Lectures on Optimization, Image Processing,and Interpolation Theory
Bj¨orn Jawerth and Mario Milman

Lecture notes from Prague

Chapter 5

The Fundamental Lemma and Multiscale Approximation of Images

5.1 Introduction

The motivation for our discussion in this lecture comes from Image Processing. Let us agree that generally speaking we can represent the greyscale of given image as a function $f \in L^2(R^2)$. Further characteristics of an image can be found by splitting f in terms of functions in different subspaces of L^2. A known strategy for analysis of images, apparently initiated by Munford-Shah [39], are the so called "$u + v$ models". An example of such a model was given in a celebrated papers by Rudin-Osher-Fatemi [47] and Rudin-Osher [46] In these models, the authors split an image as , $f = u + v$, so that the contour discontinuities are carried by $u \in BV =$ (bounded variation[1]) and the noisy part is carried by v. Apparently independently from interpolation/approximation theory, Rudin-Osher-Fatemi set up a penalty scheme in order to offset one characteristic versus the other. Their penalty has the form

$$OR(\lambda) = \inf_{f=u+v} \{\|u\|_{BV} + \lambda \|v\|_{L^2}^2\}, \tag{5.1}$$

where λ is some fixed parameter, which acts as a threshold or filter to cut off the noisy part of f. Of course the reader of these notes will recognize $OR(\lambda)$ as

[1]Using the homogeneous norm

$$\|f\|_{BV} = \sup_h \frac{\|f(x+h) - f(x)\|_{L^1}}{|h|}$$

a K−functional, namely[2]

$$OR(\lambda) = K(\lambda, f, BV, (L^2)^2).$$

In their scheme Rudin-Osher-Fatemi[3] call $\|v\|_{L^2}^2$ the"fidelity" term and $\|u\|_{BV}$ is the regularizing term. For a given λ, this model produces a unique minimizer (cf. [36])

$$f = u_\lambda + v_\lambda.$$

To proceed further one needs to fine tune the value of the crucial parameter λ. For example, if the value of the parameter λ is too small most of the image will be kept in the L^2 component, v_λ that will be very close to the original f and as a result we will achieve very little compression. On the other hand if λ is very large we will end up keeping only a cartoon like representation of the image $u_\lambda \in BV$ while the texture and oscillatory details will be kept in the v_λ component. Good choices of λ can be estimated using statistical information on the noise if is known. The limitation of the scheme is that one is limited by one scale determined by λ.

For the computation of $OR(\lambda)$ a very precise analysis of the optimal decomposition was given by Meyer [36]

Theorem 28 *(i) Let $f \in L^2(R^2)$, and suppose that*

$$\|f\|_{W_\infty^{-1}} = \inf\{\||g|\|_\infty : f = \partial_x g_1 + \partial_y g_2, g = (g_1, g_2), |g| = (g_1^2 + g_2^2)^{1/2}\} \leq \frac{1}{2\lambda}.$$

Then the optimal decomposition of f achieving (5.1) is given by

$$f = 0 + f.$$

(ii) If on the other hand $\|f\|_{W_\infty^{-1}} > \frac{1}{2\lambda}$, then the optimal decomposition of $f = u + v$ achieving (5.1) is characterized by

$$\|v\|_{W_\infty^{-1}} = \frac{1}{2\lambda}, \int u(x)v(x)dx = \frac{1}{2\lambda}\|u\|_{BV}.$$

This result plays a fundamental role in the next development due to Tadmor-Nezzar-Vese [51]. These authors deviced a multiscale approach to the optimization problem (5.1). The idea of these authors is that if we have a decomposition at level λ

$$f = u_\lambda + v_\lambda, \tag{5.2}$$

[2]The main reason for using $(L^2)^2$ rather than L^2 is that the variational problem is easier to work with $(L^2)^2$. We also note that while $(L^2)^2$ is not normed: $\|u\|_2^2$ is not homogeneous and fails the triangle inequality. But $\|u\|_2^2$ is a "quasi norm" (cf. [1]): $\|-u\|_2^2 = \|u\|_2^2 \geq 0$, $\|u\|_2^2 = 0 \Leftrightarrow u = 0$, and $\|u+v\|_2^2 \leq 2\left(\|u\|_2^2 + \|v\|_2^2\right)$. Such objects have been studied intensively and incorporated to Interpolation theory (cf. [1]). In fact by means of using the Aoki-Rolewicz theorem one can always replace a quasi-norm by another one that satisfies the triangle inequality.

[3]Meyer [36] has interpreted this method observing that u_λ extracts the edges from f while v_λ captures the textures.

we can further improve the resolution by extracting the edges of the L^2 component v_λ at the next level say 2λ (remember that this increases the filter and thus makes the next L^2 component smaller by placing more of the decomposition (edges) on the BV component: in other words starting with (5.2) we proceed to find an optimal decomposition of v_λ :

$$v_\lambda = u_{2\lambda} + v_{2\lambda} \text{ for } K(2\lambda, v_\lambda, BV, (L^2)^2) = \inf_{v_\lambda = u+v} \{\|u\|_{BV} + 2\lambda \|v\|_{L^2}^2\}.$$

Continuing in this fashion leads to a "multiscale hierarchical decomposition" of the image f

$$f = \sum u_j, \tag{5.3}$$

where the "dyadic blocks" $u_j = u_j(f)$, capture different scales of the original image. In their work [51] Tadmor-Nezzar-Vese have quantified the multiscale nature of the expansion and studied its convergence. The analysis of the convergence of the algorithm is based on careful applications of Theorem 28.

In this lecture we take up to further extend and clarify the analysis of [51]. First we will show that very much like the Rudin-Osher-Fatemi scheme $OR(\lambda)$ is a $K-$functional penalty method, the Tadmor-Nezzar-Vese is based on a variant of the $J-$method which we develop in detail. As a consequence we treat the multilevel decomposition as a variant of the Fundamental Lemma of Interpolation Theory. As a consequence we can treat general pairs of spaces rather than the pair (BV, L^2) and we avoid the use of Hilbert methods or duality (cf. Theorem 28 above).

Moreover, the realization that the approximation scheme of [51] is connected with the Fundamental Lemma has other advantages. In particular we compare the Tadmor et al method with the usual Fundamental Lemma. The experimental results show apparently little difference between the Tadmor et al fundamental lemma and the usual fundamental lemma. In particular we note that the usual fundamental lemma allows us to achieve rapid convergence given the telescoping nature of the decompositions.

One drawback of the results of Tadmor-Nezzar-Vese [51] is that the convergence of the "multiscale hierarchical decomposition" given by (5.3) can only be achieved for f that belong to real interpolation spaces between BV and L^2, the convergence for general $f \in L^2$ was left open. Using the strong form of the fundamental lemma one can easily achieve convergence even in the limiting case $f \in L^2$, thus resolving a question asked in [51].

The last sections of this Chapter include the results of the numerical experiments. We refer to a forthcoming paper [14] for further details.

5.2 Multiscale Decompositions and the Fundamental Lemma

In this section we set up a general real interpolation/approximation theory approach to the analysis of convergence of "multiscale hierarchical decomposition" of Tadmor-Nezzar-Vese [51].

We start by briefly reviewing the basic idea of [51]. As noted above in their work [51] Tadmor-Nezzar-Vese deal with the pair $(BV(R^2), L^2(R^2))$. Let us also remark that by the Sobolev embedding theorem we have $BV(R^2) \subset L^2(R^2)$. Moreover (cf. [36], for each $f \in L^2$, and for each $t > 0$ we can always find an optimal decomposition $f = f_{bv,t} + f_{2,t}$ such that

$$K(t, f; BV(R^2), L^2(R^2)) = \|f_{bv,t}\|_{BV} + t\|f_{2,t}\|_{L^2}. \tag{5.4}$$

Let us now replace the pair $(BV(R^2), L^2(R^2))$ by a pair of quasi-Banach spaces (X_0, X_1), such that $X_0 \subset X_1$. We will assume that the quasi norms satisfy the 1−triangle inequality; moreover we assume that for each $x \in X_1, t > 0$, we can always find an optimal decomposition such that $x = x_{0,t} + x_{1,t}$,

$$K(t, x; X_0, X_1) = \|x_{0,t}\|_{X_0} + t\|x_{1,t}\|_{X_1}.$$

In fact, it is easy to eliminate these extra restrictions (cf. [14] for more details). For a fixed $\lambda > 0$, let $f \in X_0$ and let $f = x_{0,\lambda} + x_{1,\lambda}$ be an optimal decomposition for the computation of

$$K(\lambda, f; X_0, X_1) = \inf_{f=x_0+x_1}\{\|x_0\|_{X_0} + \lambda\|x_1\|_{X_1}\}.$$

We with this decomposition at hand we continue the spliting by solving

$$K(2\lambda, x_{1,\lambda}; X_0, X_1) = \|x_{0,2\lambda}\|_{X_0} + \lambda\|x_{1,2\lambda}\|_{X_1}, x_{1,\lambda} = x_{0,2\lambda} + x_{1,2\lambda},\ x_{i,2\lambda} \in X_i, i = 0, 1.$$

Continuing in this fashion we obtain a sequence $x_{1,2^{k-1}\lambda} = x_{0,2^k\lambda} + x_{1,2^k\lambda}, x_{i,2^k\lambda} \in X_i, i = 0, 1, k = 1, ...$, such that

$$K(2^k\lambda, x_{1,2^{k-1}\lambda}; X_0, X_1) = \|x_{0,2^k\lambda}\|_{X_0} + \lambda 2^k\|x_{1,2^k\lambda}\|_{X_1}, \tag{5.5}$$

For future reference we also note that form the definitions it follows that

$$K(2^k\lambda, x_{1,2^{k-1}\lambda}; X_0, X_1) \leq 2^k\lambda\|x_{1,2^{k-1}\lambda}\|_{X_1}, k = 1, ... \tag{5.6}$$

Observe that

$$x_{0,\lambda} + x_{1,\lambda} = x_{0,\lambda} + x_{0,2\lambda} + x_{1,2\lambda}$$

in other words

$$x_{1,\lambda} - x_{1,2\lambda} = x_{0,2\lambda}.$$

More generally we have

$$x_{1,2^{k-1}\lambda} - x_{1,2^k\lambda} = x_{0,2^k\lambda}, k = 1, ... \tag{5.7}$$

Therefore following the usual construction of the fundamental lemma we are led to define

$$u_k = x_{1,2^{k-1}\lambda} - x_{1,2^k\lambda} \in X_0, k = 1, ...$$

Moreover, since $X_0 \subset X_1$, we complete the sequence by setting

$$u_0 = x_{0,\lambda}.$$

Then we have

$$\sum_{k=0}^{N} x_{0,2^k\lambda} = \sum_{k=0}^{N} u_k = x_{0,\lambda} + \sum_{k=1}^{N}(x_{1,2^{k-1}\lambda} - x_{1,2^k\lambda}) = x_{0,\lambda} + x_{1,\lambda} - x_{1,2^N\lambda}. \quad (5.8)$$

Therefore,

$$f - \sum_{k=0}^{N} u_k = x_{1,2^N\lambda}, \quad (5.9)$$

Therefore the convergence of the decomposition

$$f \sim \sum_{k=0}^{\infty} u_k, \quad (5.10)$$

hinges on the study of the convergence of $x_{1,2^N\lambda X_1}$ to zero in suitable norms. We now study the convergence in the sum space (in our case X_1),

$$\left\| f - \sum_{k=0}^{N} u_k \right\|_{X_1} = \left\| x_{1,2^N\lambda} \right\|_{X_1}.$$

We collect a number of inequalities that will be useful in our analysis

Lemma 29 *(i) Let $f \in X_1$, then the series $\sum_{k=0}^{\infty} \frac{\|x_{0,2^k\lambda}\|_{X_0}}{2^k\lambda}$ is convergent and in fact*

$$\sum_{k=0}^{N} \frac{\left\| x_{0,2^k\lambda} \right\|_{X_0}}{2^k\lambda} \leq \|f\|_{X_1}, N = 1..... \quad (5.11)$$

(ii) For all $k = 0, 1,$

$$\left\| x_{1,2^k\lambda} \right\|_{X_1} \leq C \|f\|_{X_1}. \quad (5.12)$$

(iii) If $f \in X_0$, then

$$\left\| x_{1,2^k\lambda} \right\|_{X_0} \leq C 2^k \lambda \|f\|_{X_0}. \quad (5.13)$$

(iv) If $f \in X_{\theta,q}$, then

$$\left\| x_{1,2^k\lambda} \right\|_{X_{\theta,q}} \leq C 2^k \lambda \|f\|_{X_{\theta,q}}. \quad (5.14)$$

(v) [Holmstedt's formula]

$$K(t, f, X_{\theta,q}, X_1) \approx \left\{ \int_0^{t^{1/1-\theta}} (K(s, f; X_0, X_1) s^{-\theta})^q \frac{ds}{s} \right\}^{1/q}.$$

In particular,

$$t^\theta K(t^{1-\theta}, f; X_{\theta,q}, X_1) \succeq K(t, f; X_0, X_1) \quad (5.15)$$

Proof. (i) Combining (5.5) and (5.6) we see that

$$\left\|x_{0,2^k\lambda}\right\|_{X_0} + \lambda 2^k \left\|x_{1,2^k\lambda}\right\|_{X_1} \leq 2^k\lambda \left\|x_{1,2^{k-1}\lambda}\right\|_{X_1}. \tag{5.16}$$

Therefore,

$$\begin{aligned}
\sum_{k=0}^{N} \frac{\left\|x_{0,2^k\lambda}\right\|_{X_0}}{2^k\lambda} &\leq \left\|x_{0,\lambda}\right\|_{X_0} + \sum_{k=1}^{N-1}\left(\left\|x_{1,2^{k-1}\lambda}\right\|_{X_1} - \left\|x_{1,2^k\lambda}\right\|_{X_1}\right) \\
&= \frac{\left\|x_{0,\lambda}\right\|_{X_0}}{\lambda} + \left\|x_{1,\lambda}\right\|_{X_1} - \left\|x_{1,2^{N-1}\lambda}\right\|_{X_1} \\
&= \frac{K(\lambda,f)}{\lambda} - \left\|x_{1,2^{N-1}\lambda}\right\|_{X_1} \\
&\leq \|f\|_{X_1}.
\end{aligned}$$

(ii) To prove (5.12) we proceed by induction. The case $k=0$ is trivial since we have

$$\|x_{1,\lambda}\|_{X_1} \leq \frac{K(\lambda,f)}{\lambda} \leq \|f\|_{X_1}. \tag{5.17}$$

Using (5.17) we find

$$2\lambda \left\|x_{1,2\lambda}\right\|_{X_1} \leq K(2\lambda, x_{1,\lambda}) \leq 2\lambda \left\|x_{1,\lambda}\right\|_{X_1} \leq 2K(\lambda,f).$$

Thus,

$$2\lambda \left\|x_{1,2\lambda}\right\|_{X_1} \leq 2K(\lambda,f),$$

proving the case $k=1$. The argument that $k-1$ implies k is the same for arbitrary k.

(iii) When $f \in X_0$ then from the definitions it follows that $x_{1,2^k\lambda} \in X_0$, then from (5.9) and (5.11) we see that

$$\begin{aligned}
\left\|x_{1,2^k\lambda}\right\|_{X_0} &\leq \|f\|_{X_0} + \left\|\sum_{j=0}^{k} x_{0,2^j\lambda}\right\|_{X_0} \\
&\leq \|f\|_{X_0} + 2^k\lambda \sum_{j=0}^{k} \frac{\left\|x_{0,2^j\lambda}\right\|_{X_0}}{2^j\lambda} \\
&\leq \|f\|_{X_0} + 2^k\lambda \|f\|_{X_1} \\
&\leq C2^k\lambda \|f\|_{X_0}.
\end{aligned}$$

(iv) Follows *mutatis mutandi* the proof of (iii).

(v) For Holmstedt's formula see [1]. The inequality (5.15) follows readily by monotonicity. ■

Proposition 30 *Suppose that* $f \in X_0$, *then* $\left\|f - \sum_{k=0}^{2N} u_k\right\|_{X_1} \to 0.$

Proof. Repeat (5.9) at level $2N$, then

$$x_{1,2^{2N}\lambda} + \sum_{k=0}^{2N} u_k = x_{1,2^N\lambda} + \sum_{k=0}^{N} u_k,$$

and therefore we find that

$$x_{1,2^{2N}\lambda} = -\sum_{k=N+1}^{2N} u_k + x_{1,2^N\lambda}. \tag{5.18}$$

Now from (5.9) and (5.18) write

$$\begin{aligned} f - \sum_{k=0}^{2N} u_k &= x_{1,2^{2N}\lambda} \\ &= -\sum_{k=N+1}^{2N} u_k + x_{1,2^N\lambda} \end{aligned}$$

Also from (5.5) we have

$$\left\|x_{0,2^{2N}\lambda}\right\|_{X_0} + 2^{2N}\left\|x_{1,2^{2N}\lambda}\right\|_{X_1} = K(2^{2N}\lambda, x_{1,2^{2N-1}\lambda}).$$

Therefore since

$$2^{2N}\left\|x_{1,2^{2N}\lambda}\right\|_{X_1} \leq K(2^{2N}\lambda, x_{1,2^{2N-1}\lambda}) \tag{5.19}$$

we can continue with

$$\begin{aligned} K(2^{2N}\lambda, x_{1,2^{2N-1}\lambda}) &= K(2^{2N}\lambda, -\sum_{k=N+1}^{2N-1} u_k + x_{1,2^N\lambda}) \qquad (5.20) \\ &\leq \left\| -\sum_{k=N+1}^{2N-1} u_k + x_{1,2^N\lambda}\right\|_{X_0} \\ &\leq \left\|\sum_{k=N+1}^{2N-1} u_k\right\|_{X_0} + \left\|x_{1,2^N\lambda}\right\|_{X_0} \\ &\leq \sum_{k=N+1}^{2N-1} \|u_k\|_{X_0} + \left\|x_{1,2^N\lambda}\right\|_{X_0}. \end{aligned}$$

Putting together these estimates we have

$$\begin{aligned} \left\|x_{1,2^{2N}\lambda}\right\|_{X_1} &\leq \frac{\sum_{k=N+1}^{2N-1} \|u_k\|_{X_0}}{2^{2N}} + \frac{\left\|x_{1,2^N\lambda}\right\|_{X_0}}{2^{2N}} \\ &= I + II. \end{aligned}$$

Now by Lemma 29 (iii) we have

$$II \leq \frac{C2^N\lambda \|f\|_{X_0}}{2^{2N}} \to 0, \text{ as } N \to \infty.$$

while on account of (5.11) we also have

$$\frac{1}{2^{2N}\lambda}\sum_{k=N+1}^{2N-1}\|u_k\|_{X_0} \leq \sum_{k=N+1}^{2N-1}\frac{\|u_k\|_{X_0}}{2^k\lambda} \to 0 \text{ as } N \to \infty.$$

This concludes the proof. ■

We now consider convergence in the case when $f \in X_{\theta,q}$. It will be convenient to collect the corresponding estimates in a Lemma

Proposition 31 *Let $0 < \theta < 1$, and let $f \in X_{\theta,q}$. Then*

$$\left\| f - \sum_{k=0}^{N} u_k \right\|_{X_{\theta,q}} \to 0.$$

Proof. Assume first that $\theta < 1/2$, we shall indicate later the trivial modifications to deal with the general case. We start once again with the representation

$$x_{1,2^{2N-1}\lambda} = -\sum_{k=N}^{2N-1} u_k + x_{1,2^N\lambda}$$

Then, proceeding as in the proof of Proposition 30 (cf. (5.20)), we have

$$\begin{aligned}
\|x_{1,2^{2N-1}\lambda}\|_{X_1} &\preceq \frac{K(2^{2N}\lambda, -\sum_{k=N}^{2N-1} u_k + x_{1,2^N\lambda})}{2^{2N}\lambda} \\
&\leq \frac{K(2^{2N}\lambda, -\sum_{k=N}^{2N-1} u_k)}{2^{2N}\lambda} + \frac{K(2^{2N}\lambda, x_{1,2^N\lambda})}{2^{2N}\lambda} \\
&= I + II.
\end{aligned}$$

As before on account of Lemma 29 (i)

$$\begin{aligned}
I &\leq \frac{1}{2^{2N}}\left\|\sum_{k=N}^{2N-1} u_k\right\|_{X_0} \\
&\leq \sum_{k=N}^{2N-1}\frac{\|u_k\|_{X_0}}{2^{2N}} \to 0.
\end{aligned}$$

On the other hand by Lemma 29 (v)

$$\begin{aligned}
II &\leq 2^{2N\theta}\lambda^{\theta}\frac{K(2^{2N(1-\theta)}\lambda, x_{1,2^N\lambda}, X_{\theta,q}, X_1)}{2^{2N}\lambda} \\
&\leq 2^{2N\theta-2N}\lambda^{\theta-1}\|x_{1,2^N\lambda}\|_{X_{\theta,q}} \\
&\leq 2^{2N\theta-2N}\lambda^{\theta-1}2^N \ \|f\|_{X_{\theta,q}} \text{ (by Lemma 29 (iv))}.
\end{aligned}$$

The right hand side goes to zero since $2N\theta - 2N + N < 0$.

To deal with the case $\theta \geq 1/2$ we simply apply the same argument writing $x_{1,2^{2N-1}\lambda} = -\sum_{k=N\alpha}^{2N-1} u_k + x_{1,2^{N\alpha}\lambda}$, where α is chosen so that $2N\alpha\theta - 2N\alpha + N\alpha < 0$. ■