown to be only a little suboptimal.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

Appendix: Proofs

Proof of Proposition 1 For part (a), a proof is straightforward and can be found in the literature; see, e.g., [30]. For part (b) note that $\Lambda(1) = I_2$ implies

$$H(1) = R(\theta_n) \cdots R(\theta_1) \Lambda(-1) = R(\theta_1 + \cdots + \theta_n) \Lambda(-1)$$

which is then combined with part (a) to give Eq. (16). □

Proof of Proposition 2 From relationship (5) we have by differentiation:

$$\begin{pmatrix} C'(z) \\ D'(z) \end{pmatrix} = 2zH'(z^2) \begin{pmatrix} 1 \\ z^{-1} \end{pmatrix} + H(z^2) \begin{pmatrix} 0 \\ -z^{-2} \end{pmatrix}.$$

With $z = 1$ and $z = -1$ this gives:

$$\begin{pmatrix} C'(1) & C'(-1) \\ D'(1) & D'(-1) \end{pmatrix} = 2H'(1) \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix} + H(1) \begin{pmatrix} 0 & 0 \\ -1 & -1 \end{pmatrix}.$$

With Eq. (15) for $H(1)$, condition (a) now follows directly from (25), and condition (b) from (23). They are equivalent, because (25) and (23) are.

Differentiation of definition (17) for $\tilde{H}(z)$ shows that $\tilde{H}^{(m)}(z) = H^{(m)}(z)H(1)^{-1}$ for all $m \in \mathbb{N}$. We may therefore write:

$$\begin{pmatrix} C'(1) & C'(-1) \\ D'(1) & D'(-1) \end{pmatrix} = 2\sqrt{2}\tilde{H}'(1) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} + \frac{1}{2}\sqrt{2} \begin{pmatrix} -1 & -1 \\ 1 & 1 \end{pmatrix},$$

which shows (d) and (e).

From Eq. (12) we have by differentiation:

$$H'(z) = \sum_{k=1}^{n-1} [R(\theta_n)\Lambda(z) \cdots R(\theta_{k+1})\Lambda'(z)R(\theta_k)\Lambda(z) \cdots R(\theta_1)\Lambda(-1)],$$

with $\Lambda'(z) = \begin{pmatrix} 0 & 0 \\ 0 & -z^{-2} \end{pmatrix}$. Using the fact that $\Lambda(1) = I_2$ and the properties of rotation matrices (such as $R(\alpha)R(\beta) = R(\alpha + \beta)$), it follows that

$$H'(1) = \sum_{k=1}^{n-1} \left[R(\theta_{k+1} + \cdots + \theta_n) \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} R(\theta_1 + \cdots + \theta_k) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right].$$

With definition (20) and property (16), which allows us to write $R(\theta_{k+1} + \cdots + \theta_n) = R(\frac{\pi}{4} - \phi_k)$, we find that

$$H'(1) = \sum_{k=1}^{n-1} \begin{pmatrix} \sin(\frac{\pi}{4} - \phi_k) \\ -\cos(\frac{\pi}{4} - \phi_k) \end{pmatrix} (\sin(\phi_k) - \cos(\phi_k)).$$

This gives:

$$H'_{11}(1) - H'_{12}(1) = \sum_{k=1}^{n-1} \sin\left(\frac{\pi}{4} - \phi_k\right) (\sin(\phi_k) + \cos(\phi_k)).$$

Because $\sin(\phi_k) + \cos(\phi_k) = \sqrt{2} \cos(\frac{\pi}{4} - \phi_k)$ and $\sin(\frac{\pi}{4} - \phi_k) \cos(\frac{\pi}{4} - \phi_k) = \frac{1}{2} \cos(2\phi_k)$, it follows that

$$H'_{11}(1) - H'_{12}(1) = \frac{1}{2} \sqrt{2} \sum_{k=1}^{n-1} \cos(2\phi_k).$$

According to condition (a) this equals $-\frac{1}{4} \sqrt{2}$, from which (f) now equivalently follows. Finally, with $\cos(2\phi_k) = \sin(2(\frac{\pi}{4} - \phi_k)) = \sin(2(\theta_{k+1} + \dots + \theta_n))$ also (c) follows. □

Proof of Proposition 3 Differentiating relationship (5) twice gives:

$$\begin{pmatrix} C''(z) \\ D''(z) \end{pmatrix} = 4z^2 H''(z^2) \begin{pmatrix} 1 \\ z^{-1} \end{pmatrix} + 2H'(z^2) \begin{pmatrix} 1 \\ -z^{-1} \end{pmatrix} + 2H(z^2) \begin{pmatrix} 0 \\ z^{-3} \end{pmatrix}.$$

With $z = 1$ and $z = -1$ this gives:

$$\begin{pmatrix} C''(1) & C''(-1) \\ D''(1) & D''(-1) \end{pmatrix} = 4H''(1) \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} + 2H'(1) \begin{pmatrix} 1 & 1 \\ -1 & 1 \end{pmatrix} + 2H(1) \begin{pmatrix} 0 & 0 \\ 1 & -1 \end{pmatrix},$$

in which $H(1)$ is given by Eq. (15). Now: $C''(-1) = 0$ takes the form $4H''_{11}(1) - 4H''_{12}(1) + 2H'_{11}(1) + 2H'_{12}(1) - \sqrt{2} = 0$, yielding (a). Likewise, $D''(1) = 0$ gives: $4H''_{21}(1) + 4H''_{22}(1) + 2H'_{21}(1) - 2H'_{22}(1) - \sqrt{2} = 0$, yielding (b).

From Eq. (17) we have that $\tilde{H}^{(m)}(z) = H^{(m)}(z)H(1)^{-1}$ for all $m \in \mathbb{N}$. Therefore: $H''_{11}(1) - H''_{12}(1) = \sqrt{2}\tilde{H}''_{12}(1)$ and $H'_{11}(1) + H'_{12}(1) = \sqrt{2}\tilde{H}'_{11}(1)$, which shows that (d) is equivalent to (a). Also: $H''_{21}(1) + H''_{22}(1) = \sqrt{2}\tilde{H}''_{21}(1)$ and $H'_{21}(1) - H'_{22}(1) = \sqrt{2}\tilde{H}'_{22}(1)$, whence (e) is equivalent to (b).

To address statements (c) and (f), we differentiate the expression for $H'(z)$ obtained in the proof of Proposition 2. This gives:

$$\begin{aligned} H''(z) = & \sum_{k=1}^{n-1} [R(\theta_n)\Lambda(z) \cdots R(\theta_{k+1})\Lambda''(z)R(\theta_k)\Lambda(z) \cdots R(\theta_1)\Lambda(-1)] \\ & + 2 \sum_{1 \leq k < \ell \leq n-1} [R(\theta_n)\Lambda(z) \cdots R(\theta_{\ell+1})\Lambda'(z)R(\theta_\ell)\Lambda(z) \\ & \cdots R(\theta_{k+1})\Lambda'(z)R(\theta_k)\Lambda(z) \cdots R(\theta_1)\Lambda(-1)]. \end{aligned}$$

Here we have that $\Lambda''(z) = \begin{pmatrix} 0 & 0 \\ 0 & 2z^{-3} \end{pmatrix}$. Now:

$$\begin{aligned}
 H''(1) = & \sum_{k=1}^{n-1} \left[R(\theta_{k+1} + \dots + \theta_n) \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix} R(\theta_1 + \dots + \theta_k) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right] \\
 & + 2 \sum_{1 \leq k < \ell \leq n-1} \left[R(\theta_{\ell+1} + \dots + \theta_n) \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} R(\theta_{k+1} + \dots + \theta_\ell) \begin{pmatrix} 0 & 0 \\ 0 & -1 \end{pmatrix} \right. \\
 & \left. \times R(\theta_1 + \dots + \theta_k) \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right]
 \end{aligned}$$

which can be rewritten, using the definition (20) and property (16), as:

$$\begin{aligned}
 H''(1) = & 2 \sum_{k=1}^{n-1} \left[\begin{pmatrix} -\sin(\frac{\pi}{4} - \phi_k) \\ \cos(\frac{\pi}{4} - \phi_k) \end{pmatrix} (\sin(\phi_k) - \cos(\phi_k)) \right] \\
 & + 2 \sum_{1 \leq k < \ell \leq n-1} \left[\begin{pmatrix} -\sin(\frac{\pi}{4} - \phi_\ell) \\ \cos(\frac{\pi}{4} - \phi_\ell) \end{pmatrix} \cos(\phi_\ell - \phi_k) (\sin(\phi_k) - \cos(\phi_k)) \right].
 \end{aligned}$$

Therefore: $H''_{11}(1) - H''_{12}(1) = -2 \sum_{k=1}^{n-1} \sin(\frac{\pi}{4} - \phi_k)(\sin(\phi_k) + \cos(\phi_k)) - 2 \sum_{1 \leq k < \ell \leq n-1} \sin(\frac{\pi}{4} - \phi_\ell) \cos(\phi_\ell - \phi_k)(\sin(\phi_k) + \cos(\phi_k))$. For the first term of this expression we can write: $-2\sqrt{2} \sum_{k=1}^{n-1} \sin(\frac{\pi}{4} - \phi_k) \cos(\frac{\pi}{4} - \phi_k) = -\sqrt{2} \sum_{k=1}^{n-1} \sin(\frac{\pi}{2} - 2\phi_k) = -\sqrt{2} \sum_{k=1}^{n-1} \cos(2\phi_k)$, which because of part (f) of Proposition 2 equals $\frac{1}{2}\sqrt{2}$. For the second term we have: $-2\sqrt{2} \sum_{1 \leq k < \ell \leq n-1} \sin(\frac{\pi}{4} - \phi_\ell) \cos(\phi_\ell - \phi_k) \cos(\frac{\pi}{4} - \phi_k) = -\sqrt{2} \sum_{1 \leq k < \ell \leq n-1} \cos(\phi_\ell - \phi_k)(\sin(\frac{\pi}{2} - \phi_\ell - \phi_k) + \sin(-\phi_\ell + \phi_k)) = -\sqrt{2} \sum_{1 \leq k < \ell \leq n-1} \cos(\phi_\ell - \phi_k) \cos(\phi_\ell + \phi_k) + \frac{1}{2}\sqrt{2} \sum_{1 \leq k < \ell \leq n-1} \sin(2(\phi_\ell - \phi_k))$. Here we have that $\cos(\phi_\ell - \phi_k) \cos(\phi_\ell + \phi_k) = \frac{1}{2}(\cos(2\phi_\ell) + \cos(2\phi_k))$, and it holds that $\sum_{1 \leq k < \ell \leq n-1} (\cos(2\phi_\ell) + \cos(2\phi_k)) = (n - 2) \sum_{k=1}^{n-1} \cos(2\phi_k)$, which equals $-\frac{1}{2}(n - 2)$ because of part (f) of Proposition 2. Therefore:

$$H''_{11}(1) - H''_{12}(1) = \frac{1}{2}\sqrt{2} + \frac{1}{4}(n - 2)\sqrt{2} + \frac{1}{2}\sqrt{2} \sum_{1 \leq k < \ell \leq n-1} \sin(2(\phi_\ell - \phi_k)).$$

Also we have that $\frac{1}{2}(H'_{11}(1) + H'_{12}(1)) = \frac{1}{2} \sum_{k=1}^{n-1} \sin(\frac{\pi}{4} - \phi_k)(\sin(\phi_k) - \cos(\phi_k))$, in which $\sin(\phi_k) - \cos(\phi_k) = -\sqrt{2} \sin(\frac{\pi}{4} - \phi_k)$. With the identity $\sin^2(\frac{\pi}{4} - \phi_k) = \frac{1}{2} - \frac{1}{2} \sin(2\phi_k)$ it follows that:

$$\frac{1}{2}(H'_{11}(1) + H'_{12}(1)) = -\frac{1}{4}(n - 1)\sqrt{2} + \frac{1}{4}\sqrt{2} \sum_{k=1}^{n-1} \sin(2\phi_k).$$

Taking those expressions together, it follows from part (a) that $\frac{1}{4}\sqrt{2} + \frac{1}{2}\sqrt{2} \sum_{1 \leq k < \ell \leq n-1} \sin(2(\phi_\ell - \phi_k)) + \frac{1}{4}\sqrt{2} \sum_{k=1}^{n-1} \sin(2\phi_k) = \frac{1}{4}\sqrt{2}$, which implies (f). Finally, the equivalence of (c) and (f) follows from the definition (20) and property (16). □

- Proof of Theorem 1* (1) Let $E = \sum_{k=1}^m |s_k|^2 = \sum_{k=1}^m |w_k|^2$ be the energy contained in the signal s . The variance of the vector of absolute values $\{|w_k|\}$ is given by $\frac{1}{m} \sum_{k=1}^m |w_k|^2 - \left(\frac{1}{m} \sum_{k=1}^m |w_k|\right)^2 = \frac{1}{m} E - \frac{1}{m^2} \|w\|_1^2 = \frac{1}{m} E - \frac{1}{m^2} V_1^2$. Since E and m are constant, maximization of this quantity with respect to the free filter parameters is equivalent to minimization of V_1 .
- (2) The variance of the vector of energies $\{|w_k|^2\}$ is given by $\frac{1}{m} \sum_{k=1}^m |w_k|^4 - \left(\frac{1}{m} \sum_{k=1}^m |w_k|^2\right)^2 = \frac{1}{m} V_4^2 - \frac{E^2}{m^2}$, in which E and m are constant regardless of the choice of filter parameters. Hence maximization of this quantity is equivalent to maximization of V_4 . □

References

1. D. Alpay, L. Baratchart, A. Gombani, in *On the Differential Structure of Matrix-valued Rational Inner Functions* ed. by A. Feintuch, I. Gohberg, Nonselfadjoint Operators and Related Topics, Operator Theory: Advances and Applications, vol 73 (Birkhuser, Basel, 1994), pp. 30–66 . https://doi.org/10.1007/978-3-0348-8522-5_2
2. N. Ansari, A. Gupta, Signal-matched wavelet design via lifting using optimization techniques, in *IEEE International Conference on Digital Signal Processing (DSP)*, pp. 863–867 (2015). <https://doi.org/10.1109/ICDSP.2015.7251999>
3. S. Bugenhagen, A. Cowley Jr., D. Beard, Identifying physiological origins of baroreflex dysfunction in salt-sensitive hypertension in the Dahl SS rat. *Physiol. Genomics* **42**(1), 23–41 (2010). <https://doi.org/10.1152/physiolgenomics.00027.2010>
4. E.J. Candès, M. Wakin, An introduction to compressive sampling. *IEEE Signal Process. Mag.* **25**(2), 21–30 (2008). <https://doi.org/10.1109/MSP.2007.914731>
5. E.J. Candès, M. Wakin, S. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization. *J. Fourier Anal. Appl.* **14**(5), 877–905 (2007). <https://doi.org/10.1007/s00041-008-9045-x>
6. J. Chapa, R. Rao, Algorithms for designing wavelets to match a specified signal. *IEEE Trans. Signal Process.* **48**(12), 3395–3406 (2000). <https://doi.org/10.1109/78.887001>
7. I. Daubechies, Orthonormal bases of compactly supported wavelets. *Commun. Pure Appl. Math.* **41**(7), 909–996 (1988). <https://doi.org/10.1002/cpa.3160410705>
8. I. Daubechies, J. Lagarias, Two-scale difference equations. II. Local regularity, infinite products of matrices and fractals. *SIAM J. Math. Anal.* **24**(4), 1031–1079 (1992). <https://doi.org/10.1137/0523059>
9. D.L. Donoho, For most underdetermined systems of linear equations, the minimal ℓ^1 -norm near-solution approximates the sparsest near-solution. *Commun. Pure Appl. Math.* **59**(6), 797–829 (2006). <https://doi.org/10.1002/cpa.20132>
10. D.L. Donoho, M. Elad, Optimally sparse representation from overcomplete dictionaries via ℓ^1 norm minimization. *Proc. Natl. Acad. Sci. USA* **100**(5), 2197–2202 (2003). <https://doi.org/10.1073/pnas.0437847100>
11. A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C. Peng, H. Stanley, Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation* **101**(23), e215–e220 (2000). <https://doi.org/10.1161/01.CIR.101.23.e215>
12. A. Gupta, S.D. Joshi, S. Prasad, A new method of estimating wavelet with desired features from a given signal. *Sig. Process.* **85**(1), 147–161 (2005). <https://doi.org/10.1016/j.sigpro.2004.09.008>
13. B. Hanzon, M. Olivi, R. Peeters, Balanced realizations of discrete-time stable all-pass systems and the tangential Schur algorithm. *Linear Algebra Appl.* **418**(2–3), 793–820 (2006). <https://doi.org/10.1016/j.laa.2006.03.027>
14. C. Heil, G. Strang, *Continuity of the joint spectral radius: application to wavelets*, ed. A. Bojanczyk, G. Cybenko *Linear Algebra for Signal Processing*, IMA Vol. Math. Appl., vol. 69 (Springer, New York, 1995), pp. 51–61. https://doi.org/10.1007/978-1-4612-4228-4_4
15. J. Karel. A wavelet approach to cardiac signal processing for low-power hardware applications. Ph.D. thesis, Maastricht University (2009). ISBN: 978-90-5278-887-6

16. J. Karel, R. Peeters, R. Westra, K. Moermans, S. Haddad, W. Serdijn, Optimal discrete wavelet design for cardiac signal processing, in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 2769–2772 (2005). <https://doi.org/10.1109/IEMBS.2005.1617046>
17. J. Karel, R. Senden, J. Janssen, H. Savelberg, B. Grimm, I. Heyligers, R. Peeters, K. Meijer, Towards unobtrusive in vivo monitoring of patients prone to falling, in *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 5018–5021 (2010). <https://doi.org/10.1109/IEMBS.2010.5626232>
18. S. Mallat, Multiresolution approximations and wavelet orthonormal bases for $L^2(\mathbb{R})$. *Trans. Am. Math. Soc.* **315**(1), 69–87 (1989)
19. S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation. *IEEE Trans. Pattern Anal.* **11**(7), 674–693 (1989). <https://doi.org/10.1109/34.192463>
20. S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*, 3rd edn. (Academic Press, Cambridge, 2008)
21. M. Mansour, On the design of matched orthogonal wavelets with compact support, in *IEEE International Conference Acoustics, Speech and Signal Processing (ICASSP)*, pp. 4388–4391 (2011). <https://doi.org/10.1109/ICASSP.2011.5947326>
22. D. Monro, B. Bassil, G. Dickson, Orthonormal wavelets with balanced uncertainty. *Proc. Int. Conf. Image Proc.* **2**, 581–584 (1996). <https://doi.org/10.1109/ICIP.1996.560928>
23. S. Murugesan, D.B.H. Tay, Design of almost symmetric orthogonal wavelet filter bank via direct optimization. *IEEE Trans. Image Proc.* **21**(5), 2474–2480 (2012). <https://doi.org/10.1109/TIP.2012.2188037>
24. G.P. Nason, B.W. Silverman, in *The Stationary Wavelet Transform and Some Statistical Applications*, ed. A. Antoniadis, G. Oppenheim, Wavelets and Statistics, Lecture Notes in Statistics, vol. 103 (Springer, New York, 1995), pp. 281–299. https://doi.org/10.1007/978-1-4612-2544-7_17
25. R. Peeters, M. Olivi, B. Hanzon, Parametrization of matrix-valued lossless functions based on boundary interpolation, in *Proceedings of 19th International Symposium on Mathematical Theory Networks and Systems (MTNS)*, pp. 563–570 (2010)
26. V.P. Potapov, The multiplicative structure of J-contractive matrix functions. *Trudy Moskov. Mat. Obšč* **4**, 125–236 (1955)
27. P. Ravier, O. Buttelli, Robust detection of QRS complex using Klauder wavelets, in *XII. European Signal Processing Conference (EUSIPCO)*, pp. 2199–2202 (2004)
28. S. Saxena, V. Kumar, S. Hamde, QRS detection using new wavelets. *J. Med. Eng. Technol* **26**(1), 7–15 (2002). <https://doi.org/10.1080/03091900110096038>
29. J.L. Starck, J. Fadili, F. Murtagh, The undecimated wavelet decomposition and its reconstruction. *IEEE Trans. Image Proc.* **16**(2), 297–309 (2007). <https://doi.org/10.1109/TIP.2006.887733>
30. G. Strang, T. Nguyen, *Wavelets and Filter Banks* (Wellesley-Cambridge Press, Wellesley, MA, 1996)
31. W. Sweldens, The lifting scheme: a custom-design construction of biorthogonal wavelets. *Appl. Comput. Harmon. Anal.* **3**(2), 186–200 (1996). <https://doi.org/10.1006/acha.1996.0015>
32. P. Vaidyanathan, Theory and design of M-channel maximally decimated quadrature mirror filters with arbitrary M, having the perfect-reconstruction property. *IEEE Trans. Acoust. Speech Signal Proc.* **35**(4), 476–492 (1987). <https://doi.org/10.1109/TASSP.1987.1165155>
33. P. Vaidyanathan, Z. Doğanata, The role of lossless systems in modern digital signal processing: a tutorial. *IEEE Trans. Educ.* **32**(3), 181–197 (1989). <https://doi.org/10.1109/13.34150>