A COMMON FORMULATION FOR INTERPOLATION, PREDICTION, AND UPDATE LIFTING DESIGN

Joel Solé and Philippe Salembier

Technical University of Catalonia (UPC)
Dept. of Signal Theory and Communications, Campus Nord, Barcelona, Spain
{joel, philippe}@gps.tsc.upc.edu

ABSTRACT

The optimization of a quadratic objective function with linear constraints is useful for interpolation purposes. This formulation may be employed to derive an initial prediction in the lifting scheme domain in order to construct wavelet transforms. We modify the formulation to design final prediction and update lifting steps. The linear constraints relate wavelet bases and coefficients with the underlying signal. The objective function is the detail signal energy for the prediction lifting design and the gradient of the approximation signal for the update. To report concrete results and the power of the approach, we derive update steps using an auto-regressive image model that show better performance than the 5/3 wavelet for the compression of several image classes.

1. INTRODUCTION

Wavelet transforms have widely shown their usefulness in image compression. The lifting scheme [1] is a method to create biorthogonal wavelet filters from other ones. Basically, lifting consists of prediction and update steps. We briefly introduce lifting in section 2.

The prediction step extracts the redundancy existing in the odd samples from the even samples. Interpolative functions are a reasonable choice as initial prediction lifting step. An example is the family of Deslauriers-Dubuc interpolating wavelets, which are constructed via lifting using two steps. An interesting adaptive quadratic interpolation method is proposed in [2]. We outline it in section 3. The interpolation signal is found in [2] by means of the optimal recovery theory. We have observed that the problem statement may be reformulated as a simple minimization of a quadratic function with linear equality constraints. This insight provides all the resources and flexibility coming from the convex optimization theory to solve the problem. Furthermore, the initial problem statement may be modified in many different ways and convex optimization theory still offers solutions.

In this paper, we employ the new found flexibility to design lifting steps with different criteria than the usual vanishing moments and spectral considerations. First, linear constraints are changed. Transformed coefficients are the inner product of wavelet basis vectors with the signal data. These products are new linear constraints introduced in the formulation. This allows to construct initial prediction steps, as well as the subsequent predictions for which the spatial interpolation interpretation is not straightforward. Section 4 explains this point. Second, the objective function is modified to construct the gradient-minimized update steps in section 5. This case is further developed to offer the practical results in section 6. Auto-regressive image models of first and second order are used to derive specific update steps that successfully compare to the 5/3 wavelet for image compression purposes.

Notation: boldface upper-case letters denote matrices, boldface lower-case letters denote column vectors, upper-case italics denote sets, and lower-case italics denote scalars. Indexes are omitted for short when they are clear from the context.

2. LIFTING SCHEME

Lifting scheme comprises the following parts:

(a) Lazy wavelet transform of the input data $\mathbf{x}$ into two subsignals:
   - An approximation or low-pass signal $\mathbf{l}_0$ formed by the even samples of $\mathbf{x}$.
   - A detail or high-pass signal $\mathbf{h}_0$ formed by the odd samples of $\mathbf{x}$.

(b) Lifting steps, $i = 1..L$.
   - Prediction P_i of the detail signal with the $\mathbf{l}_{i-1}$ samples (1).
   - Update U_i of the approximation signal with the $\mathbf{h}_i$ samples (2).

(c) Output data: the transform coefficients $\mathbf{l}_L$ and $\mathbf{h}_L$.

$$h_i[n] = h_{i-1}[n] - P_i(\mathbf{l}_{i-1}[n]) \quad (1)$$

$$l_i[n] = l_{i-1}[n] + U_i(\mathbf{h}_i[n]) \quad (2)$$

Lifting steps improve the initial lazy wavelet transform properties. Eventually, input data may be any other wavelet transform with some properties we want to improve. Several prediction and updates ($L > 1$) may be concatenated in order to reach the desired properties for the wavelet basis. A multi-resolution decomposition of $\mathbf{x}$ (3) is attained by plugging the approximated signal $\mathbf{l}_L$ into another lifting step block, obtaining $\mathbf{l}^{(2)}$ and $\mathbf{h}^{(2)}$. The process is iterated on $\mathbf{l}^{(k)}$.

$$\begin{aligned} \mathbf{x} \rightarrow (\mathbf{l},\, \mathbf{h}) \rightarrow (\mathbf{l}^{(2)},\, \mathbf{h}^{(2)},\, \mathbf{h}) \rightarrow \ldots \rightarrow \\ \rightarrow (\mathbf{l}^{(K)},\, \mathbf{h}^{(K)},\, \mathbf{h}^{(K-1)}, \ldots,\, \mathbf{h}) \end{aligned} \qquad (3)$$

3. QUADRATIC INTERPOLATION

An interpolation method based on the quadratic signal class determined from the local image behavior is presented in [2]. The quadratic signal class is determined by a set of patches $\mathcal{S} = \{x_1, \ldots, x_m\}$ representative of the local data. Patches may extracted from an up-sampling and filtering of the image or from other images. Patches are high-density, i.e., they have the same resolution as the interpolation. The quadratic class is defined by a matrix $\mathbf{Q}$ for which the ellipsoid

$$\mathbf{x}^T \mathbf{Q} \mathbf{x} \leq \epsilon \qquad (4)$$

must be representative of the training set $\mathcal{S}$, i.e., $\mathbf{Q}$ must be a matrix such that when an image patch $\mathbf{y}$ is similar to the vectors in $\mathcal{S}$, then (4) holds for $\mathbf{y}$. Matrix $\mathbf{S}$ is formed by arranging the image patches in $\mathcal{S}$ as columns: $\mathbf{S} = (\mathbf{x}_1 \ldots \mathbf{x}_m)$.

The image patch $\mathbf{y}$ is imposed to be a linear combination of the training set $\mathcal{S}$ through a column vector $\mathbf{c}$:

$$\mathbf{S}\mathbf{c} = \mathbf{y}. \qquad (5)$$

Vectors in $\mathbf{S}$ are similar among them and $\mathbf{y}$ is similar to them when $\mathbf{c}$ has small energy,

$$\|\mathbf{c}\|^2 = \mathbf{c}^T\mathbf{c} = \mathbf{y}^T(\mathbf{S}\mathbf{S}^T)^{-1}\mathbf{y} = \mathbf{y}^T\mathbf{Q}\mathbf{y} \leq \epsilon,$$

where $\mathbf{Q}$ is the pseudo-inverse of $\mathbf{S}\mathbf{S}^T$. In this sense, good interpolators $\mathbf{y}$ for the quadratic class determined by $\mathbf{Q}$ are expanded with the weighting vectors $\mathbf{c}$ of energy bounded by some ϵ. Once the high density class $\mathcal{S}$ is determined, the optimal interpolated vector $\mathbf{x}$ can be simply seen as the solution of (6), instead of using the optimal recovery theory as in [2]. The solution of (6) is the minimum energy $\mathbf{c}$ subject to the patches linear constraint (5). This interpretation is very useful for the lifting design covered in the subsequent sections.

$$\begin{aligned} &\underset{\mathbf{x},\mathbf{c}}{\text{minimize}} && \|\mathbf{c}\|^2 \\ &\text{subject to} && \mathbf{S}\mathbf{c} = \mathbf{x} \end{aligned} \qquad (6)$$

There is more information that may be used to improve the solution. Previous knowledge about $\mathbf{x}$ is available. Typically, one of every two elements of $\mathbf{x}$ are already known in the interpolation of $\mathbf{h}_0$ with the samples $\mathbf{l}_0$. It may be also known that the original high density pixels have been averaged before a decimation by two. Both cases impose a linear constraint on the data, denoted by $\mathbf{A}^T\mathbf{x} = \mathbf{b}$. In the first case, the columns of matrix $\mathbf{A}$ are formed by vectors $\mathbf{e}_i$, being the one located at the position of the known sample. The respective position of vector $\mathbf{b}$ has the sample value. An illustrative example for the second case is the following. Assume that the pixel value is the average of four high density neighbors, then there would be a $1/4$ at each of their corresponding positions in a column of $\mathbf{A}$. Whatever the linear constraints, they are included in (6) to reach the formulation

$$\begin{aligned} &\underset{\mathbf{x},\mathbf{c}}{\text{minimize}} && \|\mathbf{c}\|^2 \\ &\text{subject to} && \mathbf{S}\mathbf{c} = \mathbf{x} \\ & && \mathbf{A}^T\mathbf{x} = \mathbf{b}. \end{aligned} \qquad (7)$$

The solution of this problem is

$$\mathbf{x}^\star = \mathbf{S}\mathbf{S}^T\mathbf{A}(\mathbf{A}^T\mathbf{S}\mathbf{S}^T\mathbf{A})^{-1}\mathbf{b}, \qquad (8)$$

which is the least-square solution for the quadratic norm determined by $\mathbf{S}\mathbf{S}^T$ and the linear constraints $\mathbf{A}^T\mathbf{x} = \mathbf{b}$. Taking the expectation in (8) the formulation is made *global*. In this case, the quadratic class is determined by the correlation matrix $\mathbf{R} = \mathbb{E}\left[\mathbf{S}\mathbf{S}^T\right]$. The equivalent global formulation of (7) is

$$\begin{aligned} &\underset{\mathbf{x}}{\text{minimize}} && \mathbf{x}^T\mathbf{R}^{-1}\mathbf{x} \\ &\text{subject to} && \mathbf{A}^T\mathbf{x} = \mathbf{b} \end{aligned} \qquad (9)$$

and the corresponding solution is

$$\mathbf{x}^\star = \mathbf{R}\mathbf{A}(\mathbf{A}^T\mathbf{R}\mathbf{A})^{-1}\mathbf{b}. \qquad (10)$$

With this common formulation, local adapted and global interpolative predictions may be constructed and the available knowledge about the input signal may be incorporated by adding constraints to refine the result. In the next section, the linear constraints in (9) are modified to include the transform coefficients inner products in order to construct final prediction steps.

4. LINEAR PREDICTION STEP DESIGN

A transform coefficient i is the inner product of a wavelet or scaling basis vector $\mathbf{w}_i$ with the input signal. Using this notation, coefficients $h[n]$ and $l[n]$ arise from $h[n] = \mathbf{w}_{h[n]}^T\mathbf{x}$ and $l[n] = \mathbf{w}_{l[n]}^T\mathbf{x}$, respectively. A second prediction step P_2 predicts a coefficient $h_1[n]$ using a set of neighboring approximate samples, which are denoted by $\mathbf{l}_1[n]$ with some notation abuse. The operators are linear and so, we have

$$h_2[n] = h_1[n] - \hat{h}_1[n] = h_1[n] - P_2(\mathbf{l}_1[n]) = h_1[n] - \mathbf{p}_2^T\mathbf{l}_1[n].$$

The approximate coefficients linear constraints are included in the formulation. Therefore, matrix $\mathbf{A}$ columns are formed by $\mathbf{w}_{l_1[n]}$, which are the basis vectors of each neighbor $l_1[n]$ employed for the prediction. The independent term is $\mathbf{b} = \mathbf{l}_1[n]$. According to the established notation and constraints, the predicted value $\hat{h}_1[n]$ is found by using the optimal interpolation vector (10),

$$\hat{h}_1[n] = \mathbf{w}_{h_1[n]}^T\mathbf{x}^\star = \mathbf{w}_{h_1[n]}^T\mathbf{R}\mathbf{A}(\mathbf{A}^T\mathbf{R}\mathbf{A})^{-1}\mathbf{b} = \mathbf{p}_2^T\mathbf{b},$$

from which the optimal prediction filter is

$$\mathbf{p}_2^\star = (\mathbf{A}^T\mathbf{R}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{R}\mathbf{w}_{h_1[n]}. \quad (11)$$

Interestingly, this filter (11) is equivalent to the one in [3] that minimizes the MSE of the second prediction, that is,

$$\mathbf{p}_2^\star = \arg\min_{\mathbf{p}_2} f_0(\mathbf{p}_2) = \mathbb{E}[(h_1[n] - \hat{h}_1[n])^2]. \quad (12)$$

However, the convex optimization theory approach permits modifications that allow the inclusion of more knowledge in the formulation, with the use of other objective functions (as the update designs of section 5) or with the addition of constraints. For instance, equality and inequality linear constraints on the smoothness of the signal or on its lower and upper bounds may be included. In general, prediction step is easier to design than the update because the spatial interpretation of the prediction filtering is more direct. For this reason, we devote the rest of the paper to the design of update steps and to report the results obtained from their application.

5. LINEAR UPDATE STEP DESIGN

Previous formulation with an appropriate objective function is applied to design update lifting steps. The next two subsections develop two different objective functions to construct update steps that are used to perform the experiments in section 6.

5.1. First design

Coefficient $l[n]$ is updated with $\tilde{l}[n] = \mathbf{u}^T\mathbf{b}$. The objective function is set to be the l^2-norm of the substraction between the updated coefficient $l[n] + \tilde{l}[n]$ and the set $\mathcal{I}$ of the neighboring scaling coefficients. This objective function leads to a smooth approximate signal that helps the prediction to perform better in the next resolution level. Formally stated, the goal is to find $\mathbf{u}$ such that

$$\mathbf{u}^\star = \arg\min_{\mathbf{u}} f_0(\mathbf{u}),$$

with

$$f_0(\mathbf{u}) = \mathbb{E}\left[\sum_{i\,\in\,\mathcal{I}} (l[i] - (l[n] + \tilde{l}[n]))^2\right], \quad (13)$$

which is equivalent to

$$f_0(\mathbf{u}) = \sum_{i\,\in\,\mathcal{I}} \mathbb{E}\left[(\mathbf{w}_{l[i]}^T\mathbf{x} - \mathbf{w}_{l[n]}^T\mathbf{x} - \mathbf{u}^T\mathbf{b})^2\right], \quad (14)$$

where now $\mathbf{b} = \mathbf{h}_1[n]$. Expression (14) is developed. Then differentiated with respect to $\mathbf{u}$. After that, the linear constraints $\mathbf{A}^T\mathbf{x} = \mathbf{b}$ are introduced and the definition of correlation matrix used to reach the expression

$$\nabla_{\mathbf{u}} f_0 = 2\sum_{i\,\in\,\mathcal{I}} \mathbf{u}^T\mathbf{A}^T\mathbf{R}\mathbf{A} + \mathbf{w}_{l[n]}^T\mathbf{R}\mathbf{A} - \mathbf{w}_{l[i]}^T\mathbf{R}\mathbf{A}.$$

Let denote the mean of the neighboring approximate signal basis vectors employed to update as

$$\overline{\mathbf{w}}_{\mathcal{I}} = \frac{1}{|\mathcal{I}|}\sum_{i\,\in\,\mathcal{I}} \mathbf{w}_{l[i]},$$

where $|\mathcal{I}|$ denotes the cardinal of the set $\mathcal{I}$. Equalling the derivative to zero, the optimal update filter minimizing the local gradient is found to be

$$\mathbf{u}^\star = (\mathbf{A}^T\mathbf{R}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{R}(\overline{\mathbf{w}}_{\mathcal{I}} - \mathbf{w}_{l[n]}), \quad (15)$$

and the optimally updated coefficient is

$$l_1[n] = l_0[n] + (\overline{\mathbf{w}}_{\mathcal{I}} - \mathbf{w}_{l[n]})^T\mathbf{R}\mathbf{A}(\mathbf{A}^T\mathbf{R}\mathbf{A})^{-1}\mathbf{b}. \quad (16)$$

Again, the interpretation relying on the optimal interpolation of $\mathbf{x}$, $l_1[n] = l_0[n] + \mathbf{u}^{\star T}\mathbf{b} = l_0[n] + (\overline{\mathbf{w}}_{\mathcal{I}} - \mathbf{w}_{l[n]})^T\mathbf{x}^\star$, is practical because it allows the use of additional knowledge. A related construction is developed in the next subsection.

5.2. Second design

An additional consideration on the set of approximation signal neighbors $\mathcal{I}$ may be included to the previous gradient-minimization design. As each sample in $\mathcal{I}$ is also updated, it is interesting to consider the minimization of the gradient of $l[n] + \tilde{l}[n]$ with respect to the updated samples $l[i] + \tilde{l}[i]$, $i \in \mathcal{I}$, through a still unknown update filter. To this goal, the objective function (13) is modified in order to find the optimal update with this criterion:

$$f_0(\mathbf{u}) = \mathbb{E}\left[\sum_{i\,\in\,\mathcal{I}} ((l[i] + \tilde{l}[i]) - (l[n] + \tilde{l}[n]))^2\right].$$

The objective function is expanded taking into account the updated coefficients bases $\tilde{\mathbf{w}}_{l[i]} = \mathbf{w}_{l[i]} + \mathbf{A}_{l[i]}\mathbf{u}$, being $\mathbf{A}_{l[i]}$ the constraint matrix relative to the position of sample $l[i]$ and $\mathbf{A} = \mathbf{A}_{l[n]}$. The algebraic manipulation is similar to the previous case. The optimal solution is described by expression

$$\mathbf{u}^\star = \mathbf{M}^{-1}(\mathbf{A}^T\mathbf{R}(\overline{\mathbf{w}}_{\mathcal{I}} - \mathbf{w}_{l[n]}) + \overline{\mathbf{A}}_{\mathcal{I}}^T\mathbf{R}\mathbf{w}_{l[n]} - \overline{\mathbf{b}}_{\mathcal{I}}), \quad (17)$$

being

$$\mathbf{M} = \mathbf{A}^T\mathbf{R}(\mathbf{A} - 2\overline{\mathbf{A}}_{\mathcal{I}}) + \overline{\mathbf{R}}_{\mathcal{I}},$$

where the mean of the different products of the bases and matrices are denoted by

$$\begin{aligned}\overline{\mathbf{A}}_{\mathcal{I}} &= \frac{1}{|\mathcal{I}|}\sum_{i\,\in\,\mathcal{I}} \mathbf{A}_{l[i]},\\ \overline{\mathbf{R}}_{\mathcal{I}} &= \frac{1}{|\mathcal{I}|}\sum_{i\,\in\,\mathcal{I}} \mathbf{A}_{l[i]}^T\mathbf{R}\mathbf{A}_{l[i]},\\ \overline{\mathbf{b}}_{\mathcal{I}} &= \frac{1}{|\mathcal{I}|}\sum_{i\,\in\,\mathcal{I}} \mathbf{A}_{l_0[i]}^T\mathbf{R}\mathbf{w}_{l_0[i]}.\end{aligned}$$

Equation (17) is very simple to compute in practice. The only differences w.r.t. (15) are the additional terms concerning the mean of the neighbors basis vectors, which are known.

	5/3 wavelet	AR-1 model
Synthetic	3.832	3.508
SST	3.252	3.123
Mammography	2.349	2.358

Table 1. Compression results with Jpeg2000 using the standard 5/3 wavelet and the proposed optimal update with the AR-1 model for the synthetic, mammography, and SST image classes. Results are given in bits per pixel.

6. EXPERIMENTS

The framework developed in this paper allows the construction of 2-D nonseparable filters. However, for the experiments we restrict ourselves to 1-D separable decompositions. Concretely, we derive update steps for the prediction $\mathbf{p}_1 = (\ 1/2 \quad 1/2\ )^T$ of the 5/3 wavelet used for lossy-to-lossless compression in the Jpeg2000 standard [4]. The 5/3 wavelet update is $\mathbf{u}_1 = (\ 1/4 \quad 1/4\ )^T$. For fair comparison, we also employ two neighbors for the update and so, in practice this application simply reduces to propose a coefficient different from $1/4$ for the update filter. Even in this simple case, the proposal attains noticeable improvements.

In the first experiment, a second order auto-regressive model (AR-2) is used to determine the local image behavior. For a subset of the AR-2 parameters values, the resulting optimal update coefficient (17) coincides with the 5/3 update, but not for other possible values. Figure 1 highlights this fact. It relates the update coefficient with the AR-2 parameters. Therefore, for many images the usual update is far from being optimal in the sense of (17). An adaptive update filter is constructed by estimating the AR-2 parameters for each line in an image and using the filter given by the second design (17). To assess the performance, the energy of the coarser level detail signal $\mathbf{h}_1^{(2)}$ is computed. This comes from the assumption that a good approximation signal for compression provides small prediction energy. For a wide set of natural images, like Lenna and Cameraman, the energy is up to 25% smaller for the adaptive optimal update step w.r.t. the 5/3 update.

The second experiment derives filters applicable to a more global setting. The AR model of first order is estimated for three image classes and so, each model is useful for a whole corpus of images instead of being local. 15 synthetic images, 15 mammography and 6 sea surface temperature (SST) images are used. The correlation matrix is determined by the AR-1 parameter and then it is plugged into the first design (15) to obtain an update filter used for all the images in the class. Image compression is performed with a four resolution level decomposition within the Jpeg2000 coder environment. Numerical results appear in table 1 compared to the 5/3 wavelet. For the mammography, compression slightly worsens, maybe due to the background statistics, but for the two other classes, the proposal results improve those of the 5/3.

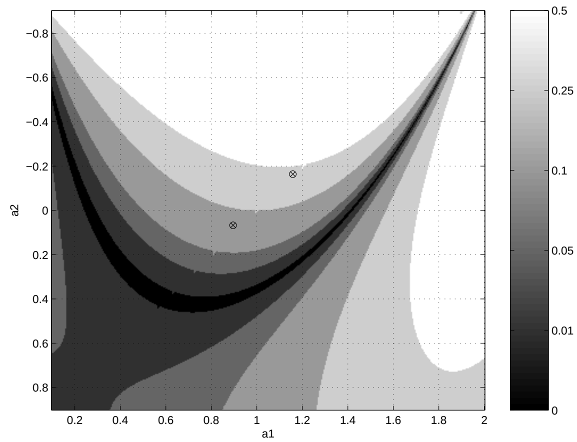

Fig. 1. Six level-sets of a function of the update coefficient with respect to the AR-2 parameters. The function is the absolute value of the update coefficient minus 0.25 (the 5/3 wavelet update coefficient). Thus, the resulting filter is very similar to the 5/3 in the dark areas and different in the light areas. The two circles depict the mean AR-2 parameters for the synthetic and SST image classes.

7. CONCLUSIONS

The relation between interpolation and lifting design has been highlighted. The common setting is employed to derive optimal lifting steps with different criteria. We show two different ways to obtain useful updates and apply the results to a concrete case with success. Both, the local adaptive and the fixed update provide encouraging results. Furthermore, from the developed viewpoint, it is demonstrated that the widespread 5/3 wavelet is not optimal for many images. Despite of these examples, the framework offers additional flexibility that deserves to be deeply studied and employed in all its extension in a future work.

8. REFERENCES

[1] W. Sweldens, “The lifting scheme: A custom-design construction of biorthogonal wavelets,” *Appl. Comput. Harmon. Anal.*, vol. 3, no. 2, pp. 186–200, 1996.

[2] D. D. Muresan and T. W. Parks, “Adaptively Quadratic (AQua) Image Interpolation,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 690–698, May 2004.

[3] A. T. Deever and S. S. Hemami, “Lossless image compression with projection-based and adaptive reversible integer wavelet transforms,” *IEEE Transactions on Image Processing*, vol. 12, no. 5, pp. 489–499, May 2003.

[4] ISO/IEC, *ISO/IEC 15444-1: JPEG 2000 image coding system*, 2000.