Fast algorithm for directional time-scale analysis using wavelets

Rob A. Zuidwijk[a] and Paul M. de Zeeuw[a]

[a]Center for Mathematics & Computer Science (CWI)

P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

ABSTRACT

Fast algorithms performing time-scale analysis of multivariate functions are discussed. The algorithms employ univariate wavelets and involve a directional parameter, namely the angle of rotation. Both the rotation steps and the wavelet analysis/synthesis steps in the algorithms require a number of computations proportional to the number of data involved.

The rotation and wavelet techniques are used for the segregation of wanted and unwanted components in a seismic signal. As an illustration, the rotation and wavelet methods are applied to a synthetic shot record.

Keywords: (bi-)orthogonal wavelets, fast wavelet transform, interpolation, rotation, seismic data processing

1. INTRODUCTION

The topic of this paper is the wavelet analysis of functions in several variables using univariate wavelets. In particular, we study fast algorithms to perform such analyses. Observe that the well-known pyramid algorithm for multivariate functions[12] , which uses tensor products of univariate wavelets, is such a method. An important aspect of the methods in this paper is the introduction of a directional parameter, namely the angle of rotation. Besides a combination of the usual pyramid scheme and rotation, we also study a method based on the wavelet X-ray transform.

The *wavelet X-ray transform*[10,13,21] given by

$$P_\psi f(\theta, x, b, a) = \int_{\mathbb{R}} f(x + t\theta) \frac{1}{\sqrt{a}} \overline{\psi\left(\frac{t-b}{a}\right)}\, dt$$

computes a one-dimensional wavelet transform of the restriction of $f \in L^2(\mathbb{R}^n)$ to the line $\{x + t\theta \mid t \in \mathbb{R}\}$, where $\theta \in \mathbb{R}^n$ is a unit vector and $x \in \theta^\perp = \{y \in \mathbb{R}^n \mid y \perp \theta\}$. The *translation, dilation* parameter pair (b, a) is taken from the open upper half plane $\{(x, y) \in \mathbb{R}^2 \mid y > 0\}$, just as for the usual 1-D wavelet transform. The wavelet ψ satisfies additional conditions so that the function f can be reconstructed from its wavelet X-ray coefficients $P_\psi f(\theta, x, b, a)$.

Properties of the wavelet X-ray transform were discussed in[10,13] . The transform was discretized by means of Fourier methods in[18] and used to detect linear events in SAR images there. The transform has received further attention in[21,22] , where an alternative discretization was proposed. A fast implementation of this discretization simply comes down to the computation of the 1-D fast wavelet transform along parallel gridlines in a rotated grid. Recall that the 1-D fast wavelet transform computes wavelet coefficients of a function starting with function values given at equidistant points. In this paper, we shall not recite the 1-D fast wavelet transform and refer to[4,11,12] for details. We mention only the immediate fact that computing fast wavelet transforms along parallel gridlines requires a number of computations proportional to the number of gridpoints.

In practice, a function f is usually given on the standard Euclidean grid by means of values at the grid points. These values can be interpreted as point evaluations or as certain averages. The latter interpretation fits in the context of biorthogonal Riesz systems. Some remarks on biorthogonal Riesz systems are given in Section 2.

The application of the fast wavelet X-ray transform or a rotated version of the 2-D wavelet transform requires values at the grid points of a rotated (and possibly dilated) version of the Euclidean grid. As a consequence, the following question comes up: Given values of f at the standard Euclidean grid, what are the *best* values of f at the grid points of the rotated and dilated grid? Clearly, this is an interpolation issue, and it will be dealt with in this paper. Optimal values of f -in least squares sense- at the rotated and dilated grid are described in Sections 3 and

Other author information: R.A.Zuidwijk@cwi.nl, Paul.de.Zeeuw@cwi.nl, dbs.cwi.nl/cwwwi/owa/cwwwi.print_projects?ID=64

4. Moreover, in Section 4, it will be shown that the number of computations required for the rotation procedure is proportional to the number of data.

Finally, we consider an application from the field of seismic data processing. For obvious reasons geophysicist are after geometric information on the stratification of impedance beneath the earth's surface. Acquisition of data is done by exciting a signal at the surface. Waves are reflected at interfaces (i.e. where the impedances changes rapidly). Seismometers at the surface record the groundmotion as the reflecting waves arrive. As the deflections are measured in time and for an array of seismometers, we can represent the recorded signals as a two-dimensional gridfunction (i.e. a rectangular uniform grid where each gridpoint has been assigned a real number). The information of interest is then constituted by hyperbolic shaped events. Unfortunately, these events are overshadowed by groundroll and surface waves directly stemming from the excitative source. Indeed, these waves often have dominant amplitude and completely blur the picture. It is a major challenge to filter out the surface waves from such given seismic data. We propose a synthetic and strongly simplified testproblem that might serve as a benchmark. We sketch a numerical algorithm for the solution of this testproblem in Section 5.

2. BIORTHOGONAL RIESZ SYSTEMS

2.1. Biorthogonality and Hilbert space geometry

We introduce Riesz systems in Hilbert space and focus on geometric issues related to biorthogonal pairs of Riesz systems. For general reading on Riesz systems, we refer to[17] , although many wavelet textbooks such as[1,4,5,9] contain material on the subject relevant to wavelets as well. Recall that a *Riesz system* in a Hilbert space H is a sequence of vectors $(x_k)_{k=1}^{\infty}$ with two positive constants $0 < A \leq B$ such that

$$A \sum_{k=1}^{N} |a_k|^2 \leq \left\| \sum_{k=1}^{N} a_k x_k \right\|_H^2 \leq B \sum_{k=1}^{N} |a_k|^2$$

for all finite sequences $a_1, \ldots, a_N$. If the constants A, B are chosen optimally for the Riesz system under consideration, then A, B are called the *Riesz bounds* of the Riesz system. Let $(x_k)_{k=1}^{\infty}$ be a Riesz system in the Hilbert space H with closed linear span $V = \mathrm{span}(x_k)_{k=1}^{\infty}$. We first remark that there exists a *unique* sequence of vectors $(\widetilde{x}_k)_{k=1}^{\infty}$ in V such that the *biorthogonality condition*

$$\langle x_k, \widetilde{x}_l \rangle_H = \delta_{k,l} \tag{1}$$

is satisfied. It turns out that $(\widetilde{x}_k)_{k=1}^{\infty}$ is a Riesz system and we get $\mathrm{span}(\widetilde{x}_k)_{k=1}^{\infty} = V$. Finally, if the Riesz bounds of $(x_k)_{k=1}^{\infty}$ are given by A, B, then the Riesz bounds of $(\widetilde{x}_k)_{k=1}^{\infty}$ are given by $\widetilde{A} = B^{-1}, \widetilde{B} = A^{-1}$. A Riesz system $(\widetilde{x}_k)_{k=1}^{\infty}$ which satisfies (1) will be called a *dual Riesz system* with respect to $(x_k)_{k=1}^{\infty}$. Observe that a Riesz system is its own dual if and only if it is orthonormal. This particular situation corresponds to the case when $A = B = 1$. If we allow the sequence $(\widetilde{x}_k)_{k=1}^{\infty}$ to span a subspace $\widetilde{V} \subseteq H$ which is different from the subspace V, then the biorthogonality condition (1) does not imply that $(\widetilde{x}_k)_{k=1}^{\infty}$ is a Riesz system. However, we can state the following results[23] .

THEOREM 2.1. *Let $V, \widetilde{V} \subseteq H$ be subspaces in Hilbert space, and assume that V contains a Riesz basis $(x_k)_{k=1}^{\infty}$ with Riesz bounds A, B. Then $\widetilde{V}$ contains a Riesz basis $(\widetilde{x}_k)_{k=1}^{\infty}$, with Riesz bounds $\widetilde{A}, \widetilde{B}$, biorthogonal to $(x_k)_{k=1}^{\infty}$ if and only if $\widetilde{V} \oplus V^{\perp} = H$, i.e., if and only if there exists a bounded projection P onto $\widetilde{V}$ along $V^{\perp}$. In that case, the Riesz basis $(\widetilde{x}_k)_{k=1}^{\infty}$ in $\widetilde{V}$ is uniquely determined and the Riesz bounds of the systems are subject to*

$$\alpha^2 \leq \widetilde{A}B \leq \beta^2, \quad \alpha^2 \leq A\widetilde{B} \leq \beta^2,$$

where

$$\alpha = \inf_{0 \neq x \in \widetilde{V}} \frac{\|Px\|}{\|x\|} = \inf_{0 \neq y \in V} \frac{\|P^* y\|}{\|y\|}, \quad \beta = \sup_{0 \neq x \in \widetilde{V}} \frac{\|Px\|}{\|x\|} = \sup_{0 \neq y \in V} \frac{\|P^* y\|}{\|y\|}.$$

.

Observe that in the case when $V = \widetilde{V}$, the theorem reproduces the fact that $\widetilde{A} = B^{-1}$ and $\widetilde{B} = A^{-1}$. The bounds in the theorem are not sharp but can be attained in specific cases, as can be seen even in finite dimensions[23] . In the papers[14–16] , the constant β as in Theorem 2.1 arises as the reciprocal cosine of the angle between V and $\widetilde{V}$ and is used there to obtain an important instance of the following result. This instance is discussed in the next subsection. Observe that the case when $\beta = 1$ corresponds to the case when $V = \widetilde{V}$.

THEOREM 2.2. *Let $V, \widetilde{V} \subseteq H$ be subspaces in Hilbert space such that $V \oplus \widetilde{V}^{\perp} = H$. Let P be the projection onto V along $\widetilde{V}^{\perp}$ and let Π be the orthoprojector onto V. For all $g \in H$, we get*

$$\|g - \Pi g\|_H \leq \|g - Pg\|_H \leq \beta \|g - \Pi g\|_H,$$

where

$$\beta = \sup_{0 \neq k \in V} \frac{\|P^* k\|_H}{\|k\|_H}.$$

2.2. Multiresolution analysis

Riesz systems consisting of integer translates of a single function $\varphi \in L^2(\mathbb{R})$ are of particular interest here. More explicitly, we will assume that $\varphi \in L^2(\mathbb{R})$ is a function such that the sequence $(\varphi(\cdot - k))_{k \in \mathbb{Z}}$ is a Riesz system in $L^2(\mathbb{R})$. If V is the closed linear span of this Riesz system, then there exists a unique Riesz system in V biorthogonal to $(\varphi(\cdot - k))_{k \in \mathbb{Z}}$ which is a basis in V. In particular, for the given function φ, one may construct a function $\widetilde{\varphi} \in V$, such that the dual Riesz system of $(\varphi(\cdot - k))_{k \in \mathbb{Z}}$ is given by $(\widetilde{\varphi}(\cdot - k))_{k \in \mathbb{Z}}$. We mention that if φ has compact support and is bounded, then $\widetilde{\varphi}$ has exponential decay at infinity always and in some cases compact support[4] .

We shall make further assumptions on the functions φ and $\widetilde{\varphi}$. Indeed, we will assume that these functions induce multiresolution analyses. Multiresolution analysis forms one of the key issues in wavelet theory; see.[4,5,12] A *multiresolution analysis* (MRA) in $L^2(\mathbb{R})$ is a sequence of subspaces $(V_j)_{j \in \mathbb{Z}}$ in $L^2(\mathbb{R})$ with the following properties: For each $j \in \mathbb{Z}$, we have

(i) $V_j \subset V_{j+1}$,

(ii) $f \in V_j \;\Leftrightarrow\; f(2\cdot) \in V_{j+1}$,

and

(iii) $\bigcap_{j \in \mathbb{Z}} V_j = (0)$,

(iv) $\bigcup_{j \in \mathbb{Z}} V_j \subseteq L^2(\mathbb{R})$ dense,

(v) there exists $\varphi \in V_0$ such that $(\varphi(\cdot - k))_{k \in \mathbb{Z}}$ is a Riesz basis in V_0.

A function $\varphi \in L^2(\mathbb{R})$ which gives rise to a multiresolution analysis as above will be called a *scaling function*. Examples of such function are spline functions together with various dual functions[1,3,4] .

We remark that a biorthogonal pair of scaling functions leads to the definition of a biorthogonal pair of wavelets and a (fast) wavelet transform. We refer to[1,4,11] .

3. LEAST SQUARES METHOD AND DUAL METHOD

Throughout this section, we will assume that φ induces a multiresolution analysis as described in the previous section, i.e., φ is a scaling function. We will also fix a dual scaling function $\widetilde{\varphi} \in L^2(\mathbb{R})$.

Given an orthonormal basis $\theta_1, \ldots, \theta_n$ in $\mathbb{R}^n$, we consider the (rotated and dilated) grid

$$\mathbb{G}_{\theta,d} = \{\sum_{r=1}^{n} p_r d_r \theta_r \mid (p_1, \ldots, p_n) \in \mathbb{Z}^n\},$$

where $d_1, \ldots, d_n$ are positive real numbers. The standard basis in $\mathbb{R}^n$ will be denoted by $e_1, \ldots, e_n$ and the standard grid by $\mathbb{G}_e$ accordingly. Observe that $\mathbb{G}_{\theta,d}$ is the image of $\mathbb{G}_e$ under permutation, rotation and dilation. In fact, $\mathbb{G}_{\theta,d} = DRQ\mathbb{G}_e$, where Q is an $n \times n$ permutation matrix, R is an $n \times n$ rotation matrix and D is an $n \times n$ diagonal matrix with diagonal $d = (d_1, \ldots, d_n)^T$.

In order to deal with functions on $\mathbb{R}^n$ for $n \geq 2$, we shall construct multivariate functions by means of products of univariate ones. Indeed, given an orthonormal basis $\theta_1, \ldots, \theta_n$ in $\mathbb{R}^n$, we shall write

$$\Phi_{\theta,d}(y) = \prod_{r=1}^{n} d_r^{-1/2} \varphi(d_r^{-1} \langle y, \theta_r \rangle), \quad \text{almost all } y \in \mathbb{R}^n,$$

where φ is a scaling function. In the same fashion, one defines $\widetilde{\Phi}_{\theta,d}$ using the dual scaling function $\widetilde{\varphi}$. The next two lemmas justify that $\Phi_{\theta,d}$ will be called a *multivariate scaling function*.

LEMMA 3.1. *The system $(\Phi_{\theta,d}(\cdot - p))_{p \in \mathbb{G}_{\theta,d}}$ is a Riesz system in $L^2(\mathbb{R}^n)$. The dual Riesz system is given by $(\widetilde{\Phi}_{\theta,d}(\cdot - p))_{p \in \mathbb{G}_{\theta,d}}$. In particular,*

$$\langle \Phi_{\theta,d}(\cdot - p), \widetilde{\Phi}_{\theta,d}(\cdot - q) \rangle_{L^2(\mathbb{R}^n)} = \delta_{p,q}.$$

The Riesz bounds of the system are given by A^n, B^n.

A *multiresolution analysis in $L^2(\mathbb{R}^n)$ associated with a grid* $\mathbb{G}_{\theta,d}$ is a sequence of subspaces $(V_{\theta,d,j})_{j \in \mathbb{Z}}$ in $L^2(\mathbb{R}^n)$ with the properties: For each $j \in \mathbb{Z}$, we have

(i) $V_{\theta,d,j} \subset V_{\theta,d,j+1}$,

(ii) $f \in V_{\theta,d,j} \;\Leftrightarrow\; f(2\cdot) \in V_{\theta,d,j+1}$,

and

(iii) $\bigcap_{j \in \mathbb{Z}} V_{\theta,d,j} = (0)$,

(iv) $\bigcup_{j \in \mathbb{Z}} V_{\theta,d,j} \subseteq L^2(\mathbb{R}^n)$ dense,

(v) there exists $\Phi \in V_{\theta,d,0}$ such that $(\Phi(\cdot - p))_{p \in \mathbb{G}_{\theta,d}}$ is a Riesz basis in $V_{\theta,d,0}$.

LEMMA 3.2. *If φ induces a multiresolution analysis in $L^2(\mathbb{R})$, then $\Phi_{\theta,d}$ induces a multiresolution analysis in $L^2(\mathbb{R}^n)$ associated with the grid $\mathbb{G}_{\theta,d}$.*

We shall introduce a multivariate interpolation operator which incorporates the Riesz system induced by the multivariate scaling function. Define $L_{\theta,d} : \ell^2(\mathbb{G}_{\theta,d}) \to L^2(\mathbb{R}^n)$ by

$$L_{\theta,d}(a_p)_{p \in \mathbb{G}_{\theta,d}}(y) = \sum_{p \in \mathbb{G}_{\theta,d}} a_p \Phi_{\theta,d}(y - p)$$

for almost all $y \in \mathbb{R}^n$. The operator $\widetilde{L}_{\theta,d}$ is defined in the same way using the dual scaling function. Observe that by Lemma 3.1, we get

$$A^n \|(a_p)_{p \in \mathbb{G}_{\theta,d}}\|^2_{\ell^2(\mathbb{G}_{\theta,d})} \leq \|L_{\theta,d}(a_p)_{p \in \mathbb{G}_{\theta,d}}\|^2_{L^2(\mathbb{R}^n)} \leq B^n \|(a_p)_{p \in \mathbb{G}_{\theta,d}}\|^2_{\ell^2(\mathbb{G}_{\theta,d})}.$$

This implies that $L_{\theta,d} : \ell^2(\mathbb{G}_{\theta,d}) \to L^2(\mathbb{R}^n)$ is a bounded injective operator with closed range. It follows that $L_{\theta,d}$ has a bounded left-inverse. Indeed, the adjoint operator $L^*_{\theta,d} : L^2(\mathbb{R}^n) \to \ell^2(\mathbb{G}_{\theta,d})$ of $L_{\theta,d}$ is given by

$$(L^*_{\theta,d} g) = \left(\langle g, \Phi_{\theta,d}(\cdot - p) \rangle_{L^2(\mathbb{R}^n)} \right)_{p \in \mathbb{G}_{\theta,d}}, \quad g \in L^2(\mathbb{R}^n),$$

and now we can identify a left inverse of $L_{\theta,d}$.

THEOREM 3.3. *The operator $L_{\theta,d} : \ell^2(\mathbb{G}_{\theta,d}) \to L^2(\mathbb{R}^n)$ has $\widetilde{L}^*_{\theta,d} : L^2(\mathbb{R}^n) \to \ell^2(\mathbb{G}_{\theta,d})$ as a left inverse.*

Proof. Note that Lemma 3.1 and the continuity of the inner product implies

$$\langle \widetilde{L}^*_{\theta,d} L_{\theta,d}(a_p)_{p \in \mathbb{G}_{\theta,d}}, (b_q)_{q \in \mathbb{G}_{\theta,d}} \rangle_{\ell^2(\mathbb{G}_{\theta,d})} = \langle L_{\theta,d}(a_p)_{p \in \mathbb{G}_{\theta,d}}, \widetilde{L}_{\theta,d}(b_q)_{q \in \mathbb{G}_{\theta,d}} \rangle_{L^2(\mathbb{R}^n)} =$$

$$\langle \sum_{p \in \mathbb{G}_{\theta,d}} a_p \Phi_{\theta,d}(\cdot - p), \sum_{q \in \mathbb{G}_{\theta,d}} b_q \widetilde{\Phi}_{\theta,d}(\cdot - q) \rangle_{L^2(\mathbb{R}^n)} = \sum_{p \in \mathbb{G}_{\theta,d}} a_p \overline{b_p}.$$

□

In the particular case when $\Phi_{\theta,d} = \widetilde{\Phi}_{\theta,d}$, i.e., in the orthonormal case, we obviously get $L_{\theta,d} = \widetilde{L}_{\theta,d}$ and by Theorem 3.3, we see that in this case $L^*_{\theta,d}L_{\theta,d} = I_{\ell^2(\mathbb{G}_{\theta,d})}$. In the general situation, the self-adjoint operator $L^*_{\theta,d}L_{\theta,d}$ will not be the identity, although the following holds true.

LEMMA 3.4. *The operator $L^*_{\theta,d}L_{\theta,d} : \ell^2(\mathbb{G}_{\theta,d}) \to \ell^2(\, GG_{\theta,d})$ is a strictly positive, hence boundedly invertible, operator which satisfies the estimates (in the sense of self-adjoint operators)*

$$A^n\, I \leq L^*_{\theta,d}L_{\theta,d} \leq B^n\, I.$$

Observe that the expression

$$(f_p)_{p\in\mathbb{G}_{\theta,d}} = (L^*_{\theta,d}L_{\theta,d})^{-1}L^*_{\theta,d}g \tag{2}$$

makes sense and provides the least-squares solution to the equation

$$L_{\theta,d}(f_p)_{p\in\mathbb{G}_{\theta,d}} = g \tag{3}$$

where $g \in L^2(\mathbb{R}^n)$ is a given function. Another approximate solution to (3) is given by

$$(f_p)_{p\in\mathbb{G}_{\theta,d}} = \widetilde{L}^*_{\theta,d}g. \tag{4}$$

We will now look at equation (3) and its approximate solutions (2) and (4) more closely. If we apply the operator $L_{\theta,d}$ to the right hand side of (2), we get $L_{\theta,d}(L^*_{\theta,d}L_{\theta,d})^{-1}L^*_{\theta,d}g$. The operator $\Pi_{\theta,d} = L_{\theta,d}(L^*_{\theta,d}L_{\theta,d})^{-1}L^*_{\theta,d}$ is the orthogonal projection onto ran $L_{\theta,d}$. The *least squares solution method* (2) produces an error $\|g-\Pi_{\theta,d}g\|_{L^2(\mathbb{R}^n)}$ which equals zero whenever $g \in$ ran $L_{\theta,d}$. In general, this error is minimal among all attainable errors, since $\|g-\Pi_{\theta,d}g\|_{L^2(\mathbb{R}^n)}$ equals the distance between g and ran $L_{\theta,d}$.

On the other hand, if we apply $L_{\theta,d}$ to the right hand side of (4), we get $L_{\theta,d}\widetilde{L}^*_{\theta,d}g$. The operator $P_{\theta,d} = L_{\theta,d}\widetilde{L}^*_{\theta,d}$ is a not necessarily orthogonal projection onto ran $L_{\theta,d}$. This *dual solution method* produces an error $\|g-P_{\theta,d}g\|_{L^2(\mathbb{R}^n)}$. Since $g - \Pi_{\theta,d}g \perp \Pi_{\theta,d}g - P_{\theta,d}g$, we get

$$\|g - P_{\theta,d}g\|^2_{L^2(\mathbb{R}^n)} = \|P_{\theta,d}g - \Pi_{\theta,d}g\|^2_{L^2(\mathbb{R}^n)} + \|g - \Pi_{\theta,d}g\|^2_{L^2(\mathbb{R}^n)}.$$

It is immediate that the error caused by the least squares solution method is majorized by the error caused by the dual method. Moreover, the difference between the two errors can be measured by $\|P_{\theta,d}g - \Pi_{\theta,d}g\|_{L^2(\mathbb{R}^n)}$.

THEOREM 3.5. *Let $\Phi_{\theta,d}$ and $\widetilde{\Phi}_{\theta,d}$ be multivariate scaling functions as defined before which satisfy the biorthogonality condition*

$$\langle \Phi_{\theta,d}(\cdot - p), \widetilde{\Phi}_{\theta,d}(\cdot - q) = \delta_{p,q}, \quad p, q \in \mathbb{G}_{\theta,d},$$

and let β be defined as in Theorem 2.2. The closed linear span of $(\Phi_{\theta,d}(\cdot - p))_{p\in\mathbb{G}_{\theta,d}}$ is denoted by $V_{\theta,d}$ and the closed linear span of $(\widetilde{\Phi}_{\theta,d}(\cdot - p))_{p\in\mathbb{G}_{\theta,d}}$ is denoted by $\widetilde{V}_{\theta,d}$. The orthoprojector onto $V_{\theta,d}$ is given by $\Pi_{\theta,d}$, and the projection onto $V_{\theta,d}$ along $\widetilde{V}^{\perp}_{\theta,d}$ reads $P_{\theta,d}$. We get for $g \in L^2(\mathbb{R}^n)$,

$$\|g - \Pi_{\theta,d}g\|_{L^2(\mathbb{R}^n)} \leq \|g - P_{\theta,d}g\|_{L^2(\mathbb{R}^n)} \leq \beta^n \|g - \Pi_{\theta,d}g\|_{L^2(\mathbb{R}^n)}.$$

4. INTERPOLATION BETWEEN ROTATED GRIDS

We shall now apply the considerations in the preceding section to interpolation between dilated and rotated Cartesian grids. The situation is as follows. A collection of data (complex numbers) $(a_q)_{q\in\mathbb{G}_e}$ is given and a function $g \in L^2(\mathbb{R}^n)$ is attached to this collection by means of interpolation. Indeed, we assume that

$$g(y) = \sum_{q\in\mathbb{G}_e} a_q \Phi_e(y - q), \quad \text{almost all } y \in \mathbb{R}^n.$$

Note that we have constructed a function g such that

$$a_q = \langle g, \widetilde{\Phi}_e(\cdot - q)\rangle, \quad q \in \mathbb{G}_e.$$

In words, the function g reproduces the original data when inner products are taken with translates along the standard grid $\mathbb{G}_e$ of the dual multivariate function $\widetilde{\Phi}_e$. We now search for a collection of data $(f_p)_{p\in\mathbb{G}_{\theta,d}}$ associated with the dilated and rotated grid $\mathbb{G}_{\theta,d}$. This data should reproduce the function g as an interpolant of $(f_p)_{p\in\mathbb{G}_{\theta,d}}$. In general, this is not possible. Therefore, we shall write

$$f(y) = \sum_{p\in\mathbb{G}_{\theta,d}} f_p \Phi_{\theta,d}(y-p), \quad \text{almost all } y \in \mathbb{R}^n,$$

and we shall try to minimize $\|f-g\|_{L^2(\mathbb{R}^n)}$.

In terms of interpolation operators, we have the following situation. Given the data $(a_q)_{q\in\mathbb{G}_e} \in \ell^2(\mathbb{G}_e)$, we construct $g = L_e(a_q)_{q\in\mathbb{G}_e}$. The least squares solution to

$$L_{\theta,d}(f_p)_{p\in\mathbb{G}_{\theta,d}} = g = L_e(a_q)_{q\in\mathbb{G}_e}$$

provides us with data $(f_p)_{p\in\mathbb{G}_{\theta,d}}$ on the rotated and dilated grid, such that $f = L_{\theta,d}(f_p)_{p\in\mathbb{G}_{\theta,d}}$ minimizes $\|f-g\|_{L^2(\mathbb{R}^n)}$ among all possible solutions. The least squares solution can be obtained with the least squares method

$$(f_p)_{p\in\mathbb{G}_{\theta,d}} = (L^*_{\theta,d}L_{\theta,d})^{-1}L^*_{\theta,d}g = (L^*_{\theta,d}L_{\theta,d})^{-1}L^*_{\theta,d}L_e(a_q)_{q\in\mathbb{G}_e}.$$

Here, the least squares method involves the solution of a linear system with a sparse matrix. With complete factorization the number of (floating point) operations required for the solution is at least proportional to the square of the number of data, see e.g.[7] .

In order to reduce the number of computations, the dual method

$$(f_p)_{p\in\mathbb{G}_{\theta,d}} = \widetilde{L}^*_{\theta,d}g = \widetilde{L}^*_{\theta,d}L_e(a_q)_{q\in\mathbb{G}_e}$$

is proposed. Moreover, the solution of the dual method can be taken arbitrarily close or equal to the least squares solution. We describe the dual method. The operator $\widetilde{L}^*_{\theta,d}L_e : \ell^2(\mathbb{G}_e) \to \ell^2(\mathbb{G}_{\theta,d})$ is applied to the data $(a_q)_{q\in\mathbb{G}_e}$ as follows: Observe that for $p \in \mathbb{G}_{\theta,d}$,

$$\left(\widetilde{L}^*_{\theta,d}L_e(a_q)_{q\in\mathbb{G}_e}\right)_p = \int_{\mathbb{R}^n} L_e(a_q)_{q\in\mathbb{G}_e}(y)\overline{\widetilde{\Phi}_{\theta,d}(y-p)}\,dy =$$

$$\int_{\mathbb{R}^n}\sum_{q\in\mathbb{G}_e} a_q\Phi_e(y-q)\overline{\widetilde{\Phi}_{\theta,d}(y-p)}\,dy = \sum_{q\in\mathbb{G}_e} a_q\langle\Phi_e(\cdot-q),\widetilde{\Phi}_{\theta,d}(\cdot-p)\rangle_{L^2(\mathbb{R}^n)}.$$

The infinite matrix $M = (M_{p,q})_{p\in\mathbb{G}_{\theta,d},q\in\mathbb{G}_e}$, with matrix elements given by $M_{p,q} = \langle\Phi_e(\cdot-q),\widetilde{\Phi}_{\theta,d}(\cdot-p)\rangle_{L^2(\mathbb{R}^n)}$ provides the solution to the dual method as follows:

$$(f_p)_{p\in\mathbb{G}_{\theta,d}} = M\,(a_q)_{q\in\mathbb{G}_e}.$$

In order to describe the number of calculations required to perform the dual method, we consider the case when the dataset $(a_q)_{q\in\mathbb{G}_e}$ is finite. In particular, we denote the unit cube in $\mathbb{R}^n$ by $K = [-1,1]^n$ and assume that $a_q = 0$ for $q \notin Q\cdot K$ where Q is a fixed positive integer. Further, we will assume that the underlying scaling functions φ and $\widetilde{\varphi}$ have compact support. Indeed, assume that $\rho,\widetilde{\rho} > 0$ are chosen such that supp $\varphi \subseteq [-\rho,\rho]$ and supp $\widetilde{\varphi} \subseteq [-\widetilde{\rho},\widetilde{\rho}]$. Let $(p_1,\ldots,p_n) \in \mathbb{Z}^n$ and $(q_1,\ldots,q_n) \in \mathbb{Z}^n$, then $q = \sum_{r=1}^n q_r e_r \in \mathbb{G}_e$ and $p = \sum_{r=1}^n p_r\theta_r d_r \in \mathbb{G}_{\theta,d}$. The matrix element

$$M_{p,q} = \int_{\mathbb{R}^n}\prod_{r=1}^n \varphi(\langle y,e_r\rangle - q_r)d_r^{-1/2}\overline{\widetilde{\varphi}(d_r^{-1}\langle y,\theta_r\rangle - p_r)}\,dy$$

is nonzero, only if

$$|\langle y,e_r\rangle - q_r| \le \rho, \quad |d_r^{-1}\langle y,\theta_r\rangle - p_r| \le \widetilde{\rho}, \quad r = 1,\ldots,n.$$

We get

$$\|y-q\|^2 = \sum_{r=1}^{n} |\langle y, e_r\rangle - q_r|^2 \le \rho^2 n,$$

$$\|y-p\|^2 = \sum_{r=1}^{n} |\langle y, \theta_r\rangle - d_r p_r|^2 = \sum_{r=1}^{n} d_r^2 |d_r^{-1}\langle y, \theta_r\rangle - p_r|^2 \le \widetilde{\rho}^2 \|d\|^2,$$

where $\|d\|$ is the Euclidean norm of $d = (d_1, \ldots, d_n)^T$. This provides

$$0 = \|(y-q)+(q-p)+(p-y)\| \ge \|p-q\| - \|y-p\| - \|y-q\| \ge \|p-q\| - (\rho\sqrt{n} + \widetilde{\rho}\|d\|).$$

As a result, we see that

$$\|p-q\| \le (\rho\sqrt{n} + \widetilde{\rho}\|d\|).$$

We may conclude that if $\|p-q\| > (\rho\sqrt{n} + \widetilde{\rho}\|d\|)$, then $M_{p,q} = 0$. This means that the matrix M has a *band-like structure* with *band width* majorized by $(\rho\sqrt{n} + \widetilde{\rho}\|d\|)$. The number of data N, i.e., the number of gridpoints of $\mathbb{G}_{\theta,d}$ in $Q \cdot K$, is proportional to $(2Q)^n \Delta^{-1}$, where $\Delta = d_1 \ldots d_n$. The number of computations required to calculate $(f_p)_{p\in\mathbb{G}_{\theta,d}}$ is majorized by the number $N(\rho\sqrt{n} + \widetilde{\rho}\|d\|)$ which is of the same order as the number of data involved.

5. AN APPLICATION IN SEISMIC DATA PROCESSING

In this section we consider an application of the theory described in the previous section. We propose a testproblem and sketch the outline of a numerical algorithm to solve it.

5.1. An idealized testproblem

We propose a simple synthetic shotrecord that is depicted by the picture on the left of Figure 1. It shows a simple

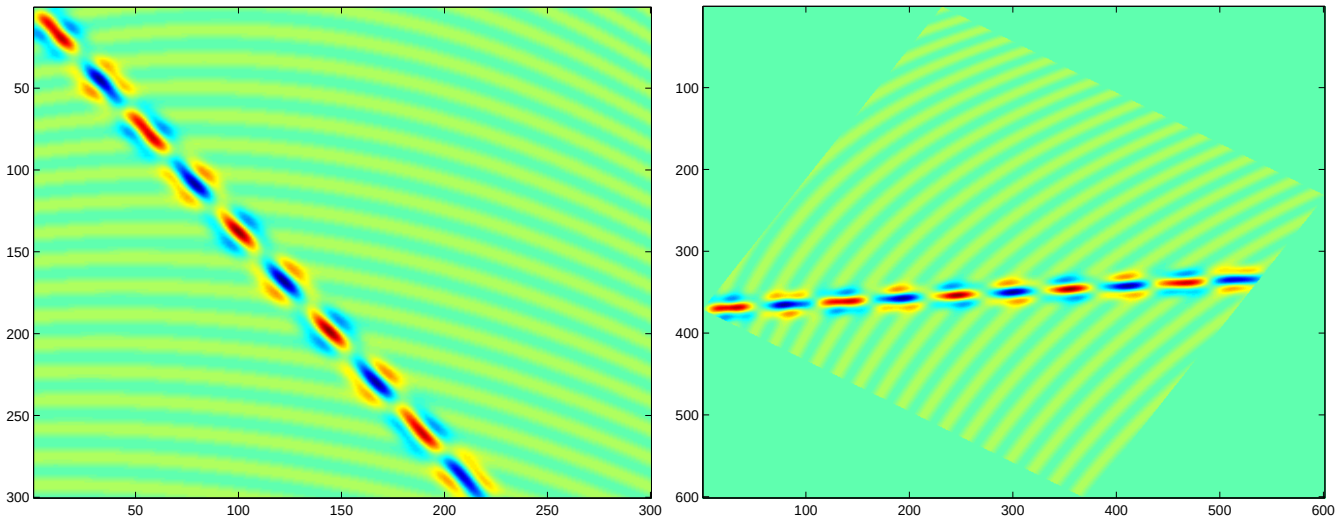

Figure 1. Idealized testproblem (frequency= 4)

synthetic dataset on a grid of dimensions 300×300. The dataset consists of low-amplitude hyperbolic events on which a high-amplitude function g has been superimposed. The picture is a caricature of a seismic shotrecord. The vertical axis corresponds to time, the horizontal axis to a line at the earth's surface. The hyperbolic events would correspond to measurements of seismometers recording waves excited at the earth's surface and reflected at interfaces (i.e. where the impedances changes rapidly). The challenge consists of segregating the high-amplitude distortion from the low-amplitude events of interest. The problem might serve as a benchmark for testing numerical algorithms and is therefore made available[19] at WWW. The dimensions of the grid, the angle and frequency of the distortion (and its width), and other relevant parameters can all be controlled.

5.2. A segregation template

We investigate the applicability of wavelet transforms as an alternative to the Fourier transform. Results of earlier research in this respect are reported in[2,6,8] . What's new in our method is determined by: rotation/interpolation, the transform (of wavelet-type) and the filtering. We devise numerical algorithms for the solution of problems as described in Section 5.1. The methods we discuss can be characterized by the following template with seven subsequent steps:

step 1 Flatten data.

step 2 Rotate data.

step 3 Transform data.

step 4 Adapt (*limit, mute, interpolate*) appropriate coefficients, i.e. filtering.

step 5 Apply backtransform.

step 6 De-rotate data.

step 7 De-flatten data.

Apart from step 2 (and its counterpart step 6) this is quite a conventional framework. Most or all present methods fit in this framework where often 'fast Fourier' is used as transform (step 3). Step 1 (and its counterpart step 7) may be a simple method to compensate for the exponential decay in time of the signal, but more sophisticated preprocessing at this point is quite common. Step 2 (and its counterpart step 6) is new. For reasons to be explained later, we rotate the data such that they more or less align with either horizontal or vertical gridlines. The actual filtering takes place at step 4, it is the heart of the algorithm. In a way, the combination of steps 1–3 can be seen as one big transform with the aim of making the undesired components explicit. The purpose of steps $1-3$ is to transform a signal step by step in such a manner that the coefficients with respect to the new basis of representation can be divided into a set corresponding to the desired components and another set corresponding to the undesired components. Ideally, we then could simply mute the coefficients going with the undesired components. By numerical experience our ways of adapting (step 4) evolved from $\mathcal{A}$ to $\mathcal{C}$:

$\mathcal{A}$ $c := \max(\min(c, \text{upperbound}), \text{lowerbound})$;

$\mathcal{B}$ **if** c satisfies mute-criterion **then** $c := 0$ (mute c);

$\mathcal{C}$ **if** c satisfies mute-criterion **then** replace c by interpolation between nearby coefficients that do not satisfy the mute-criterion.

Here $\mathcal{A}$ is called *limiting*, $\mathcal{B}$ is called *muting*, $\mathcal{C}$ is called *interpolation*. The values 'lowerbound' and 'upperbound' are defined in a problem-dependent way. Often $\mathcal{A}$ turns out to be too crude: the groundroll remains too large. This is improved by $\mathcal{B}$ which is more radical by muting coefficients completely. An important difference of $\mathcal{C}$ with both $\mathcal{A}$ and $\mathcal{B}$ is that the coefficients which need to be adapted are replaced by interpolation between remaining nearby coefficients that (are supposed to) correspond to the wanted components. The mute-criterion in $\mathcal{B}$, $\mathcal{C}$ still needs to be filled in. It can be simple like:

$$\text{mute-criterion} = (c > \text{upperbound} \textbf{ or } c < \text{lowerbound})$$

be it that the values 'lowerbound' and 'upperbound' still need to be determined. More details and results can be found in[20] .

5.2.1. Rotation

Why rotate the data? As with every transform in the context of seismic data processing the aim is to make a better separation between the unwanted components and the events of interest. For the two-dimensional wavelet-decomposition, it has been shown[2] that if the angle between groundroll and the gridlines is within well-defined bounds a bias exists for distribution of the energy towards either horizontal detail or vertical detail wavelet coefficients. These findings supported our research to investigate the effects of rotation. Figure 1 shows how a simple synthetic dataset on a 300×300-grid is rotated (interpolated) onto a 600×600-grid (both grids are uniform). The "right" angle of rotation is chosen visually, in practice we need to rely on a numerical approach to detect directional bias in a dataset. In order to add a touch of realism, the function g is not completely aligned with the grid. Note that the rotated grid wants a smaller meshwidth than the original if we do not allow for loss of information. Generally, we need to be concerned about costs and accuracy of the rotation, see Section 4 and[20] .

5.2.2. A numerical result

We apply the above segregation template to the idealized testproblem of Section 5.1. Steps 1 and 7 are omitted. We choose a compactly supported two-dimensional separable biorthogonal spline wavelet and use a corresponding decomposition of 6 levels. Muted coefficients at step 4 are replaced by interpolation ($\mathcal{C}$). Figure 2 shows the results before and after de-rotation (step 6). If we compare Figure 2 with Figure 1 we observe a reasonable result: the

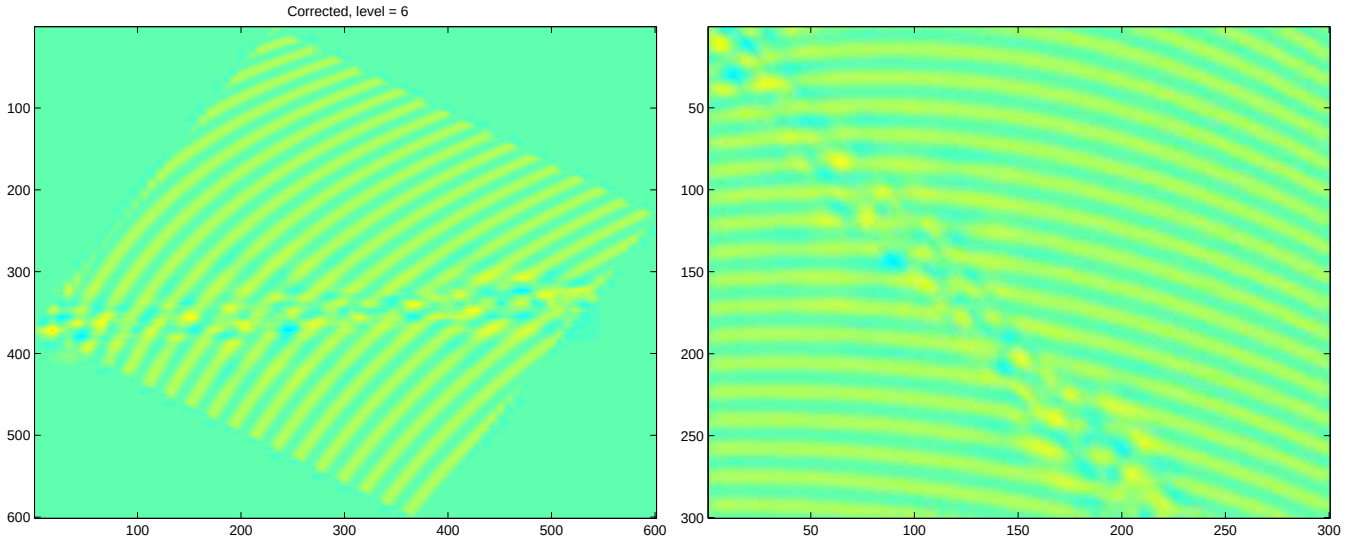

Figure 2. Numerical segregation result for the idealized testproblem (frequency= 4)

high-amplitude distortion has been removed effectively and (nearly) all of the hyperbolic events retain coherence.

6. CONCLUSIONS

The notion of a biorthogonal pair of Riesz systems which induce multiresolution analyses can be defined in the multivariate case. This leads to a fast rotation algorithm which uses biorthogonal (spline) wavelets.

By employing rotation and wavelet decomposition we were able to devise and demonstrate a numerical algorithm that can segregate a high-amplitude distortion with directional bias from low-amplitude hyperbolic events.

ACKNOWLEDGMENTS

This work was supported financially by the Technology Foundation (STW), project no. CWI44.3403.

REFERENCES

1. C.K. Chui, *An introduction to wavelets*, Academic Press, Boston, 1992.
2. J.C. Cohen and T. Chen, "Fundamentals of the Discrete Wavelet Transform for Seismic Data Processing," Preprint, 1993, `www.mathsoft.com/wavelets.html`
3. A. Cohen, I. Daubechies, J.-C. Feauveau, "Biorthogonal bases of compactly supported wavelets," *Comm. Pure Appl. Math.* 45: 485–560, 1992.
4. A. Cohen, R.D. Ryan, *Wavelets and multiscale signal processing*, Chapmann & Hall, London, 1995.
5. I. Daubechies, *Ten lectures on wavelets*, SIAM-CBMS Series 61, Philadelphia, PA, 1992.
6. Bingwen Du, Larry Lines, "Wavelet filtering of tube waves in the Glenn Pool crosswell seismic survey," *M.U.S.I.C. Annual report on wavelet application*, Memorial University of Newfoundland, 1998.
7. I.S. Duff, A.M. Erisman, J.K. Reid, *Direct Methods for Sparse Matrices*, Monographs on Numerical Analysis, Clarendon Press, Oxford, 1986.
8. Liu Faqi, M.M. Nurul Kabir and D.J. Verschuur, "Seismic processing using the wavelet and the Radon transform," *J. of Seismic Exploration* 4, 375–390, 1995.
9. M. Holschneider, *Wavelets: An analysis tool*, Clarendon Press, Oxford, 1995.
10. G. Kaiser, R.F. Streater, "Windowed Radon transforms, analytic signals, and the wave equation," *Wavelets: A tutorial in theory and applications*, C.K. Chui (ed.), pp. 399–441, Academic Press, 1992.
11. A.K. Louis, P. Maass, A. Rieder, *Wavelets: Theory and Applications*, John Wiley & Sons, Chichester, 1997.
12. S.G. Mallat, "A theory for multiresolution signal decomposition,"*IEEE Tr. Pattern Anal. Machine Intelligence* 11: 674–693, 1989.
13. T. Takiguchi, "On invertibility of the windowed Radon transform," *J. Math. Sci. Univ. Tokyo* 2: 621–636, 1995.
14. M. Unser, "Quasi-orthogonality and quasi-projections," *Appl. Comp. Harm. Anal.* 3: 201–214, 1996.
15. M. Unser, "Approximation power of biorthogonal wavelet expansions," *IEEE Tr. Signal Proc.* 44(3): 519–527, 1996.
16. M. Unser, A. Aldroubi, " A general sampling theory for nonideal acquisition devices," *IEEE Tr. Signal Proc.* 42 (11): 2915–2925, 1994.
17. R.M. Young, *An introduction to nonharmonic Fourier series*, Academic Press, New York, 1980.
18. A.L. Warrick, P.A. Delaney, "A wavelet localized Radon transform," SPIE Proceedings 2569, *Wavelet Applications in Signal and Image Processing* III, pp. 632–643, 1995.
19. P.M. de Zeeuw, "A benchmark problem representing a strongly simplified and synthetic shotrecord", `www.cwi.nl/ftp/pauldz/Demos/SimpleSet`
20. P.M. de Zeeuw, R.A. Zuidwijk, "Numerical methods for decomposition of 2D signals by rotation and wavelet techniques", CWI-report PNA-R98xx (in preparation), 1998.
21. R.A. Zuidwijk, "The wavelet X-ray transform", CWI-report PNA-R9703, 1997.
22. R.A. Zuidwijk, "The discrete and continuous wavelet X-ray transform," SPIE proceedings 3169, *Wavelet Applications in Signal and Image Processing* V, pp. 357–366, 1997.
23. R.A. Zuidwijk, P.M. de Zeeuw, "The fast wavelet X-ray transform," CWI-report PNA-R98xx (in preparation), 1998.