Wavelet-based Besov space penalty regularization

Samuli Siltanen February 2015.

1 Haar wavelet transform in 1D

1.1 The Haar transform for functions

Consider real-valued functions defined on the interval $[0, 1]$. There are two especially important functions, namely the *scaling function* $\varphi(x)$ and the *mother wavelet* $\psi(x)$, defined as follows:

$$\varphi(x) \equiv 1, \qquad \psi(x) = \begin{cases} 1, & \text{for } 0 \leq x < 1/2, \\ -1 & \text{for } 1/2 \leq x \leq 1. \end{cases}$$

Also, let us define *wavelets* as scaled and translated versions of the mother wavelet:

$$\psi_{jk}(x) := 2^{j/2}\psi(2^j x - k) \qquad \text{for } j \geq 0 \text{ and } 0 \leq k \leq 2^j - 1.$$

Let $f, g : [0, 1] \to \mathbb{R}$. Define the inner product between f and g by

$$\langle f, g\rangle := \int_0^1 f(x)\overline{g(x)}\, dx. \tag{1}$$

Then we have orthogonality:

$$\langle \psi_{jk}, \psi_{j'k'}\rangle = \begin{cases} 1 & \text{if } j = j' \text{ and } k = k', \\ 0 & \text{otherwise.} \end{cases}$$

Define the "detail" wavelet coefficients of a function f as follows:

$$d_{jk} := \langle f, \psi_{jk}\rangle, \qquad \text{for } j \geq 0 \text{ and } 0 \leq k \leq 2^j - 1, \tag{2}$$

and the average coefficient as

$$c_0 := \langle f, \varphi\rangle. \tag{3}$$

Then we can express f in terms of wavelets like this:

$$f(x) = c_0\varphi(x) + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} d_{jk}\psi_{jk}. \tag{4}$$

1.2 The Haar transform for discrete signals

Let $n = 2^m$. Given a function $f : [0,1] \to \mathbb{R}$, denote its samples at n points as follows:

$$\mathbf{f}_\nu := f(x_\nu), \qquad \text{with } x_\nu = \frac{\nu - 1}{n} \text{ for } \nu = 1, \ldots, n. \tag{5}$$

We will use the vector notation $\mathbf{f} = [\mathbf{f}_1, \ldots, \mathbf{f}_n]^T$. Then the wavelet transform can be implemented as a matrix-vector product $\mathbf{w} = W\mathbf{f}$.

Let us illustrate the structure of the vector $\mathbf{w}$ by a small example. Take $n = 8$. Then

$$\mathbf{w} = [d_{20}, d_{21}, d_{22}, d_{23}; d_{10}, d_{11}; d_{00}; c_0]^T = W\mathbf{f}. \tag{6}$$

1.3 Besov space norms

The general Besov space norm can be written as [7]

$$\|f\|_{B^s_{pq}} := \left(|c_0|^q + \sum_{j=0}^{\infty} 2^{jq(s+\frac{1}{2}-\frac{1}{p})} \left(\sum_{k=0}^{2^j-1} |d_{j,k}|^p \right)^{\frac{q}{p}} \right)^{\frac{1}{q}},$$

where $s \in \mathbb{R}$ and $1 \leq p, q < \infty$. Actually the parameter s has to satisfy $s < r$ where r is the regularity of the mother wavelet. However, we will not care about this below.

Our main interest here will be the space B^1_{11} in dimension 1, whose norm is

$$\|f\|_{B^1_{11}} = |c_0| + \sum_{j=0}^{\infty} \sum_{k=0}^{2^j-1} 2^{j/2} |d_{j,k}|. \tag{7}$$

Now the Haar basis is not smooth enough for the theory to hold, but we do not care.

In the discrete case (7) takes the form

$$\|\mathbf{f}\|_{B^1_{11}} = \|B\mathbf{w}\|_1. \tag{8}$$

Let us illustrate the structure of the weight matrix B using example (6). It is then

$$B = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \sqrt{2} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \sqrt{2} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

2 The minimization problem

2.1 Wavelet-based approach

Consider the minimization problem

$$\widetilde{\mathbf{f}} := \arg\min_{\mathbf{f}\in\mathbb{R}^n} \Big\{ \|A\mathbf{f}-\mathbf{m}\|_2^2 + \alpha' \|\mathbf{f}\|_{B^1_{11}} \Big\}, \tag{9}$$

where $\|\cdot\|_{B^1_{11}}$ denotes Besov space norm. This is interesting since the Besov space B^1_{11} is related to the Total Variation space but has different properties. See [2, 5, 4, 3] and [6, Chapter 7].

Roughly speaking, we can determine the wavelet coefficient vector $\widetilde{\mathbf{w}} = W\widetilde{\mathbf{f}}$ by solving this minimization problem:

$$\widetilde{\mathbf{w}} := \arg\min_{\mathbf{w}\in\mathbb{R}^n} \Big\{ \|AW^T\mathbf{w}-\mathbf{m}\|_2^2 + \alpha \|B\mathbf{w}\|_1 \Big\}, \tag{10}$$

where B is a diagonal weight matrix.

We do not discuss the relationship between the regularization parameters $\alpha > 0$ and $\alpha' > 0$ further in this short note; since there are many equivalent norms for the space B^1_{11}, it is not straightforward how α and α' should be related for $\widetilde{\mathbf{w}} = W\widetilde{\mathbf{f}}$ to hold.

See [1] for more information on wavelets.

2.2 Quadratic reformulation

We want to determine numerically the vector $\widetilde{\mathbf{w}} \in \mathbb{R}^n$ that solves (10). We write the vector $\mathbf{w} \in \mathbb{R}^n$ in the form

$$B\mathbf{w} = \mathbf{v}_+ - \mathbf{v}_-,$$

where $\mathbf{v}_\pm$ are nonnegative vectors: $\mathbf{v}_\pm \in \mathbb{R}^n_+$, or $(\mathbf{v}_\pm)_j \geq 0$ for all $j = 1, \ldots, n$. Now minimizing (10) is equivalent to minimizing

$$\|AW^T\mathbf{w}\|_2^2 - 2\mathbf{m}^T AW^T\mathbf{w} + \alpha\mathbf{1}^T\mathbf{v}_+ + \alpha\mathbf{1}^T\mathbf{v}_-,$$

where $\mathbf{1}$ is the vector with all elements equal to one: $\mathbf{1} = \begin{bmatrix} 1 & 1 & \cdots & 1 \end{bmatrix}^T \in \mathbb{R}^n$, and the minimization is taken over $y \in \mathbb{R}^{3n}$ defined by

$$\mathbf{y} = \begin{bmatrix} \mathbf{w} \\ \mathbf{v}_+ \\ \mathbf{v}_- \end{bmatrix}, \quad \text{where} \quad \begin{array}{l} \mathbf{w} \in \mathbb{R}^n \\ \mathbf{v}_+ \in \mathbb{R}^n_+ \\ \mathbf{v}_- \in \mathbb{R}^n_+ \end{array}.$$

Note the identity $\|AW^T\mathbf{w}\|_2^2 = \mathbf{w}^T WA^T AW^T\mathbf{w}$ and write

$$H = \begin{bmatrix} 2WA^TAW^T & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \qquad \mathbf{h} = \begin{bmatrix} -2WA^T\mathbf{m} \\ \alpha\mathbf{1} \\ \alpha\mathbf{1} \end{bmatrix}.$$

We then have the quadratic optimization problem in standard form

$$\arg\min_{\mathbf{y}} \left\{ \frac{1}{2}\mathbf{y}^T H \mathbf{y} + \mathbf{h}^T \mathbf{y} \right\} \tag{11}$$

with the constraints

$$\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} y_{n+1} \\ \vdots \\ y_{2n} \end{bmatrix} - \begin{bmatrix} y_{2n+1} \\ \vdots \\ y_{3n} \end{bmatrix} \tag{12}$$

and

$$y_j \geq 0 \text{ for } j = n+1, \ldots, 3n. \tag{13}$$

References

[1] I Daubechies. *Ten lectures on wavelets (Ninth printing, 2006)*, volume 61 of *CBMS-NSF Regional conference series in applied mathematics*. SIAM, 2006.

[2] I. Daubechies, M. Defrise, and C. De Mol. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Communications on pure and applied mathematics*, 57(11):1413–1457, 2004.

[3] Keijo Hämäläinen, Aki Kallonen, Ville Kolehmainen, Matti Lassas, Kati Niinimaki, and Samuli Siltanen. Sparse tomography. *SIAM Journal on Scientific Computing*, 35(3):B644–B665, 2013.

[4] V. Kolehmainen, M. Lassas, K. Niinimäki, and S. Siltanen. Sparsity-promoting Bayesian inversion. *Inverse Problems*, 28(2):025005, 2012.

[5] M. Lassas, E. Saksman, and S. Siltanen. Discretization-invariant Bayesian inversion and Besov space priors. *Inverse Problems and Imaging*, 3:87–122, 2009.

[6] J.L. Mueller and S. Siltanen. *Linear and Nonlinear Inverse Problems with Practical Applications*. SIAM, 2012.

[7] H. Triebel. *Function spaces and wavelets on domains*, volume 7 of *Tracts in Mathematics*. European Mathematical Society, 2008.