21st International Symposium on
Mathematical Theory of Networks and Systems
July 7-11, 2014. Groningen, The Netherlands

Data driven design of an orthogonal wavelet with vanishing moments

Ralf Peeters[1] and Joël Karel[1]

Abstract— **We present a framework to design an orthogonal wavelet with compact support and vanishing moments, tuned to a given application. This is achieved by optimizing a criterion, such that a prototype signal, which is characteristic for the application, becomes sparse in the wavelet domain. This approach is beneficial for compression and detection purposes. A parameterization is developed for which orthogonality and compact support are built in, and in terms of which we can express the vanishing moment conditions conveniently. The approach is developed for critically sampled wavelet transforms as well as for the stationary wavelet transform. Several examples illustrate the methods.**

I. Introduction

Wavelets have gained popularity over the past few decades and have proven to be useful for processing both stationary and non-stationary signals [1], [2]. Compared to Fourier transforms, wavelets have the benefit of allowing for a *choice of basis*. When one has decided on employing wavelets for a signal processing problem there is the problem of deciding which wavelet to use [3] for the problem at hand. This choice can be driven by the certain properties of a wavelet, such as having as many vanishing moments as possible [4] and/or being orthogonal. Some authors employ a parameterized wavelet and tune it to the application at hand [5]. In practice, one often employs a range of wavelets and then proceeds to use the subset that provides the most promising results.

Yet another way to proceed is to use a data driven design for wavelets. In recent years a number of approaches have been developed. In [6] the amplitude and phase spectrum of a wavelet are separately matched to that of a reference signal in the Fourier domain. In [7] a design procedure for bi- and semi-orthogonal wavelets is discussed where the optimization criterion aims to maximize the energy in the approximation coefficients, but without a clear motivation.

In this paper an approach is discussed to design compactly supported orthogonal wavelets from filter banks in polyphase, having vanishing moments. Two design criteria are given that maximize sparsity and allow for a data driven design: one by minimizing an L^1-norm, the other by maximizing an L^4-norm. These two criteria are motivated by the principle of maximization of variance. The criteria are developed for the critically sampled wavelet transform, but can be adapted for use with the stationary wavelet transform by employing weighing. To build in orthogonality, compact support, and to enforce 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 which allows one to handle them conveniently. For low orders this implies that one does not have to resort to constrained optimization with strongly nonlinear constraints.

The paper is organized as follows. In Section II we introduce wavelets from filter banks, the polyphase representation, orthogonal wavelets, and their connection to lossless polyphase filters. Then parameterizations of FIR lossless filters are discussed in Section III. In Section IV we express the vanishing moment conditions in terms of these parameterizations. A few special cases, including the manifold structure when vanishing moments up to order 1 are built in, are discussed in Section V. Sparsity criteria for orthogonal wavelet design are presented in Section VI, together with a number of optimization considerations. Several examples are discussed in Section VII and conclusions are given in Section VIII. Proofs are largely omitted because of space limitations; they will be presented elsewhere.

II. Wavelets from filter banks

One particularly useful approach to wavelet analysis which we now briefly review, is by performing signal processing with filter banks. Many more details can be found in the literature, in particular in [2] on which much of the present section is based. In this set-up, a discrete-time signal is processed by a cascade of filters in the following way. In the first stage of the cascade, the signal is passed in parallel through both a FIR low-pass filter $C(z) = c_0 + c_1 z^{-1} + \ldots + c_{2n-1} z^{-(2n-1)}$ and a FIR high-pass filter $D(z) = d_0 + d_1 z^{-1} + \ldots + d_{2n-1} z^{-(2n-1)}$. Then dyadic down-sampling $\downarrow 2$ is applied, discarding all the odd-indexed values from the signals. The two resulting signals represent an approximation signal (on a coarser scale) and a detail signal (on the finest scale). Then in any further stage of the cascade, the procedure is repeated by processing the current approximation signal in the same way, using the same filter bank. This produces approximations and detail signals at increasingly coarse levels, yielding what is called a critically sampled multi-scale decomposition.

Wavelets are associated with this filtering approach by considering solutions to two equations. The *dilation equation* is a refinement equation, given by

$$\phi(t) = \sqrt{2} \sum_{k=0}^{2n-1} c_k \phi(2t+k). \tag{1}$$

The *wavelet equation* is given by

$$\psi(t) = \sqrt{2} \sum_{k=0}^{2n-1} d_k \phi(2t+k). \tag{2}$$

[1]The authors are with the Department of Knowledge Engineering, Maastricht University, P.O. Box 616, 6200 MD Maastricht, The Netherlands, {ralf.peeters,joel.karel}@maastrichtuniversity.nl

ISBN: 978-90-367-6321-9

Note that the plus signs in these equations are slightly non-standard; they make it possible to link the wavelets directly to signal processing, which would otherwise need some additional notation. For a valid wavelet interpretation, the dilation equation must have an integrable solution $\phi(t)$, the scaling function, with a nonzero integral. This requires that $\sum_{k=0}^{2n-1} c_k = \sqrt{2}$. Also, the wavelet function $\psi(t)$ must have a zero integral, whence $\sum_{k=0}^{2n-1} d_k = 0$. In this paper we focus on *orthogonal wavelets* from filter banks. This additionally imposes the so-called *double-shift orthogonality* conditions on the filter coefficients c_k and d_k. At each filtering stage, orthogonality between the detail space (which is spanned by the wavelet and its integer translates) and the approximation space (which is spanned by the scaling function and its integer translates) is captured by the well-known *alternating flip condition* which relates the high-pass filter coefficients to the low-pass filter coefficients by

$$d_k = (-1)^k c_{2n-1-k}, \qquad (k = 0, 1, \ldots, 2n-1). \tag{3}$$

The remaining $n-1$ double-shift orthogonality conditions on the low-pass filter coefficients are then specified by:

$$\sum_{k=0}^{2n-1-2\ell} c_k c_{2\ell+k} = 0, \qquad (\ell = 1, \ldots, n-1). \tag{4}$$

For wavelet orthonormality we have the additional normalizing condition

$$\sum_{k=0}^{2n-1} c_k^2 = 1. \tag{5}$$

For an orthogonal FIR filter bank in this set-up with a given filter order $2n-1$, we then have $n-1$ remaining degrees of freedom for the filter coefficients c_k. This freedom can be used to build in additional wavelet properties, and to adapt the wavelet to the application at hand. The most well-known kind of desirable wavelet properties concerns *vanishing moments*. When a wavelet has vanishing moments up to order p, then it has smoothness by being at least $p-1$ times continuously differentiable. Also, polynomials up to degree p are contained in the approximation spaces and so they do not show up in the detail spaces. Other desirable properties that have been studied in the literature are linear phase, and symmetry. However, in this paper we shall be exclusively concerned with building in a few additional vanishing moments and then optimizing over the remaining degrees of freedom to make the wavelet fit well to an application. Vanishing moments are further discussed in Section IV and optimization criteria in Section VI.

The orthogonality conditions can be conveniently incorporated into a parameterization by switching to a polyphase filter representation. This will avoid that one has to specify the nonlinear double-shift orthogonality constraints when optimization is to be performed. The polyphase filter for a filter bank with the filters $C(z)$ and $D(z)$ is a 2×2 FIR filter $H(z) = H_0 + H_1 z^{-1} + \ldots + H_{n-1} z^{-(n-1)}$ of order $n-1$, given by:

$$H(z) = \sum_{k=0}^{n-1} \begin{pmatrix} c_{2k} & c_{2k+1} \\ d_{2k} & d_{2k+1} \end{pmatrix} z^{-k}. \tag{6}$$

If the two input signals to $H(z)$ are chosen to be the even phase and the delayed odd phase of the discrete time signal to be processed, then the two output signals coincide with the approximation and detail signals in the first stage of the earlier set-up. The wavelet orthogonality conditions now happen to correspond to $H(z)$ being a lossless filter: for every z on the unit circle the matrix $H(z)$ must be unitary. Lossless filters have been extensively studied in the literature and various parameterizations are available. In Section III we present more details. For now, note that to have a valid orthonormal wavelet interpretation and to comply with the alternating flip condition, it needs to hold that

$$H(1) = \frac{1}{2}\sqrt{2} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \tag{7}$$

This follows from Eqn. (6) and the earlier conditions on top of this page, together with unitarity of $H(1)$.

III. Parameterization of lossless FIR polyphase filters

Several parameterizations of lossless systems $H(z)$ are available in the literature. Constructions by means of the tangential Schur algorithm, which are recursive in the system order and which use interpolation conditions, are described in [8], [9]. Such parameterizations allow one to enforce vanishing moment conditions by imposing interpolation conditions at the stability boundary, see [10]. Also, by choosing interpolation points at the origin, one can impose the FIR property. In this way one arrives at a Potapov factorization of $H(z)$, see also [11]. However, the combination of the FIR property and several vanishing moments jointly is hard to achieve. For the 2×2 case that we restrict to here, we therefore first present a commonly used parameterization for FIR lossless systems (without any vanishing moment conditions), see also [12], [2]. Then we propose a new parameterization which resembles the Potapov factorization more closely and which we have found to be better suited for expressing the vanishing moment conditions. It is used for that very purpose in Section IV.

A first well-known parameterization, see [2], employs the following decomposition of $H(z)$:

$$H(z) = R(\theta_n)\Lambda(z)\cdots\Lambda(z)R(\theta_2)\Lambda(z)R(\theta_1)\Lambda(-1), \tag{8}$$

in which $R(\theta)$ denotes a rotation matrix for the angle θ, and $\Lambda(z)$ is a lossless filter of order 1 which performs a time-delay on the second input channel only:

$$R(\theta) = \begin{pmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{pmatrix}, \qquad \Lambda(z) = \begin{pmatrix} 1 & 0 \\ 0 & z^{-1} \end{pmatrix}. \tag{9}$$

The right multiplying factor $\Lambda(-1)$ is included to comply with the alternating flip condition. To allow for a wavelet

interpretation it is required that $H(1)$ has the value specified in Eqn. (7), which holds if and only if

$$\theta_1 + \theta_2 + \ldots + \theta_n = \frac{\pi}{4} (\text{mod } 2\pi). \tag{10}$$

See also [2], [13]. This allows one to eliminate one parameter by expressing it in terms of the other ones. To parameterize the space of all 2×2 lossless FIR filters with this admissible value for $H(1)$, each of the remaining $n-1$ angular parameters θ_k can be chosen from an arbitrary interval of length π. In some of the literature the parameter θ_1 is eliminated, but here we find it convenient to eliminate the parameter θ_n, by expressing it as

$$\theta_n = \frac{\pi}{4} - (\theta_1 + \theta_2 + \ldots + \theta_{n-1}) (\text{mod } 2\pi). \tag{11}$$

An advantage of this, is that it becomes easier to change the filter order, and in this way to embed low order lossless FIR filters more nicely in the space of lossless FIR filters of a high order. As it happens, the vanishing moment conditions up to a given order p are not easily expressed in terms this parameterization with parameters $\theta_1, \ldots, \theta_{n-1}$. We propose the following alternative parameterization, which will turn out to be more convenient. It employs the new parameters $\phi_1, \ldots, \phi_{n-1}$ which are defined by:

$$\phi_k = \theta_1 + \ldots + \theta_k, \qquad (k = 1, 2, \ldots, n-1). \tag{12}$$

In addition, we shift attention from $H(z)$ to the lossless FIR filter $\tilde{H}(z) = H(z)H(1)^{-1}$. We then have the following result.

Proposition 1: The space of all 2×2 lossless FIR filters $H(z)$ of order $n-1$ with $H(1)$ satisfying Eqn. (7), is given by $H(z) = \tilde{H}(z)H(1)$ in which $\tilde{H}(z)$ admits the factored form

$$\tilde{H}(z) = B_{\phi_{n-1}}(z) \cdots B_{\phi_2}(z) B_{\phi_1}(z), \tag{13}$$

where $B_{\phi_k}(z)$ $(k = 1, \ldots, n-1)$ denotes an elementary lossless FIR filter given by

$$B_{\phi_k}(z) = I_2 + (z^{-1} - 1) V_{\phi_k}, \tag{14}$$

with

$$V_{\phi_k} = \frac{1}{2} \left(I_2 - \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} R(2\phi_k) \right). \tag{15}$$

Here each parameter ϕ_k ranges over an arbitrary half-open interval of length π.

Proof. The proof of this result starts from Eqn. (8) by rewriting it in terms of the new parameterization. Note that $H(1) = H(1)^T = H(1)^{-1}$. Then the result is obtained by noting that the elementary lossless factors $B_{\phi_k}(z)$ are alternatively given by $H(1)R(\phi_k)\Lambda(z)R(\phi_k)^T H(1)$. □

The factors $B_{\phi_k}(z)$ are Potapov factors of $\tilde{H}(z)$: they are lossless of order 1 having their pole at $z = 0$ and satisfying $B_{\phi_k}(1) = I_2$. These Potapov factors are conventionally represented in the form $B_{\phi_k}(z) = I_2 + (z^{-1} - 1) u_{\phi_k} u_{\phi_k}^T$ in which u_{ϕ_k} is a vector of norm 1. In this case we may choose $u_{\phi_k} = \begin{pmatrix} \sin(\phi_k - \frac{\pi}{4}) \\ \cos(\phi_k - \frac{\pi}{4}) \end{pmatrix}$.

IV. Vanishing moment conditions

By definition, a wavelet $\psi(t)$ has vanishing moments up to order p if it holds that

$$\int_{-\infty}^{\infty} t^m \psi(t) dt = 0, \qquad (m = 0, 1, \ldots, p). \tag{16}$$

This is well-known to be equivalent to the high-pass filter $D(z)$ having vanishing moments up to order p. Because the filters $C(z)$ and $D(z)$ are intimately related, this can also be cast into the form of conditions on $C(z)$. This gives the following equivalent sets of vanishing moment conditions (for $m = 0, 1, \ldots, p$): $\sum_{k=0}^{2n-1} k^m d_k = 0$, and $\sum_{k=0}^{2n-1} k^m (-1)^k c_k = 0$. In terms of the derivatives of the filters (where the m-th derivative with respect to z is denoted by a superscript $^{(m)}$) it equivalently needs to hold (for $m = 0, 1, \ldots, p$):

$$D^{(m)}(1) = 0, \qquad \text{or} \qquad C^{(m)}(-1) = 0. \tag{17}$$

To rewrite these conditions in terms $H(z)$ and its derivatives, is possible by exploiting the relationship (cf. Eqn. (6)):

$$\begin{pmatrix} C(z) \\ D(z) \end{pmatrix} = H(z^2) \begin{pmatrix} 1 \\ z^{-1} \end{pmatrix}. \tag{18}$$

For $m = 0$ the corresponding condition is implied by the condition on $H(1)$. For $m = 1, \ldots, p$ we have the following result.

Proposition 2: The m-th order derivatives of $C(z)$ and $D(z)$ are given in terms of the derivatives of $H(z)$ of orders $0, 1, \ldots, m$ by:

$$\begin{pmatrix} C^{(m)}(z) \\ D^{(m)}(z) \end{pmatrix} = \sum_{k=0}^{m} \sum_{\ell=0}^{\lfloor k/2 \rfloor} \frac{m!}{\ell!(k-2\ell)!} (2z)^{k-2\ell} H^{(k-\ell)}(z^2) \times \tag{19}$$
$$\times \left((-1)^{m-k} z^{-(m-k+1)} \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \delta_{m,k} \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right)$$

where $\delta_{m,k}$ denotes the Kronecker delta.

Proof. This follows from Eqn. (18) by a combination of three results. (i) The k-th order derivative of a product of two functions is given by: $(A(z)B(z))^{(k)} = \sum_{\ell=0}^{k} \begin{pmatrix} k \\ \ell \end{pmatrix} A^{(\ell)} B^{(k-\ell)}$. This is standard. (ii) The k-th order derivative of the function $H(z^2)$ with respect to z, is given by: $\frac{d^k}{dz^k}(H(z^2)) = \sum_{\ell=0}^{\lfloor k/2 \rfloor} \frac{k!}{\ell!(k-2\ell)!} (2z)^{k-2\ell} H^{(k-\ell)}(z^2)$. This can be proved by induction. (iii) The k-th order derivative of the vector $\begin{pmatrix} 1 \\ z^{-1} \end{pmatrix}$ with respect to z is straightforward to compute. It is given by $(-1)^k k! z^{-(k+1)} \begin{pmatrix} 0 \\ 1 \end{pmatrix} + \delta_{k,0} \begin{pmatrix} 1 \\ 0 \end{pmatrix}$. □

By choosing $z = 1$ in the expression for $D^{(m)}(z)$, or $z = -1$ in the expression for $C^{(m)}(z)$, the vanishing moment conditions are obtained in terms of the entries of the derivatives of $H(z)$ at $z = 1$. By switching to $\tilde{H}(z)$ and its derivatives, we have the following key result.

Theorem 1: The vanishing moment conditions up to the order p are given by the matrix-vector equation:

$$\begin{pmatrix} \tilde{H}_{21}^{(1)}(1) \\ \tilde{H}_{21}^{(2)}(1) \\ \vdots \\ \tilde{H}_{21}^{(p)}(1) \end{pmatrix} = \begin{pmatrix} L_{1,1} & 0 & \cdots & 0 \\ L_{2,1} & L_{2,2} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ L_{p,1} & L_{p,2} & \cdots & L_{p,p} \end{pmatrix} \begin{pmatrix} 1 \\ \tilde{H}_{22}^{(1)}(1) \\ \vdots \\ \tilde{H}_{22}^{(p-1)}(1) \end{pmatrix}, \tag{20}$$

in which the entries $L_{k,\ell}$ of the lower triangular matrix L are given for $k \geq \ell$ by:

$$L_{k,\ell} = \frac{k!}{(-4)^{k-l+1}(\ell-1)!} C_{k-\ell+1} \tag{21}$$

with $C_k = \frac{(2k)!}{k!(k+1)!}$ denoting the k-th Catalan number.
Equivalent conditions are obtained by replacing $\tilde{H}_{21}^{(m)}$ by $\tilde{H}_{12}^{(m)}$ and $\tilde{H}_{22}^{(m)}$ by $\tilde{H}_{11}^{(m)}$, for all $m = 1, \ldots, p$. Because of space limitations, a proof of the theorem and this alternative statement will be given elsewhere.

To be explicit, the sequence of Catalan numbers starts with $C_0 = 1$, $C_1 = 1$, $C_2 = 2$, $C_3 = 5$, $C_4 = 14$. The vanishing moment conditions up to order $p = 4$ are given by:

$$\tilde{H}_{21}^{(1)}(1) = -\frac{1}{4}, \tag{22}$$

$$\tilde{H}_{21}^{(2)}(1) = \frac{1}{4} - \frac{1}{2}\tilde{H}_{22}^{(1)}(1), \tag{23}$$

$$\tilde{H}_{21}^{(3)}(1) = -\frac{15}{32} + \frac{3}{4}\tilde{H}_{22}^{(1)}(1) - \frac{3}{4}\tilde{H}_{22}^{(2)}(1), \tag{24}$$

$$\tilde{H}_{21}^{(4)}(1) = \frac{21}{16} - \frac{15}{8}\tilde{H}_{22}^{(1)}(1) + \frac{3}{2}\tilde{H}_{22}^{(2)}(1) - \tilde{H}_{22}^{(3)}(1). \tag{25}$$

These conditions can be reworked in terms of the parameterization with $\phi_1, \ldots, \phi_{n-1}$. This involves several steps. Eventually this gives the following result of which we have a lengthy proof which will be presented elsewhere.

Theorem 2: In terms of the angular parameters $\phi_1, \ldots, \phi_{n-1}$, the vanishing moment conditions up to the order p are given by the matrix-vector equation:

$$\begin{pmatrix} \sum_{1 \leq k_1 \leq n-1} \cos(2\phi_{k_1}) \\ -\sum_{1 \leq k_1 < k_2 \leq n-1} \sin(2\phi_{k_2} - 2\phi_{k_1}) \\ \sum_{1 \leq k_1 < k_2 < k_3 \leq n-1} \cos(2\phi_{k_3} - 2\phi_{k_2} + 2\phi_{k_1}) \\ -\sum_{1 \leq k_1 < k_2 < k_3 < k_4 \leq n-1} \sin(2\phi_{k_4} - 2\phi_{k_3} + 2\phi_{k_2} - 2\phi_{k_1}) \\ \vdots \end{pmatrix} =$$

$$= \begin{pmatrix} M_{1,1} & 0 & \cdots & 0 \\ M_{2,1} & M_{2,2} & \ddots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ M_{p,1} & M_{p,2} & \cdots & M_{p,p} \end{pmatrix} \times \tag{26}$$

$$\times \begin{pmatrix} 1 \\ -\sum_{1 \leq k_1 \leq n-1} \sin(2\phi_{k_1}) \\ \sum_{1 \leq k_1 < k_2 \leq n-1} \cos(2\phi_{k_2} - 2\phi_{k_1}) \\ -\sum_{1 \leq k_1 < k_2 < k_3 \leq n-1} \sin(2\phi_{k_3} - 2\phi_{k_2} + 2\phi_{k_1}) \\ \vdots \end{pmatrix}$$

in which the entries $M_{k,\ell}$ of the lower triangular matrix M are given for $k \geq \ell$ by:

$$M_{k,\ell} = \begin{cases} \frac{1}{(-2)^{k-l+1}} C_{(k-\ell)/2}, & \text{if } k-\ell \text{ is even,} \\ 0, & \text{if } k-\ell \text{ is odd,} \end{cases} \tag{27}$$

and where $C_k = \frac{(2k)!}{k!(k+1)!}$ denotes the k-th Catalan number. Explicitly, the vanishing moment conditions up to order $p = 4$ are now given by

- $$1 + 2 \sum_{1 \leq k_1 \leq n-1} \cos(2\phi_{k_1}) = 0, \tag{28}$$
- $$\sum_{1 \leq k_1 \leq n-1} \sin(2\phi_{k_1}) + 2 \sum_{1 \leq k_1 < k_2 \leq n-1} \sin(2\phi_{k_2} - 2\phi_{k_1}) = 0, \tag{29}$$
- $$1 + 4 \sum_{1 \leq k_1 < k_2 \leq n-1} \cos(2\phi_{k_2} - 2\phi_{k_1}) + \\ +8 \sum_{1 \leq k_1 < k_2 < k_3 \leq n-1} \cos(2\phi_{k_3} - 2\phi_{k_2} + 2\phi_{k_1}) = 0, \tag{30}$$
- $$\sum_{1 \leq k_1 \leq n-1} \sin(2\phi_{k_1}) + \\ +4 \sum_{1 \leq k_1 < k_2 < k_3 \leq n-1} \sin(2\phi_{k_3} - 2\phi_{k_2} + 2\phi_{k_1}) + \\ +8 \sum_{1 \leq k_1 < k_2 < k_3 < k_4 \leq n-1} \sin(2\phi_{k_4} - 2\phi_{k_3} + 2\phi_{k_2} - 2\phi_{k_1}) = 0. \tag{31}$$

V. Some special cases

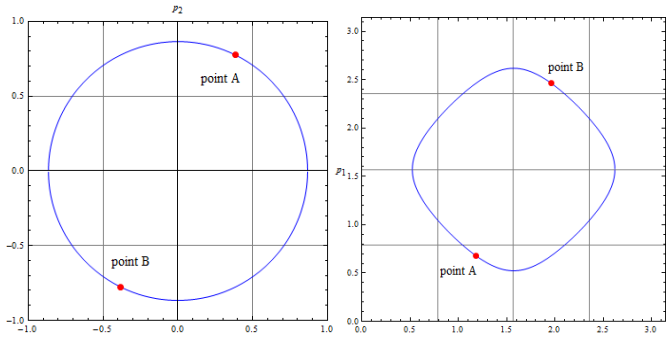
As a first special case, we address the situation in which a single vanishing moment of order $p = 1$ is additionally imposed on the filters $\tilde{H}(z)$. If we now introduce new parameters $q_k = \cos(\phi_k)$ (for $k = 1, \ldots, n-1$), which are all restricted to the interval $[-1, 1]$, then $\cos(2\phi_k) = 2q_k^2 - 1$ and so the vanishing moment constraint takes the form:

$$\sum_{k=1}^{n-1} q_k^2 = \frac{1}{2}n - \frac{3}{4}. \tag{32}$$

It follows that the manifold of lossless FIR polyphase filters $\tilde{H}(z)$ of order $n-1$ with a vanishing moment of order 1, is in fact described by that part of a hypersphere (having its center at the origin and a radius equal to $\frac{1}{2}\sqrt{2n-3}$) which falls inside the hypercube $[-1, 1]^{n-1}$ (having its center at the origin too and sides of length 2).

- For $n = 2$, the manifold consists of the two isolated points $q_1 = \pm\frac{1}{2}$.
- For $n = 3$, the manifold consists of the circle $q_1^2 + q_2^2 = \frac{3}{4}$. It is entirely contained in the square area $[-1, 1] \times [-1, 1]$.
- For $n = 4$, the manifold consists of that part of the sphere $q_1^2 + q_2^2 + q_3^2 = \frac{5}{4}$ which is contained in the cube $[-1, 1]^3$. Because the radius of the sphere exceeds 1, some parts of the sphere get truncated. Topologically we also have that opposite faces of the cube must be identified. It follows that the manifold is in fact a closed oriented surface of genus 3, i.e., it is topologically equivalent to the surface of a sphere with 3 extra handles attached to it.

The resulting manifolds for $n = 3$ for two parameterizations are shown in Figure 1. The manifold for $n = 4$ is shown in Figure 2. For $n \geq 5$ the manifold likewise consists of a



(a) (b)

Fig. 1: Manifolds of orthogonal wavelets with vanishing moments up to order 1 for $n = 3$. Left: Parameterization in terms of $q_1, q_2 \in [-1, 1]$. Right: Parameterization in terms of $\phi_1, \phi_2 \in [0, \pi]$. Points A (= Daubechies 3) and B indicate wavelets with vanishing moments up to order 2.

hypersphere truncated by a hypercube, with the intersections at opposite hyperfaces identified. It is not obvious to give a natural parameterization with $n - 2$ parameters only.

If vanishing moments up to order $p = 2$ are imposed, then the constraint Eqn. (29) must also be satisfied. For $n = 3$, the resulting submanifold consists of two isolated points on the circle $q_1^2 + q_2^2 = \frac{3}{4}$. They occur for $q_2\sqrt{1-q_2^2}(4q_1^2 - 1) + q_1\sqrt{1-q_1^2}(3 - 4q_2^2) = 0$, which gives: $(q_1, q_2) = \pm(\frac{1}{2}\sqrt{-1+\sqrt{\frac{5}{2}}}, \frac{1}{2}\sqrt{4-\sqrt{\frac{5}{2}}})$. They are plotted as points A and B in Figure 1. For $n = 4$, the submanifold consists of two isolated (non-intersecting) closed curves. See Figure 3.

VI. Wavelet Design Criteria and Wavelet Optimization

To choose an appropriate wavelet for a signal processing task is not a trivial issue. Naturally one may try a number of different wavelets from the multitude of wavelets that are available in software toolboxes and then determine experimentally which are the most promising at hand. Practical as this may seem, and even though it appears to be a quite common approach in practice, this is little efficient and it does not take full advantage of the flexibility of wavelet theory. In this paper we restrict to orthogonal wavelets generated by a parameterized FIR filter bank of a given finite order having several additional vanishing moments but still leaving a few degrees of freedom. Then a computationally appealing approach is to select a wavelet from such a class by optimizing a criterion. Here we propose that in applications where one is interested in signal compression or in the detection of features, a useful criterion is one which promotes wavelets which are able to represent typical signals from the application at hand in a sparse way. This has previously also been advocated in [14]. For compression the usefulness of such an approach is obvious. For detection purposes, we note that if a signal is sparsely represented,

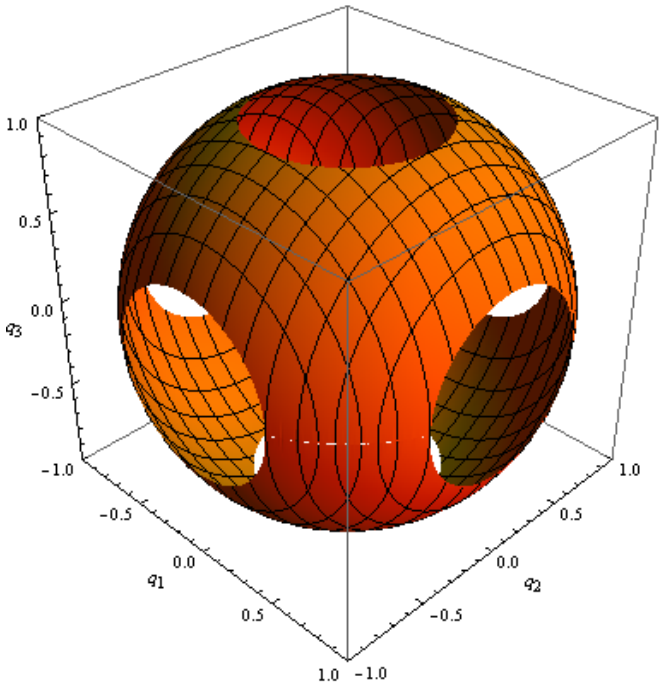


Fig. 2: Manifold of orthogonal wavelets with vanishing moments up to order 1 for $n = 4$. Parameterization in terms of $q_1, \ldots, q_3 \in [-1, 1]$. Opposite faces must be identified.

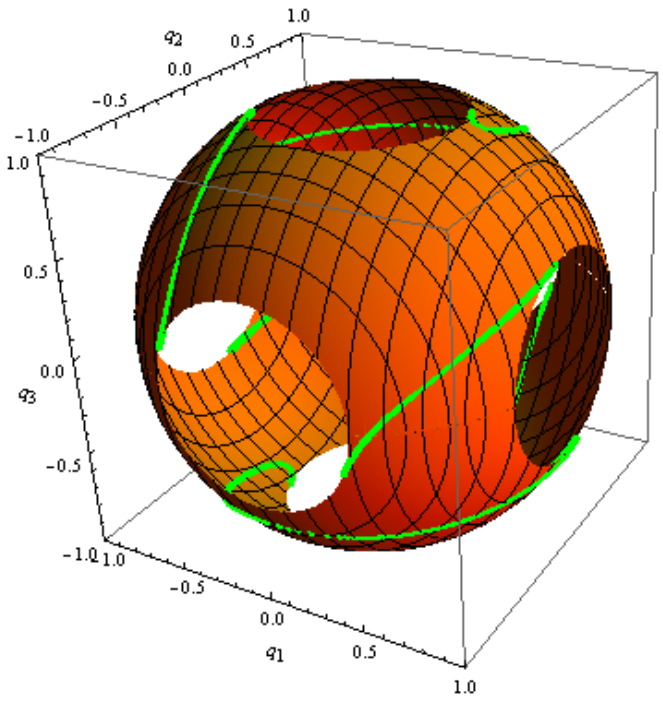
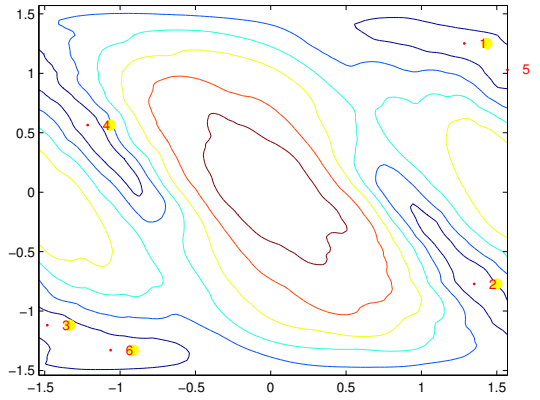


Fig. 3: Manifolds of orthogonal wavelets for $n = 4$. The surface corresponds to vanishing moments up to order 1. The curves on the surface correspond to vanishing moments up to order 2. Opposite faces must be identified; then there are two different closed curves.

then such a representation acts like a 'fingerprint' which can be searched for against a noisy background. Also, when distortions (artifacts, pathologies) occur, they can be searched for by analyzing the difference with a typical pattern.

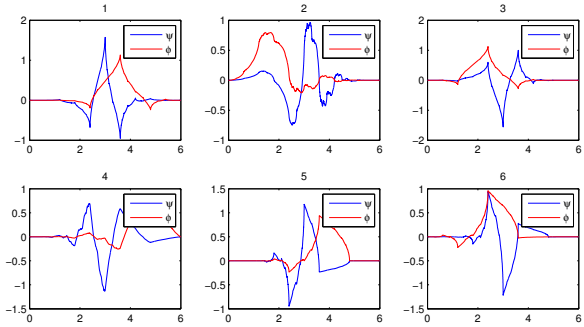
In [15], [13] we have developed several optimization criteria along such lines. They aim to design a wavelet which gives a sparse representation of a given prototype signal. In this sense, wavelet design becomes data driven, as prototype signals are usually not hard to obtain. The unifying principle to achieve sparsity is *maximization of variance*. This works because for *orthogonal* wavelets, the sum of squares of the wavelet coefficients is *fixed* and equals the sum of squares of the signal values. (This constitutes a wavelet analogue of Parseval's law, see [1], as these sums of squares correspond to the energy of a signal, in the wavelet domain and in the time domain, respectively. See also [2], [13].) We have the following result.

Theorem 3: Let s be a discrete-time signal of finite length $m = 2^k$. Using wrap around, let it be processed by a cascade of filters with down-sampling, involving ℓ consecutive levels and an orthogonal filter bank. Let w be the vector of all the wavelet coefficients at the ℓ different levels, together with the final approximation coefficients.
(i) Maximization of the variance of the sequence of absolute values $|w_i|$ of wavelet and approximation coefficients is equivalent to minimization of the L^1-norm $V_1 = \sum_{i=1}^{m} |w_i|$.
(ii) Maximization of the variance of the sequence of energies $|w_i|^2$ is equivalent to maximization of the L^4-norm $V_4 = (\sum_{i=1}^{m} |w_i|^4)^{1/4}$.
A proof of this theorem is given in [13]. Note that minimization of the L^1-norm under (i), is a well-known criterion to achieve sparsity in various other contexts, see for example [16], [17], [18], [19]. Note also that in this basic form, the theorem applies to the *critically sampled wavelet transform*, i.e., for which dyadic down-sampling is performed so that the signal s and its vector of wavelet and approximation coefficients w have the same length m. A drawback of it is that it is not invariant under translation: when the signal s is shifted the vector w changes but does not simply shift along. Another drawback is that the time resolution of the detail and approximation signals is coarser at higher levels (dyadic down-sampling makes that the number of coefficients drops by a factor 2 for every new level). These drawbacks are conveniently handled by undecimated wavelet transforms and the *stationary wavelet transform* [20] in particular. Undecimated wavelet transforms can be computed in various ways, but here we will restrict to the stationary wavelet transform. It follows that the new vector $\tilde{w}$ of all the wavelet and approximation coefficients has length $(\ell + 1)m$ for the stationary wavelet transform on ℓ levels. On the finest scale, there are two different sequences of wavelet coefficients that can be obtained: the even phase and the odd phase of the high-pass filter output before down-sampling. On the next level, there are four sequences (corresponding to the four phases), and so on. They now all show up in the vector $\tilde{w}$. It follows that, for orthogonal wavelets, the vector $\tilde{w}$ obeys a (dyadically weighted) adapted version of Parseval's law. Let



(a)

Fig. 4: L^1-minimization for two parameters ϕ_1 and ϕ_2. Contour lines of the optimization criterion are shown. Red dots mark the six local optima that were obtained.



(a)

Fig. 5: The wavelets $\psi(t)$ and scaling functions $\phi(t)$ corresponding to the six indicated local optima of Figure 4.

$\tilde{w} = \left(b_1^{(1)}, \ldots, b_m^{(1)}, \ldots, b_1^{(\ell)}, \ldots, b_m^{(\ell)}, a_1^{(\ell)}, \ldots, a_m^{(\ell)}\right)$, where $b_i^{(j)}$ denotes the i-th wavelet coefficient (of the stationary wavelet transform) at the j-th level, and $a_i^{(\ell)}$ the i-th approximation coefficient at the last (ℓ-th) level. Then we have:

$$\sum_{j=1}^{\ell} 2^{-j} \sum_{i=1}^{m} \left(b_i^{(j)}\right)^2 + \sum_{i=1}^{m} \left(a_i^{(\ell)}\right)^2 = \sum_{i=1}^{m} s_i^2. \qquad (33)$$

This property can be employed to adapt the two optimization criteria (given above for the critically sampled wavelet transform) to the stationary wavelet case, by using weighing.

Given a set of parameter values $\{\phi_1, \ldots, \phi_{n-1}\}$ and a prototype signal s, it is straightforward to calculate the wavelet decomposition of s and the values of the criteria V_1 and V_4 of Theorem 3. A local search technique can then be employed to perform numerical optimization. However, caution must be exercised due to the existence of local optima; see also [13]. For instance, the local optima we encountered in an example with two free parameters are illustrated in Figures 4-5.

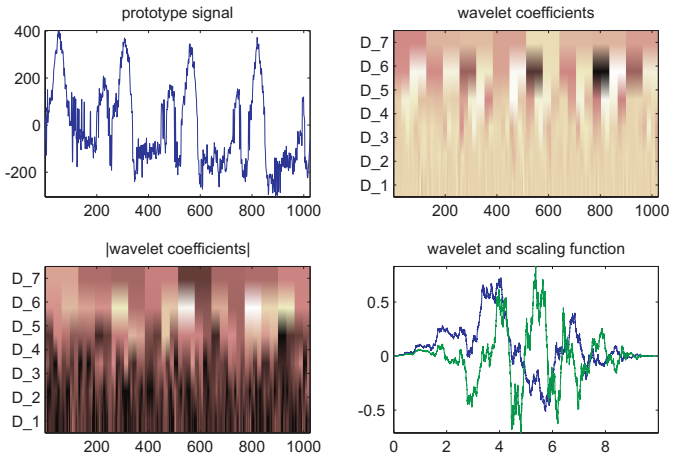


Fig. 6: Top-left: Intracardiac signal (1 kHz). Top-right: Wavelet decomposition of the intracardiac signal with an L^4-optimized wavelet with 5 parameters, vanishing moments up to order 1 and using the critically sampled wavelet transform. In the top row the coarsest detail coefficients are displayed. The color intensity of the blocks corresponds to the coefficient values. The lower rows show finer detail coefficients. Bottom-left: absolute values of the wavelet coefficients. Bottom-right: wavelet and scaling function of the optimized wavelet.

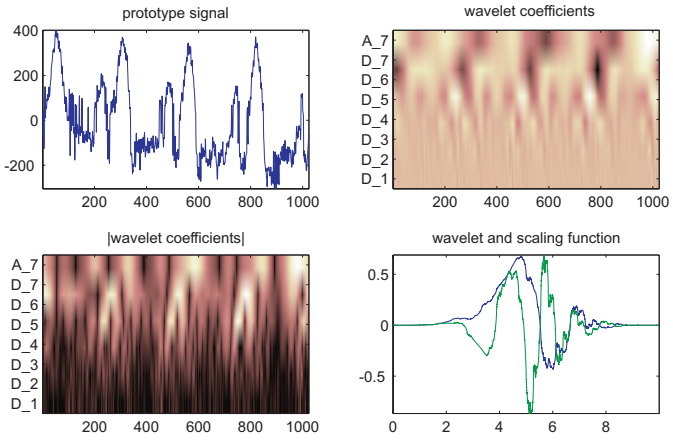


Fig. 7: L^4-optimization with 5 parameters using a stationary wavelet transform and vanishing moments up to order 1.

VII. Examples

As a first example to show which wavelets are obtained by the proposed constructions, we consider a signal from the Physionet "Intracardiac Atrial Fibrillation" database [21] (see Figure 6). A wavelet was designed to give a sparse representation for this prototype data in the wavelet domain. From Figures 6-7 we infer that for the critically sampled wavelet undesired effects may occur. Using the stationary wavelet transform certain effects can be avoided such as the effect that introducing an effective time delay with the filter is beneficial due to the dyadic sampling. In this case vanishing moments up to order 1 were enforced. The low-pass filter coefficients obtained with dyadic optimization are:

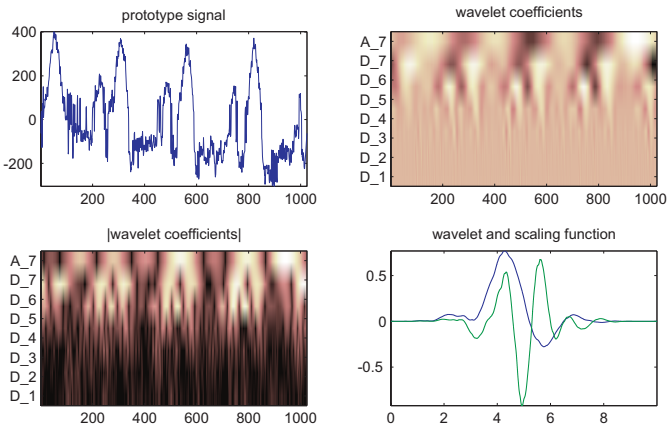


Fig. 8: L^1-optimization with 5 parameters using a stationary wavelet transform and vanishing moments up to order 1.

$$c_0 = 0.2406, c_1 = -0.1609, c_2 = -0.0412, c_3 = 0.4592,$$
$$c_4 = 0.5756, c_5 = 0.3915, c_6 = 0.0219,$$
$$c_7 = -0.4387, c_8 = -0.0898, c_9 = 0.1343.$$

The low-pass filter coefficients obtained with stationary wavelet optimization are:

$$c_0 = 0.0478, c_1 = 0.0333, c_2 = 0.0988, c_3 = 0.2233,$$
$$c_4 = 0.5122, c_5 = 0.7414, c_6 = 0.0816,$$
$$c_7 = -0.3388, c_8 = -0.0333, c_9 = 0.0478.$$

which exhibits apparent structure in the outer coefficients.

In Figure 8, L^1 optimization is used for the wavelet design. The obtained low-pass filter coefficients are:

$$c_0 = 0.0681, c_1 = 0.0352, c_2 = -0.0287, c_3 = 0.3564,$$
$$c_4 = 0.7583, c_5 = 0.4909, c_6 = -0.0734,$$
$$c_7 = -0.2086, c_8 = -0.0172, c_9 = 0.0332.$$

In this example the L^1-minimization approach produces a notably smoother wavelet than the L^4-maximization approach. It aligns well with the morphology of the peaks in the signal.

A second example which concerns an acceleration signal is shown in Figure 9. With 3 parameters and 2 enforced vanishing moments (up to order 1) that gives one degree of freedom. The optimal wavelet strongly resembles the Symlet 3 / Daubechies 3 wavelet. The low-pass filter coefficients we found are:

$$c_0 = 0.3805, c_1 = 0.8240, c_2 = 0.3881,$$
$$c_3 = -0.1453, c_4 = -0.0615, c_5 = 0.0284,$$

which gives an L^2-norm difference of only 0.0920 with the actual Symlet 3 low-pass filter coefficients.

VIII. Conclusions

In this paper we have discussed a data driven approach to design an orthogonal wavelet for an application at hand. We have presented a parameterization of lossless FIR polyphase

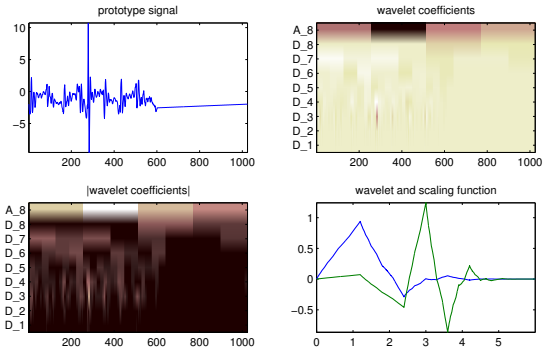


Fig. 9: L^1 wavelet optimization for an acceleration signal, with 1 free parameter (vanishing moments up to order 1).

filters, which matches the framework for orthogonal wavelets with compact support satisfying the wavelet admissibility condition. We then have given the expressions for vanishing moments up to an arbitrary order p: first in terms of the polyphase filter $\tilde{H}(z)$, then in terms of the specific parameterization with parameters $\phi_1, \ldots, \phi_{n-1}$ and for some cases for a parameterization $q_1, \ldots, q_{n-1}$. Manifolds for these orthogonal wavelets with vanishing moments have been shown. Vanishing moments are useful properties of wavelets, because they give smoothness and make that polynomials up to a certain degree end up in the approximation signal. For a small number of vanishing moments, these conditions allows us to perform constrained optimization. For wavelet design we have argued that a sparsity criterion is appropriate, especially if the goal of wavelet processing is either compression or detection. In our set-up, such a sparsity criterion takes the form of an L^1-minimization or an L^4-maximization problem. These criteria can be used both with the critically sampled wavelet transform, or in an adapted form, with the stationary wavelet transform. We have demonstrated the approach with a number of examples. The stationary wavelet transform proves useful in avoiding certain undesired effects.

A few open questions still remain. First, it is known that wavelet functions can only be computed for a given set of filter coefficients when the joint spectral radius of two particular matrices is less than 1, see [22]. This puts an additional constraint on the parameter domain, if one wants to adhere to a wavelet interpretation of the filter bank signal processing approach. It still needs to be worked out how it affects the parameter domain. Second, it is observed that several well-known wavelets tend to come out (in good approximation) for some of the examples we have investigated. This concerns the Daubechies wavelets and Symlets of certain orders, but also Coiflets have been known to turn up; see, e.g., the first local optimum in Figure 5. At present it is unclear what causes this, and whether there might be a more general principle behind this. Matters such as the principle that vanishing moments of order p suppress polynomials up to order p, and that vanishing moments ensure that effective differentiation occurs [1] are expected to be involved.

REFERENCES

[1] S. Mallat, *A Wavelet Tour of Signal Processing: The Sparse Way*. Academic Press, 2008.

[2] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Wellesley-Cambridge Press, 1996.

[3] S. Saxena, V. Kumar, and S. Hamde, "QRS detection using new wavelets," *Journal of Medical Engineering and Technology*, vol. 26, no. 1, pp. 7–15, January-February 2002.

[4] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Comm. on Pure and Appl. Math.*, vol. 41, no. 7, pp. 909–996, 1988.

[5] P. Ravier and O. Buttelli, "Robust detection of QRS complex using Klauder wavelets," in *EUSIPCO 2004*, 2004.

[6] J. Chapa and R. Rao, "Algorithms for designing wavelets to match a specified signal," *IEEE Trans. Signal Processing*, vol. 48, no. 12, pp. 3395–3406, December 2000.

[7] A. Gupta, S. D. Joshi, and S. Prasad, "A new method of estimating wavelets with desired features from a given signal," *Signal Processing*, vol. 85, no. 1, pp. 147–161, January 2005.

[8] D. Alpay, L. Baratchart, and A. Gombani, "On the differential structure of matrix-valued rational inner functions," *Operator Theory: Advances and Applications*, vol. 73, pp. 30–66, 1994.

[9] B. Hanzon, M. Olivi, and R. Peeters, "Balanced realizations of discrete-time stable all-pass systems and the tangential schur algorithm," *Linear Algebra and its Applications*, vol. 418, pp. 793–820, 2006.

[10] R. Peeters, M. Olivi, and B. Hanzon, "Parametrization of matrix-valued lossless functions based on boundary interpolation," in *Proceedings of the 19th MTNS, Budapest, Hungary*, 2010, pp. 563–570.

[11] Y. Potapov, "The multiplicative structure of j-contractive matrix functions," *Amer. Math. Soc. Transl. Ser.*, vol. 2, no. 15, pp. 131–244, 2006.

[12] P. Vaidyanathan and Z. Doğanata, "The role of lossless systems in modern digital signal processing: analog tutorial," *IEEE Trans. Educ.*, vol. 32, no. 3, pp. 181–197, August 1989.

[13] J. Karel, "A wavelet approach to cardiac signal processing for low-power hardware applications," Ph.D. dissertation, Maastricht University, December 2009, ISBN: 978-90-5278-887-6. [Online]. Available: http://dissertaties.ub.unimaas.nl/detail.aspx?id=20802

[14] S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, no. 7, pp. 674–693, July 1989.

[15] J. Karel, R. Peeters, R. Westra, K. Moermans, S. Haddad, and W. Serdijn, "Optimal discrete wavelet design for cardiac signal processing," in *Proc. 27th Ann. Int. Conf. of the IEEE Eng. in Medicine and Biology Society (EMBC 2005), Shanghai, China, September 1-4*, 2005.

[16] D. L. Donoho, "For most underdetermined systems of linear equations, the minimal ℓ_1-norm near-solution approximates the sparsest near-solution." *Comm. Pure Appl. Math.*, vol. 59, no. 6, pp. 797–829, June 2006.

[17] D. L. Donoho and M. Elad, "Optimally sparse representation from overcomplete dictionaries via ℓ_1 norm minimization," *Proc. Natl. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2002, March 2003.

[18] E. J. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted l1 minimization," *J. Fourier Anal. Appl.*, vol. 14, pp. 877–905, 2007.

[19] E. J. Candès and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 21–30, March 2008.

[20] G. P. Nason and B. W. Silverman, "The stationary wavelet transform and some statistical applications," in *Lecture Notes in Statistics: Wavelets and Statistics*, ser. Lecture Notes in Statistics, A. Antoniadis and G. Oppenheim, Eds., vol. 103. Springer-Verlag, 1995, pp. 281–299.

[21] A. Goldberger, L. Amaral, L. Glass, J. Hausdorff, P. Ivanov, R. Mark, J. Mietus, G. Moody, C.-K. Peng, and H. Stanley, "Physiobank, physiotoolkit, and physionet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, no. 23, pp. e215–e220, 2000.

[22] I. Daubechies and J. C. Lagarias, "Two-scale difference equations II. local regularity, infinite products of matrices and fractals," *SIAM J. Math. Anal.*, vol. 23, no. 4, pp. 1031–1079, 1992.