A Tight Frame Algorithm in Image Inpainting

CHENG, Kei Tsi Daniel

A Thesis Submitted in Partial Fulfilment

of the Requirements for the Degree of

Master of Philosophy

in

Mathematics

July 2007

Thesis/ Assessment Committee

Professor Ng Kung Fu (Chair)
Professor Chan Hon Fu Raymond (Thesis Supervisor)
Professor Feng Dejun (Committee Member)
Professor Ng Kwok Po Michael (External Examiner)

Abstract of thesis entitled:

A Tight Frame Algorithm in Image Inpainting

Submitted by CHENG, Kei Tsi Daniel

for the degree of Master of Philosophy in Mathematics

at The Chinese University of Hong Kong in July 2007

In this thesis, we work on image inpainting. In particular, we propose a new iteration algorithm based on framelet systems, a generalization of wavelets. Hence the regularity of the restored image is guaranteed. By comparing the differences and similarities of existing algorithms as constrained minimization problems, we obtain a new iteration algorithm and prove its convergence in the context of convex analysis and optimization theory. Recently a method has been proposed in solving the total variation minimization problem. We also include a similar method as part of our convergence proof. We find that the convergence limit satisfies some regularization properties. The minimization problem imposes a sparsity on the canonical framelet coefficients of the underlying solution. This would also restricts the Besov norm of the underlying solution, and hence the rough-

of the solution is under control. Finally, we compare the new algorithm with existing ones in numerical experiments.

摘要

香港中文大學碩士論文摘要
論文提目:

圖像修補中的 Tight Frame 算法

鄭基賜
二零零七年七月

我們在這份論文研究圖像修補。根據推廣了小波的 framelet 系統，我們提出了一種新疊代算法。因此保證了被恢復的圖像的規律性。我們透過將現有的算法考慮成拘束的最小化問題，比較他們的相似性和區別。我們從這角度獲得一種新算法，並用凸分析和最優化理論證明它的收斂。最近有人提出了一個解決總變差最小化問題的算法。我們將一個類似的算法包括在我們的算法和收斂證明的裏面。我們發現得出來的解滿足一些正規化的特性。此最小化問題令標準 framelet 系數變得稀疏，這並且會限制解答的 Besov 範數，而解答的粗糙度會受到控制。最後，我們將新算法與現有的算法進行比較。

Acknowledgement

First of all, I wish to express my deepest gratitude to my supervisor, Professor Raymond H. Chan. I would like to thank for his invaluable advice. His continual guidance and emphasis on elegance in research have helped me a lot in preparing this thesis.

Next I would like thank Prof. M. Dai, Prof. Z. W. Shen, Prof. W. K. Wong and Prof. Q. Zhang for their suggestions. My thanks are also due to my colleagues and friends Dr. J. F. Cai, Mr. H. F. Chan, Ms. Y.Q. Dong, Mr. W. Wang, Mr. Z. X. Wang, Mr. C. Y. Wong and Mr Y. S. Wong for their many helpful discussions.

Finally, I would like to thank my family for their endless encouragement and support. I am particularly grateful to my parents, who enable me to devote all my energy to research.

To My Parents

Contents

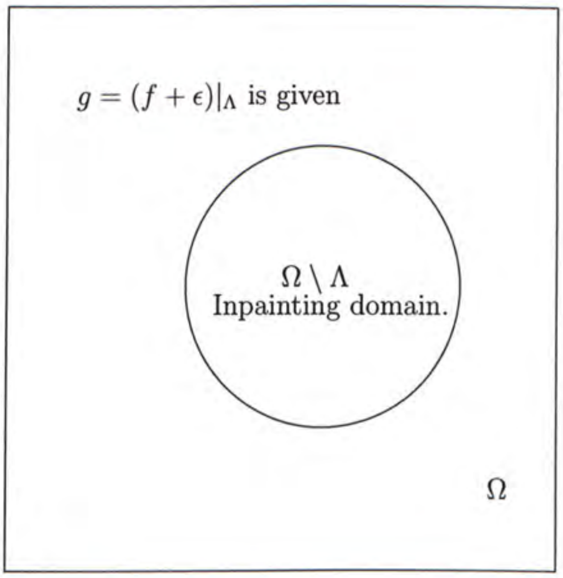
Chapter 1

Introduction

The problem of inpainting [3] occurs when part of the pixel data in a picture is lost or overwritten by other means. This problem arises from mimicking the restoration of ancient arts or drawings, where a portion of the picture is missing or damaged due to aging, scratch, etc. Inpainting problem later found its applications. For examples, we can apply inpainting algorithms in automatic scratch removal in digital photos and old films [3, 11]; in creating special effects such as disocclusion [30], object removal [3, 10], zooming and super-resolution[2, 11, 41]; or in the case when an image is transmitted through a noisy passageway. On the other hand, inpainting has been closely related to earlier works in engineering literature such as image interpolation [1, 26, 27] and error concealment [12, 25, 28] in communication technology.

The task of inpainting is to recover the missing region from the incomplete data observed. The basic idea is to fill in missing region with data available from the surrounding. Ideally, an image is expected to possess shapes and patterns consistent to the given data in human vision. That is, we should extract information such as edges and texture from the observed data and "put" these information into the corrupted part in a way that it would look

Figure 1.1: The image data is loss in the inpainting domain $\Omega \setminus \Lambda$, while the observed portion is often noisy.



like a natural extension from the view of human. In most cases, the available data of the original image is noisy and thus it is necessary to eliminate the noise while filling in the corrupted part.

The mathematical model of image inpainting can be formulated as follows. Denotes $\Omega = [0,1]^2$ to be the image domain in $\mathbb{R}^2$, x to be a general point in Ω and $\mathbf{f}(x)$ to be the original image that we are interested. Part of the image in the domain is corrupted and only those inside a subset $\Lambda \subsetneq \Omega$ is observable. The observed region Λ is non-empty. The observed image $\mathbf{g}$ would then be

$$\mathbf{g}(x) = \begin{cases} \mathbf{f}(x) + \epsilon(x), & x \in \Lambda, \\ \text{arbitary}, & \text{otherwise.} \end{cases} \tag{1.1}$$

where ϵ is the noise component. The aim of image inpainting is to recover $\mathbf{f}$

from $\mathbf{g}$. Throughout this thesis, we assume ϵ to be additive white Gaussian noise.

In parallel to its wide applications, there are various methods existing for inpainting and related problems, including the non-linear filtering method, the variational method, the Bayesian method, the learning-and-growing method, and the PDE method. The Mumford-Shah-Euler image model [20] we are going to present belongs to the PDE method. Similar to most variational inpainting models, given an observed image $\mathbf{g}$, the model tries to minimize the functional

$$J(\mathbf{f}) = \frac{\mu}{2} \int_{\Lambda} (\mathbf{f} - \mathbf{g})^2 + E(\mathbf{f}), \tag{1.2}$$

where $E(\mathbf{f})$ encodes the image model. For example, $E(\mathbf{f})$ can be $\int_{\Omega} \Delta \mathbf{f}^2$ or the total variation norm $\int_{\Omega} |\nabla \mathbf{f}|$ in Rudin, Osher and Fatemi's model [37, 38]. Chan and Shen [11] and Tsai et al. [41] used the one that Mumford and Shah proposed for image segmentation [33]:

$$E(\mathbf{f}, \Gamma) = \frac{\gamma}{2} \int_{\Omega \backslash \Gamma} |\nabla \mathbf{f}|^2 dx + \alpha \cdot \text{length}(\Gamma), \tag{1.3}$$

where Γ denotes the edge collection, and length(Γ) measures the length of the edges in Γ.

There are some deficiencies when applying Mumford and Shah's object-edge image model in inpainting. These shortcomings are first discovered by Chan and Shen [11]. To start with, one can see that penalizing the length of the edges of objects inside the image favors straight edges in the inpainting domain. The reason is that straight lines have the shortest length. As a result, the image model simply completes the missing segments by straight lines. These inpainted edges thus join the existing ones in a non-tangent manner that produce man-made corners. Minimizing the length of edges inside the inpainting domain may also violate the Connectivity Principle. The functional encourages connection between pairs of edge segments in which the intersections with the inpainting region are close to each other. If the size of the inpainting domain exceeds that of the object, long distance connection would cost more in the functional, and the restored image would have disconnected parts. This may contradict to the fact that many real images have structures like thin stripes or fiber-like textures.

To overcome the above artifacts, Chan, Kang and Shen [9] proposed using Euler's elastic curve model. An additional term $\beta \int_\Gamma \kappa^2 ds$ is introduced to the image model. The resulting Mumford-Shah-Euler inpainting model becomes

$$J(\mathbf{f},\Gamma) = \frac{\mu}{2}\int_\Lambda (\mathbf{f}-\mathbf{g})^2 dx + \frac{\gamma}{2}\int_{\Omega\setminus\Gamma} |\nabla u|^2 dx + \alpha \text{length}(\Gamma) + \beta \int_\Gamma \kappa^2 ds.$$

If we define $\mu_\Lambda = \mu \cdot \mathbf{1}_\Lambda(x)$, we have

$$J(\mathbf{f},\Gamma) = \frac{1}{2}\int_\Omega \mu_\Lambda (\mathbf{f}-\mathbf{g})^2 dx + \frac{\gamma}{2}\int_{\Omega\setminus\Gamma} |\nabla u|^2 dx + \int_\Gamma (\alpha + \beta\kappa^2) ds.$$

The solution from this model is found by the Euler-Lagrange equation and used the level set method of Osher and Sethian [35].

In this thesis, we derive algorithms based on tight framelet system. This system, as well as other redundant systems, has already been used in many applications. Perturbation of the framelet coefficients and built-in regularization of algorithms allow information to transmit from Λ to $\Omega \setminus \Lambda$ smoothly. It is known that by the redundancy in the system, numerical error in computing coefficients is reduced. The projection of framelet coefficients to the space of some canonical coefficients can eliminate the error component in the kernel of reconstruction operator. In [7], an algorithm based on tight framelet is proposed. The algorithm is shown to improve the PSNR by 2 to 3 dB when compared with variational approaches such as those given in [10, 11]. The advantage of the redundancy is also mentioned in [22] for its robust signal representation.

Many images such as carton images can be modeled as piecewise smooth functions. The framelet system can have a good approximation to these functions. Additionally, the weighted ℓ_1 minimization term in the coefficient domain allows sparse representations of images. With this important property, we can efficiently solve the problem in the framelet domain. Some inpainting algorithms related to ℓ_1 minimization are proposed in [7, 15, 19, 21, 39].

In this thesis, we consider some of these algorithms as constrained minimization problems, and derive new minimization problems from them. Using Combettes and Wajs's framework of proximal forward-backward splitting [14], and applying Chambolle's technique used in total variational minimization [6], we propose a new iteration algorithm for inpainting problems. Besides, we will see that the minimization problem is related to the Besov norm of its solution.

The organization of this thesis is as follows. In Chapter 2, we will review some preliminary knowledge that is useful later in the thesis. We will then consider some inpainting algorithms as the minimizers of functionals given in Chapter 3. While we can see their similarities from the minimization point of view, we also present some new minimization functionals induced from them. In the subsequent Chapters 4 and 5, we will try to find new algorithms that converge to the minimizers of these new functionals. The convergence is then discussed in Chapter 6. The numerical results are given in Chapter 7.

Chapter 2

Background Knowledge

2.1 Image Restoration using Total Variation Norm

In this section, we consider (1.2) for the case where $E(\mathbf{f})$ is the total variation norm, and the Chambolle's method [6] of solving it. We consider the method here because our approach in solving the inpainting problem will be similar to this method. Since in his paper, the method is proposed to solve image denoising and zooming, we assume $\Lambda = \Omega$ in this section for illustrating his idea.

Many choices of image model have been proposed for solving denoising problems. In 1977, Tikhonov and Arsenin [40] proposed $E(\mathbf{f})$ to be $\gamma \int_\Omega |\nabla \mathbf{f}|^2$. A very strong smoothing Laplacian operator appears in the Euler-Langrange equation

$$\mathbf{0} = \mu(\mathbf{f} - \mathbf{g}) - \Delta \mathbf{f}.$$

The L^2 norm of the gradient in the functional is very strong in denoising but does not allow sharp edges in the solution.

In [37, 38], Rudin, Osher and Fatemi proposed using L^1 norm of the gradient, also called the total variation norm, instead. After this change, (1.2) becomes

$$J(\mathbf{f}) = \frac{\mu}{2}\int_{\Omega}(\mathbf{f}-\mathbf{g})^2dx + \int_{\Omega}|\nabla\mathbf{f}|dx. \tag{2.1}$$

Multifarious approaches in solving it can be found in the literature. In the following, we are going to present some of them.

The first comes from Rudin, Osher and Fatemi [38] themselves. They made use of the Euler-Langrange equation, assuming homogeneous Neumann boundary conditions

$$\mathbf{0} = \mu(\mathbf{f}-\mathbf{g}) - \nabla\cdot\left(\frac{\nabla\mathbf{f}}{|\nabla\mathbf{f}|}\right) \tag{2.2}$$

and a time marching scheme to reach the steady state of the following parabolic equation:

$$\frac{\partial\mathbf{f}}{\partial t}(x,t) = \nabla\cdot\left(\frac{\nabla\mathbf{f}(x,t)}{|\nabla\mathbf{f}(x,t)|}\right) - \mu(x,t)(\mathbf{f}(x,t)-\mathbf{g}(x,t)), \tag{2.3}$$

with initial condition

$$\mathbf{f}(x,0) = \mathbf{g}(x,0), x\in\Omega.$$

An explicit discretization scheme is used to obtain a numerical method so that $\mathbf{f}(x,t)$ would approach the solution of the Euler-Language equation when t goes to infinity. In order to have stability, the time step should be small: $\Delta t = O(\Delta x^2)$. Equation (2.2) is degenerate due to the term $1/|\nabla\mathbf{f}|$. To overcome this difficulty, people use $\phi(|\nabla\mathbf{f}|)$ with suitable properties instead of $|\nabla\mathbf{f}|$. A common proxy is $\sqrt{|\nabla\mathbf{f}|^2+\beta}$ for $\beta > 0$, the corresponding time marching scheme is

$$\frac{\partial\mathbf{f}}{\partial t}(x,t) = \nabla\cdot\left(\frac{\nabla\mathbf{f}(x,t)}{\sqrt{|\nabla\mathbf{f}(x,t)|^2+\beta}}\right) - \mu(x,t)(\mathbf{f}(x,t)-\mathbf{g}(x,t)), \tag{2.4}$$

with initial condition

$$\mathbf{f}(x,0) = \mathbf{g}(x,0), x\in\Omega.$$

Chan, Golub and Mulet [8] introduced a new algorithm to remove the highly non-linear and non-differentiable term $\nabla\mathbf{f}/|\nabla\mathbf{f}|$ in (2.2). Their idea is to introduce a new variable $p = \nabla\mathbf{f}/|\nabla\mathbf{f}|$, and the new method is called the primal-dual method. Contrast to the simple look, the term has got rid of the singularity in the equation and provides a good global convergence property. With $p = \nabla\mathbf{f}/|\nabla\mathbf{f}|$ in mind, we can replace (2.2) by the following equivalent system of linear equations:

$$\begin{cases} \mathbf{0} = |\nabla\mathbf{f}|p - \nabla\mathbf{f} \equiv F(p, \mathbf{f}) \\ \mathbf{0} = -\nabla \cdot p + \mu(\mathbf{f} - \mathbf{g}) \equiv G(p, \mathbf{f}). \end{cases} \tag{2.5}$$

The authors of [8] suggested using Newton method to solve the system. Thus we obtain the linearization of the above system:

$$\begin{bmatrix} |\nabla\mathbf{f}| & -(I - \frac{p\nabla\mathbf{f}^T}{|\nabla\mathbf{f}|})\nabla \\ -\nabla\cdot & \mu I \end{bmatrix} \begin{bmatrix} \delta p \\ \delta\mathbf{f} \end{bmatrix} = - \begin{bmatrix} F(p, \mathbf{f}) \\ G(p, \mathbf{f}) \end{bmatrix}. \tag{2.6}$$

To solve (2.6), first eliminate δp and then solve the resulting equation:

$$\left[-\nabla \cdot \left(\frac{1}{|\nabla\mathbf{f}|}(I - \frac{p\nabla\mathbf{f}^T}{|\nabla\mathbf{f}|})\nabla\right) + \mu I\right] \delta\mathbf{f} = -G(\mathbf{f}),$$

and δp is then computed by

$$\delta p = \frac{1}{|\nabla\mathbf{f}|}(I - \frac{p\nabla\mathbf{f}^T}{|\nabla\mathbf{f}|})\nabla\delta\mathbf{f} - p + \frac{\nabla\mathbf{f}}{|\nabla\mathbf{f}|}.$$

The last method here comes from the dual formulation. By introducing the dual variable p, and using integral by parts, minimizing (2.1) can be rewritten as

$$\inf_{\mathbf{f}} \sup_{|p|\leq 1} \left\{ \frac{\mu}{2} \int_\Omega (\mathbf{f} - \mathbf{g})^2 dx + \int_\Omega \mathbf{f}\nabla \cdot p dx \mid p \in C_C^1(\Omega) \right\}. \tag{2.7}$$

Since the function is strictly convex in $\mathbf{f}$ and concave (linear) in p, we can swap inf and sup, and obtain

$$\sup_{|p|\leq 1} \inf_{\mathbf{f}} \left\{ \frac{\mu}{2} \int_\Omega (\mathbf{f} - \mathbf{g})^2 dx + \int_\Omega \mathbf{f}\nabla \cdot p dx \mid p \in C_C^1(\Omega) \right\}. \tag{2.8}$$

Fix p, we can find the solution to $\inf_{\mathbf{f}}$ by Euler-Langrange equation. The solution is given by: $\mathbf{f} = \mathbf{g} - (\nabla \cdot p)/\mu$.

Substitute back into (2.8), and use the identity $\sup f(x) \equiv -\inf(-f(x))$ for a general function $f(x)$, we arrive at the dual problem:

$$\inf_{|p|\leq 1} \left\{ \int_{\Omega} (\nabla \cdot p - \mu \mathbf{g})^2 dx \mid p \in C_C^1(\Omega) \right\}. \tag{2.9}$$

In discrete case, the constraint $|p_i| \leq 1$ yield the Lagrange multiplier $\alpha_i \geq 0$ under the Karush-Kuhn-Tucker conditions, so that for each i,

$$-(\nabla(\nabla \cdot p - \mu \mathbf{g}))_i + \alpha_i p_i = 0, \tag{2.10}$$

with complementarity conditions:

$$(i) \quad \alpha_i > 0 \text{ and } |p_i| = 1; \text{ or}$$

$$(ii) \quad \alpha_i = 0 \text{ and } |p_i| < 1.$$

Note that after solving for p, we can recover $\mathbf{f}$: $\mathbf{f} = \mathbf{g} - (\nabla \cdot p)/\mu$.

Chambolle made an observation in [6] that in the complementarity conditions above, $\alpha_i = |(\nabla(\nabla \cdot p - \mu \mathbf{g}))_i|$ for both cases. To see that, since by (2.10), we have $\alpha_i p_i = (\nabla(\nabla \cdot p - \mu \mathbf{g}))_i$, by taking absolute value on both sides, $\alpha_i |p_i| = |(\nabla(\nabla \cdot p - \mu \mathbf{g}))_i|$. In the first case, the result follows from substituting $|p_i| = 1$, while we can see that $|(\nabla(\nabla \cdot p - \mu \mathbf{g}))_i|$ is zero in the second case. Hence the Lagrange multiplier α is eliminated and we obtain the following simplified equation:

$$-(\nabla(\nabla \cdot p - \mu \mathbf{g}))_i + |(\nabla(\nabla \cdot p - \mu \mathbf{g}))_i| p_i = 0. \tag{2.11}$$

Chambolle [6] thus proposed a semi-implicit gradient descent algorithm. First set $\tau > 0$ and $p^0 = 0$, then for any $n \geq 0$,

$$p_i^{n+1} = p_i^n - \tau(-(\nabla(\nabla \cdot p^n - \mu \mathbf{g}))_i + |(\nabla(\nabla \cdot p^n - \mu \mathbf{g}))_i| p_i^{n+1}). \tag{2.12}$$

Reducing to the explicit scheme:

$$p_i^{n+1} = \frac{p_i^n + \tau(\nabla(\nabla \cdot p^n - \mu \mathbf{g}))_i}{1 + \tau|(\nabla(\nabla \cdot p^n - \mu \mathbf{g}))_i|}. \tag{2.13}$$

We now state the main theorem in [6]:

Theorem 1 *Let* $\tau \leq 1/8$. *Then* $(\nabla \cdot p^n)/\mu$ *converges to* $\mathbf{g} - \mathbf{f}$ *as* $n \to \infty$.

2.2 An Example of Tight Frame system

In this section, we give examples of tight frame filters using the unitary extension principle given in [36]. To know what tight frame is, see Section 3.1.

We first consider the univariate case, i.e., $\mathbf{f}$ is 1D signal. Let $\hat{h}_0$ be the trigonometric polynomial $\cos^{2m}(\omega/2)$. It is the refinement symbol of the B-spline

$$\hat{\varphi}(\omega) = \frac{\sin^{2m}(\omega/2)}{(\omega/2)^{2m}}$$

of order $2m$. The corresponding wavelets are given by $\hat{\psi}_n(\omega) = \hat{h}_n(\omega/2)\hat{\varphi}(\omega/2)$ with wavelet masks

$$\hat{h}_n(\omega) = \sqrt{\begin{pmatrix} 2m \\ n \end{pmatrix}} \sin^n(\omega/2)\cos^{2m-n}(\omega/2),$$

for $1 \leq n \leq 2m$. It is shown in [36] that the system

$$\mathcal{X} = \{2^{k/2}\psi_n(2^k \cdot -j) : k, j \in \mathbb{Z}; n = 1, \cdots, 2m\}$$

is a tight frame system.

For example, if $m = 1$, we call it the piecewise linear tight frame systems. The corresponding filters are

$$h_0 = \frac{1}{4}[1, 2, 1], h_1 = \frac{\sqrt{2}}{4}[1, 0, -1], h_2 = \frac{1}{4}[-1, 2, -1].$$

If $m = 2$, it is the piecewise cubic tight frame systems. The corresponding filters are

$$h_0 = \frac{1}{16}[1,4,6,4,1], h_1 = \frac{1}{8}[1,2,0,-2,-1], h_2 = \frac{\sqrt{6}}{16}[-1,0,2,0,-1],$$

$$h_3 = \frac{1}{8}[-1,2,0,-2,1], h_4 = \frac{1}{16}[1,-4,6,-4,1].$$

From now on, for the ease of computation, we use the discrete version. In the discretized setting, Ω is divided into $l \times l$ square pixels. The digital data, also denoted by $\mathbf{f}$, of an image consists of $N = l^2$ pixels. Now, $\mathbf{f}$ is considered as a vector in $\mathbb{R}^N$ by concatenating the image into a column vector. The data value of a pixel is produced by averaging the function value inside that pixel.

In the following, we give the tight frame decomposition operator A. To apply the filters to finite sequence of discrete signals, boundary conditions must be amalgamated. We use the Neumann boundary conditions here, and we will get Toeplitz-plus-Hankel matrices. Using the Neumann boundary conditions have an advantage that usually the restored images have less artifacts near the boundaries, see [5, 7, 34]. Let the length of the signal be N, for any filter $h = \{h(j)\}_{j=-m}^{m}$, we define an N-by-N matrix $D(h)$ given by

$$D(h) = \begin{bmatrix} h(0) & \cdots & h(-m) & & 0 \\ \vdots & \ddots & \ddots & \ddots & \\ h(m) & \ddots & \ddots & \ddots & h(-m) \\ & \ddots & \ddots & \ddots & \vdots \\ 0 & & h(m) & \cdots & h(0) \end{bmatrix} + \left[\begin{array}{cccc|c} h(1) & h(2) & \cdots & h(m) & \\ h(2) & \cdots & \cdots & & \\ \vdots & \cdots & & & 0 \\ h(m) & & & & \\ \hline & & 0 & & 0 \end{array}\right]$$

$$+\left[\begin{array}{c|cccc} 0 & & & 0 & \\ \hline & & & & h(-m) \\ & & & \cdots & \vdots \\ 0 & & \cdots & \cdots & h(-2) \\ & h(-m) & \cdots & h(-2) & h(-1) \end{array}\right].$$

Note that the last two matrices correspond to implementing the Neumann boundary conditions to the filter h. Define $H_i \equiv D(h_i)$. The unitary principle implies that $\sum_{i=0}^{2m} H_i^* H_i = I$. Therefore, we have the simplest, single level, tight frame system $A = [H_0^*, H_1^*, \cdots, H_{2m}^*]^*$.

Next, we present the multi-level version tight frame system (framelet system). It corresponds to the framelet decomposition without down sampling. We will use this multi-level version in our numerical tests. For level l, the filter corresponding to h, which we call $h^{(l)}$, is given by

$$h^{(l)} = \{h(-m), \underbrace{0, \cdots, 0}_{2^{l-1}-1}, h(-m+1), 0, \cdots, \cdots, 0, h(m-1), \underbrace{0, \cdots, 0}_{2^{l-1}-1}, h(m)\}.$$

The $h^{(l)}$ is formed by adding $(2^{l-1}-1)$ zeros between every adjacent element in h. Let $H_i^{(l)} \equiv D(h_i^{(l)})$, and $\Pi_{l=a}^{L-1} H_0^{(L-l)} \equiv H_0^{(L-a)} \cdots H_0^{(1)}$. The multi-level version of the tight frame system of level L without down-sampling is

$$M = \begin{bmatrix} \Pi_{l=0}^{L-1} H_0^{(L-l)} \\ H_1^{(L)} \Pi_{l=1}^{L-1} H_0^{(L-l)} \\ \vdots \\ H_{2m}^{(L)} \Pi_{l=1}^{L-1} H_0^{(L-l)} \\ \vdots \\ \vdots \\ H_1^{(1)} \\ \vdots \\ H_{2m}^{(1)} \end{bmatrix}. \tag{2.14}$$

By the perfect decomposition and reconstruction formula, one has $M^*M = I$.

When $\mathbf{f}$ is $2D$, the tight frame systems are constructed using the tensor products of the filters in $1D$, and define in analogous to (2.14), see [16, 29].

2.3 Sparse and compressed representation

In this section, we present the relation between the minimization of ℓ_p-norm of the transform coefficients for $0 < p \leq 1$ and the sparsity of transform coefficients, see [18]. The approximation ability allows us to incorporate the minimization of ℓ_p-norm into denoising and inpainting problems.

As the modern civilization advance, we have to deal with ever-increasing amount of data. In the process of acquiring and exploiting data, people find that most of the acquired data is compressible, i.e., can be eliminated with almost no perceptual loss. In the wild success of compression that reduces the data sizes for sounds, images and specialized data, a very natural question raises: do we need to put so much effort to acquire all the data, in which most will be thrown away? Can we directly obtain the part that will be retained?

Suppose $\mathbf{x}$ is an unknown vector in $\mathbb{R}^N$ (e.g. digital image or signal). If $\mathbf{x}$ is known to be compressible by transform coding with a known transform, and can be reconstructed via some procedures, we can measure m general linear functionals of $\mathbf{x}$ and then reconstruct $\mathbf{x}$ from these measurements. The number of measurement m can be significantly smaller than the original size N. By directly acquiring just the essential information about the signal/images, we are not acquiring the part of data that would eventually be thrown away by lossy compression.

In the abstract setting, we introduce here the principle of transform sparsity, which is known to hold in many setting of signal and image processing. We suppose that the object of interest is a vector $\mathbf{x} \in \mathbb{R}^N$, which can be a technical data or image with N samples or pixels, and that there is a basis $(\psi_i : i = 1, \ldots, m)$. The transform coefficients $\theta_i = \langle \mathbf{x}, \psi_i \rangle$ are assumed to be sparse in some ℓ_p-norm, in the sense that, for some $R > 0$

$$||\theta||_p \equiv \left(\sum_i |\theta_i|^p \right)^{1/p} \leq R. \tag{2.15}$$

This constraint is obeyed on natural classes of signals and images. For example, bounded variation model is used in many image processing applications. The bounded variation views an image as a function $f(x,y)$ on the unit square $0 \leq x, y \leq 1$ which satisfies

$$\int_{0 \leq x,y \leq 1} |\nabla f| dxdy \leq R. \tag{2.16}$$

In discretized setting, the digital data of image consists of $N = l^2$ pixels produced by averaging over $1/l \times 1/l$ pixels.

The constraint (2.15) on transform coefficients above relates to its sparsity. In fact, if we keep the n largest coefficients of θ, setting the others to zero, and denote the result by θ_n, then

$$||\theta - \theta_n||_2 \leq C||\theta||_p(n+1)^{1/2-1/p}, \tag{2.17}$$

where C is a constant that depends only on p. This implies that we can keep "most" of the information by keeping only $n \sim \varepsilon^{2p/(p-2)}$ terms in θ to reconstruct an approximation of the original θ with error ε.

In the above relation, more sparsity is required as p decreases. So, the ℓ_p norm with small p appears naturally in measuring sparsity. An ℓ_p constraint using $p = 2$ requires no sparsity at all.

From the above, one may naturally turn the question to: if $\mathbf{x}$ is an arbitrary signal that its transform coefficients θ observe (2.15), is it possible that the number of measurement m is dramatically less than N, yet can give us a sensible reconstruction of the target $\mathbf{x}$? This is considered in [18]. We present some of their ideas here. Let

$$X = X_{p,N}(R) = \{\mathbf{x} : ||\theta(\mathbf{x})||_p \leq R\} \subseteq \mathbb{R}^N,$$

we are interested in looking for measurement $\Phi_m : X \to \mathbb{R}^m$ of m samples about $\mathbf{x}$, and an algorithm $\phi_m : \mathbb{R}^m \to \mathbb{R}^N$ that reconstructs an approximation of $\mathbf{x}$. Here the samples are independent of each other, i.e., each sample does not depend on previous samples of $\mathbf{x}$. The measurement takes the form

$$\Phi_m(\mathbf{x}) = (\langle \xi_1, \mathbf{x}\rangle, \ldots, \langle \xi_m, \mathbf{x}\rangle),$$

where the ξ_i are sampling kernels. As the samples are independent of each other, the kernels are non-adaptive. An example is the tight frame system presented in the last section.

We want to find an optimal procedure that minimize the ℓ_2 error of the reconstruction, so it gives rise to the following minimax error for measurement:

$$E_m(X) = \inf_{\Phi_m, \phi_m} \sup_{x \in X} ||x - \phi_m(\Phi_m(x))||_2. \tag{2.18}$$

A good news comes as one can obtain an approximation in similar quality as keeping the n largest coefficients by using $m \approx n\log(N)$ samples provided by Φ_n. Indeed, if (m, N_m) be a sequence with $m \to \infty$, $N_m \sim Am^\gamma$, $\gamma > 1, A > 0$, then for $0 < p \leq 1$, there is a constant $C_p > 0$ which depends on A and γ but not on m so that

$$E_m(X) = E_m(X_{p,N}(R)) \leq C_p R(m/\log(N_m))^{1/2-1/p}. \qquad (2.19)$$

Here we do not require information on the specific object. Even though the measurement is non-adaptive and much less than the original size N, it can do as good as knowing the n best transform coefficients. As a result, we can take the minimization of the ℓ_p-norm of the transform coefficients and yet obtain a good approximation of the original signals or images. We will use the ℓ_1 minimization to tackle the inpainting problem in coming chapters.

2.4 Existence of minimizer in convex analysis

In this section, we present a sufficient condition of the existence of a minimizer in a convex programming problem. To fix our notations, denote $\Gamma_0(H)$ the set of all lower semi-continuous convex functions from a finite dimensional Hilbert space H to $(-\infty, +\infty]$ that is not identically $+\infty$.

Theorem 2 *If a function $F(\mathbf{f})$ lies in $\Gamma_0(H)$, and that it is coercive, then $F(\mathbf{f})$ possess a minimizer.*

Proof: Since $F(\mathbf{f}) \in \Gamma_0(H)$, $\inf F < +\infty$. There is a real number $\lambda > \inf F$ such that the level set $\mathcal{S} = \{\mathbf{f} : F(\mathbf{f}) \leq \lambda\}$ is non-empty. Note that $\mathcal{S}$ is closed. If F is coercive, then there exists r so that for any $\mathbf{u}$ satisfying $r < ||\mathbf{u}||$, $F(\mathbf{u}) > \lambda$. In other words, $\mathcal{S}$ is bounded. The conclusion follows from the fact that a lower semi-continuous function on a closed and bounded non-empty set inside a finite dimensional Hilbert space admits a minimum point. □

Chapter 3

Tight Frame Based Minimization

In this chapter, we first consider some inpainting algorithms as minimization problems. From these problems, we induce other minimization problems and finally give out their similarities and differences.

3.1 Tight Frames

Let A be a K-by-N ($K \geq N$) matrix whose rows are vectors in $\mathbb{R}^N$. The system, denoted by A again, consisting of all the rows of A, is a tight frame for $\mathbb{R}^N$ if for arbitrary vector $\mathbf{x} \in \mathbb{R}^N$,

$$||\mathbf{x}||_2^2 = \sum_{\mathbf{y} \in A} |\langle \mathbf{x},\mathbf{y} \rangle|^2. \tag{3.1}$$

Note that (3.1) is equivalent to that the system A possess the perfect reconstruction property $\mathbf{x} = \sum_{\mathbf{y} \in A} \langle \mathbf{x},\mathbf{y} \rangle \mathbf{y}$. The representation is unique in the case of orthogonal bases in $\mathbb{R}^N$, but this is not true in general. The matrix A is called the analysis operator, whereas its adjoint A^* is called the synthesis operator. It is clear that the perfect reconstruction equation can be written in the form $\mathbf{x} = A^*A\mathbf{x}$. This implies A is a tight frame if and only if $A^*A = I$. Reader should be aware that in general, $AA^* \neq I$. Framelets come from multi-resolution analysis on tight frames. See [7, 16, 36] and also Section 2.2 for examples and construction of tight frames and framelets.

Partition the tight frame systems into blocks of matrices

$$A = \begin{bmatrix} A_0 \\ A_1 \\ \vdots \\ A_M \end{bmatrix}.$$

The sub-matrix A_0 represents low-pass filter, while the others represent high-pass filters. Here, from the perfect reconstruction equation, we have

$$\sum_{i=0}^{M} A_i^* A_i = A^*A = I. \tag{3.2}$$

Finally, we note that from (3.1), $||\mathbf{v}||_2 = ||A\mathbf{v}||_2$ for all $\mathbf{v} \in \mathbb{R}^N$.

3.2 Minimization Problems and Algorithms

To have a sparse decomposition for images, we seek to minimize the ℓ_0 norm of wavelet coefficients, which is seemingly intractable. So the ℓ_0 norm is replaced by ℓ_1 norm, and a minimization of $||D\alpha||_1$, where α is the vector containing the wavelet coefficients and D is a diagonal scaling matrix, is sought. Let P_Λ be the diagonal matrix with diagonal entries equal 1 for indices belongs to Λ and 0 otherwise. When there is no noise, this minimization is subject to a constraint where the pixel values of the recovered image should be equal to the given values, i.e., $P_\Lambda\mathbf{f} = P_\Lambda\mathbf{g}$. So we are required to solve

Problem 0:

$$\begin{cases} \min ||D\alpha||_1, \\ s.t. \;\; \mathbf{f} = A^*\alpha, \\ \qquad P_\Lambda \mathbf{f} = P_\Lambda \mathbf{g}. \end{cases} \tag{3.3}$$

This is the problem proposed in [39], where a linear programming is applied to solve it. When noise is present in the observed data $P_\Lambda \mathbf{g}$, we replace the equality constraint of data fitting in (3.3) by an inequality constraint and solve

Problem 1:

$$\begin{cases} \min ||D\alpha||_1, \\ s.t. \;\; \mathbf{f} = A^*\alpha, \\ \qquad ||P_\Lambda \mathbf{f} - P_\Lambda \mathbf{g}||_2^2 \leq \sigma^2, \end{cases} \tag{3.4}$$

where σ^2 is the noise level. If we use the Language multiplier method to solve it, it gives rise to

$$\min_{\alpha}\{\frac{1}{2}||P_\Lambda(A^*\alpha) - P_\Lambda \mathbf{g}||_2^2 + \eta||D\alpha||_1\}, \tag{3.5}$$

where $1/2\eta$ is the Lagrange multiplier. Since kD is a diagonal matrix for any real number k, without loss of generalities, we would omit any constant multiplying $||D\alpha||_1$. Equation (3.5) was considered in [21], where the authors tackled the minimization problem by the iteration

$$\begin{cases} \alpha_{n+1} = T_\lambda(\alpha_n + A(P_\Lambda \mathbf{g} - P_\Lambda \mathbf{f}_n)), \\ \mathbf{f}_{n+1} = A^*\alpha_{n+1}. \end{cases} \tag{3.6}$$

The operator T_λ in (3.6) is the soft thresholding operator corresponding to the shrinkage operator in [21], and is defined as follows:

$$T_\lambda([\beta_1, \beta_2, \ldots, \beta_K]^T) \equiv [t_{\lambda_1}(\beta_1), t_{\lambda_2}(\beta_2), \ldots, t_{\lambda_K}(\beta_K)]^T \tag{3.7}$$

with $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_K]^T$ is the diagonal of D, and $t_{\lambda_i}(\cdot)$ is the soft thresholding function [17]:

$$t_{\lambda_i}(\beta_i) \equiv \begin{cases} \operatorname{sgn}(\beta_i)(|\beta_i| - \lambda_i), & if |\beta_i| > \lambda_i, \\ 0, & if |\beta_i| \leq \lambda_i. \end{cases}$$

Since we do not threshold the low-pass framelet coefficients, λ is of the form

$$\lambda = (\underbrace{0, \cdots, 0}_{N_0}, \lambda_{N_0+1}, \cdots, \lambda_K)^T, \tag{3.8}$$

where N_0 is the number of rows in A_0, $\lambda_i > 0$ for $i = N_0 + 1, \cdots, K$.

Next we move to the following minimization problem which solves Problem 2:

$$\begin{cases} \min \|D\alpha\|_1, \\ s.t. \quad \|\alpha - A\mathbf{f}\|_2^2 \leq \rho^2, \\ \qquad P_\Lambda \mathbf{f} = P_\Lambda \mathbf{g}. \end{cases} \tag{3.9}$$

Its Lagrangian function is given by

$$\min_{\mathbf{f}} \{\iota_{\mathbf{S}} + \min_{\alpha} \{\frac{1}{2}\|A\mathbf{f} - \alpha\|_2^2 + \eta \|D\alpha\|_1\}\}, \tag{3.10}$$

where

$$\mathbf{S} = \{\mathbf{f} : P_\Lambda \mathbf{f} = P_\Lambda \mathbf{g}\}$$

and $\iota_{\mathbf{S}}$ is the indicator function of $\mathbf{S}$ defined by

$$\iota_{\mathbf{S}}(\mathbf{x}) \equiv \begin{cases} 0, & \mathbf{x} \in \mathbf{S}, \\ +\infty, & \mathbf{x} \notin \mathbf{S}. \end{cases} \tag{3.11}$$

The authors in [5] proved that the minimizer of (3.10) is the convergence limit of a framelet inpainting algorithm

$$\mathbf{f}_{n+1} = P_\Lambda \mathbf{g} + (I - P_\Lambda) A^* T_\lambda (A\mathbf{f}_n), \tag{3.12}$$

which was considered in [7].

3.3 Other Minimization Problems

In this section, we induce other minimization problems from Problems 1 and 2, and compare their differences. By using $||\mathbf{f} - A^*\alpha||_2 \leq \rho^2$ instead of $||\alpha - A\mathbf{f}||_2 \leq \rho^2$ in Problem 2, we obtain

Problem 3:

$$\begin{cases} \min ||D\alpha||_1, \\ s.t. \quad ||\mathbf{f} - A^*\alpha||_2^2 \leq \rho^2, \\ \qquad P_\Lambda \mathbf{f} = P_\Lambda \mathbf{g}. \end{cases} \tag{3.13}$$

Similarly, replacing $\mathbf{f} = A^*\alpha$ by $\alpha = A\mathbf{f}$ in Problem 1 gives us

Problem 4:

$$\begin{cases} \min ||D\alpha||_1, \\ s.t. \quad \alpha = A\mathbf{f}, \\ \qquad ||P_\Lambda \mathbf{f} - P_\Lambda \mathbf{g}||_2^2 \leq \sigma^2. \end{cases} \tag{3.14}$$

We note that when A is an orthogonal basis, Problem 1 and Problem 4 are identical, since if $\alpha = A\mathbf{f}$, $\mathbf{f} = A^{-1}\alpha = A^*\alpha$ and vice versa. Similarly, Problem 2 and Problem 3 are also identical. However, in general, when A is a framelet system, the relation between α and $\mathbf{f}$ is many-to-one due to the redundance of tight frame systems, so $\mathbf{f} = A^*\alpha$ does not implies $\alpha = A\mathbf{f}$. In fact, as we will show in Section 4, Problem 3 can be solved by an algorithm that is essentially the same as equation (3.6).

Besides minimizing the (weighted) ℓ_1 norm of framelet coefficients to increase sparsity, we would also like to choose a good α so that the roughness of the inpainting solution is under control. This implies that the penalty function should somehow link to the true solution via some function norms. It is shown from framelet theory (see for e.g. [4, 24]) that the (weighted) ℓ_1 norm of the canonical framelet coefficient sequence of a function is equivalent to its Besov norm in the space $B^{\sigma}_{1,1}$ under some mild conditions. As a result, we want α to be a canonical framelet coefficient sequence, i.e., in the range of A, so that we can make sure that the (weighted) ℓ_1 norm of α is linked to the Besov norm of the underlying function.

Reviewing the problem formulations above, we can see that Problems 0, 1 and 3 make no restriction on α to the range of A. Since α does not necessary equal $AA^*\alpha$, restriction on $A^*\alpha$ tells us nothing about the relationship between α and the range of A. For Problem 2, a constraint has been set up so that α can be close to the canonical coefficient sequence. Lastly, α is required to lie on the range of A in Problem 4, since it requires $\alpha = A\mathbf{f}$. So the term that minimizes the (weighted) ℓ_1-norm of the canonical coefficient sequence $\alpha = A\mathbf{f}$ contained in Problem 4 ensures that the Besov norm of the solution is under control.

Chapter 4

Algorithm from minimization problem 3

In this chapter, we derive an algorithm that can be used to solve Problem 3. To see this, we start with a well known equivalence between soft-thresholding and a minimization functional.

Lemma 1 *The soft-thresholding operator T_λ, defined by* (3.7)*, satisfies*

$$T_\lambda(\beta) = \arg\min_\alpha \{||D\alpha||_1 + \frac{1}{2}||\beta - \alpha||_2^2\}, \tag{4.1}$$

where $\alpha, \beta, \lambda \in \mathbb{R}^K$ and D is a diagonal matrix with $D_{ii} = \lambda_i$.

Proof: Following the equation (2.35) in [14],

$$t_d(b) = \arg\min_a \{|da| + \frac{1}{2}||b - a||_2^2\}, \tag{4.2}$$

where a, b are real numbers and d is a non-negative number. We note that the minimization problem in (4.1) can be decomposed into disjoint 1-dimensional minimzation problems in the form of (4.2) for each coordinate. Therefore, $T_\lambda(\beta)$ is the minimizer of the minimization problem (4.1). □

Let us see the above lemma in the context of convex analysis. Let us recall two basic definitions by Moreau [23, 31, 32]. For any convex, lower semi-continuous function φ, its *Moreau envelope* of index $\gamma \in (0, +\infty)$ is a function defined by

$$^{\gamma}\varphi(\mathbf{x}) \equiv \min_{\mathbf{y}}\{\varphi(\mathbf{y}) + \frac{1}{2\gamma}||\mathbf{x}-\mathbf{y}||_2^2\}, \tag{4.3}$$

and its *proximity operator* is defined by

$$\text{prox}_{\varphi}(\mathbf{x}) \equiv \arg\min_{\mathbf{y}}\{\varphi(\mathbf{y}) + \frac{1}{2}||\mathbf{x}-\mathbf{y}||_2^2\}. \tag{4.4}$$

By [14, Lemma 2.5], the function $^{1}\varphi$ is convex and differentiable, with its gradient

$$\nabla(^{1}\varphi(\mathbf{x})) = \mathbf{x} - \text{prox}_{\varphi}(\mathbf{x}). \tag{4.5}$$

Writing $\xi(\alpha) \equiv ||D\alpha||_1$ and substituting (4.1) into (4.4), we obtain

$$\text{prox}_{\xi} = T_{\lambda}. \tag{4.6}$$

Therefore, the soft-thresholding operator is the proximity operator of the function $\xi(\alpha) = ||D\alpha||_1$.

We now give the Lagrangian function of (3.13),

$$\min_{\alpha}\{\xi(\alpha) + \min_{\mathbf{f}}\{\iota_{\mathbf{S}}(\mathbf{f}) + \frac{1}{2}||A^*\alpha - \mathbf{f}||_2^2\}\}, \tag{4.7}$$

Substituting (4.3) in the above equation yields

$$\min\{\xi(\alpha) +^{1} \iota_{\mathbf{S}}(A^*\alpha)\}. \tag{4.8}$$

The authors of [14] suggested a fixed point iteration for solving minimization problems. We find that their idea can be applied here. We now state the main convergence theorem in [14] for the finite dimensional case.

Theorem 3 *Consider the minimization problem*

$$\min_{\mathbf{f}}\{F_1(\mathbf{f}) + F_2(\mathbf{f})\}, \tag{4.9}$$

where F_1 is a convex, lower semi-continuous function and F_2 is a convex, differentiable function with a $1/b$-Lipschitz continuous gradient. Assume a minimizer to (4.9) *exists and $b > 1/2$. Then for any initial guess $\mathbf{f}_0$, the following iteration*

$$\mathbf{f}_{n+1} = \text{prox}_{F_1}(\mathbf{f}_n - \nabla F_2(\mathbf{f}_n)) \tag{4.10}$$

converges to a minimizer of (4.9).

To apply the above theorem to our minimization problem, we define $F_1(\alpha) \equiv \xi(\alpha)$ and $F_2(\alpha) \equiv {}^1\iota_{\mathbf{S}}(A^*\alpha)$. By putting (4.5) with $\varphi = \iota_{\mathbf{S}}$ and (4.6) into (4.10), we have

$$\begin{aligned} \alpha_{n+1} &= \text{prox}_{\xi}(\alpha_n - \nabla({}^1\iota_{\mathbf{S}}(A^*\alpha_n))) \\ &= T_\lambda(\alpha_n - A(A^*\alpha_n - \text{prox}_{\iota_{\mathbf{S}}}(A^*\alpha_n))), \end{aligned} \tag{4.11}$$

where we note that the gradient is taken with respect to α_n, so by the chain rule,

$$\nabla({}^1\iota_{\mathbf{S}}(A^*\alpha_n)) = A(A^*\alpha_n - \text{prox}_{\iota_{\mathbf{S}}}(A^*\alpha_n)). \tag{4.12}$$

We are going to find what $\text{prox}_{\iota_{\mathbf{S}}}$ is. Note that the set $\mathbf{S}$ is convex since it is the kernel of P_Λ plus the vector $P_\Lambda\mathbf{g}$. Any vector $\mathbf{x} \in \mathbb{R}^N$ has a unique projection onto $\mathbf{S}$, denoted by $P_{\mathbf{S}}(\mathbf{x})$. This projection is the minimizer of the distance from $\mathbf{x}$ among all elements in $\mathbf{S}$, i.e.,

$$P_{\mathbf{S}}(\mathbf{x}) = \arg\min_{y \in \mathbf{S}} \frac{1}{2}||\mathbf{x} - \mathbf{y}||_2^2.$$

Lemma 2 *The projection $P_{\mathbf{S}}(\mathbf{x})$ satisfies*

(a)

$$P_{\mathbf{S}}(\mathbf{x}) = \arg\min_{\mathbf{y}}\{\frac{1}{2}||\mathbf{x} - \mathbf{y}||_2^2 + \iota_{\mathbf{S}}(\mathbf{y})\} = \text{prox}_{\iota_{\mathbf{S}}}(\mathbf{x}),$$

(b)

$$P_{\mathbf{S}}(\mathbf{x}) = P_\Lambda\mathbf{g} + (I - P_\Lambda)\mathbf{x}.$$

Proof: Part (a) follows from the definition of $\iota_{\mathbf{S}}$ and $\text{prox}_{\iota_{\mathbf{S}}}$. In part (b), we note that for any vector $\mathbf{y} \in \mathbf{S}$, we have

$$\begin{aligned} ||\mathbf{y}-\mathbf{x}||_2^2 &= \textstyle\sum_{i\in\Lambda}(y_i - x_i)^2 + \sum_{i\in\Omega\setminus\Lambda}(y_i - x_i)^2 \\ &= \textstyle\sum_{i\in\Lambda}(g_i - x_i)^2 + \sum_{i\in\Omega\setminus\Lambda}(y_i - x_i)^2 \\ &\geq \textstyle\sum_{i\in\Lambda}(g_i - x_i)^2 \\ &= ||P_\Lambda(\mathbf{g}-\mathbf{x})||_2^2 \\ &= ||(P_\Lambda\mathbf{g} + (I - P_\Lambda)\mathbf{x}) - \mathbf{x}||_2^2. \end{aligned}$$

Therefore, $P_\Lambda\mathbf{g} + (I - P_\Lambda)\mathbf{x}$ is the closest vector to $\mathbf{x}$ in $\mathbf{S}$. □

By Lemma 2, we can rewrite (4.11) as

$$\begin{aligned} \alpha_{n+1} &= T_\lambda(\alpha_n - A(A^*\alpha_n - (P_\Lambda\mathbf{g} + (I - P_\Lambda)A^*\alpha_n))) \\ &= T_\lambda(\alpha_n + AP_\Lambda(\mathbf{g} - A^*\alpha_n)). \end{aligned} \tag{4.13}$$

Meanwhile we can obtain the vector $\mathbf{f}_n$ by solving $\min_{\mathbf{f}}\{\iota_{\mathbf{S}}(\mathbf{f}) + \frac{1}{2}||\mathbf{f} - A^*\alpha_n||_2^2\}$ from (4.7). By Lemma 2 again, we obtain $\mathbf{f}_n = P_\Lambda\mathbf{g} + (I - P_\Lambda)A^*\alpha_n$. We have an algorithm given as follows:

Algorithm 1 .

(i) Set an initial guess $\mathbf{f}_0$*, define* $\alpha_0 = A\mathbf{f}_0$.

(ii) Iterate on n *until convergence:*

$$\alpha_{n+1} = T_\lambda(\alpha_n + A(P_\Lambda\mathbf{g} - P_\Lambda A^*\alpha_n)). \tag{4.14}$$

(iii) Let α^* *be the output of Step (ii). We set* $\mathbf{f}^* = P_\Lambda\mathbf{g} + (I - P_\Lambda)A^*\alpha^*$ *to be the solution.*

But the iteration equation (4.14) is identical to (3.6) and its convergence is given in [21].

Chapter 5

Algorithm from minimization problem 4

In this chapter, we are going to derive a new algorithm using the approach similar to the last chapter. The proof of convergence of the algorithm will be given in the next chapter.

The Lagrangian function of (3.14) is given as follows:

$$\min_{\mathbf{f}}\{||DA\mathbf{f}||_1 + \frac{1}{2}||P_\Lambda \mathbf{f} - P_\Lambda \mathbf{g}||_2^2\}, \tag{5.1}$$

To apply Theorem 3 in this case, we define $F_1(\mathbf{f}) = ||DA\mathbf{f}||_1$ and $F_2(\mathbf{f}) = ||P_\Lambda \mathbf{f} - P_\Lambda \mathbf{g}||_2^2/2$. The function $F_2(\mathbf{f})$ is differentiable and $\nabla F_2(\mathbf{f}) = P_\Lambda(\mathbf{f} - \mathbf{g})$. Combining everything in (4.10), we obtain

$$\begin{aligned} \mathbf{f}_{n+1} &= \text{prox}_{F_1}(\mathbf{f}_n - P_\Lambda(\mathbf{f}_n - \mathbf{g})) \\ &= \text{prox}_{F_1}(P_\Lambda \mathbf{g} + (I - P_\Lambda)\mathbf{f}_n) = \text{prox}_{F_1}(P_{\mathbf{S}}(\mathbf{f}_n)), \end{aligned} \tag{5.2}$$

We now find prox_{F_1} using duality formulation. Here, we use an idea which is first proposed by Chambolle on total variation minimization problem [6], see also Chapter 2. First we observe that

$$F_1(\mathbf{f}) = ||DA\mathbf{f}||_1 = \sup_{v \in \mathbf{E}} \langle DA\mathbf{f}, v\rangle,$$

where $\mathbf{E} = \{p : p \in \mathbb{R}^K, |p_i| \leq 1, \forall\ 1 \leq i \leq K\}$. This means that

$$F_1(\mathbf{f}) = \sup_{v \in \mathbf{E}} \langle \mathbf{f}, A^* Dv \rangle = \sup_{w \in \mathbf{C}} \langle \mathbf{f}, w \rangle,$$

where $\mathbf{C} = \{A^* Dp : p \in \mathbb{R}^K, |p_i| \leq 1, \forall\ 1 \leq i \leq K\}$. By (2.27) and (2.39) in [14], $\text{prox}_{F_1}(\mathbf{f})$ is given by

$$\text{prox}_{F_1}(\mathbf{f}) = \mathbf{f} - P_{\mathbf{C}}(\mathbf{f}). \tag{5.3}$$

The projection $P_{\mathbf{C}}(\mathbf{f})$ is the minimizer of the distance from $\mathbf{f}$ among all elements in $\mathbf{C}$. Finding $P_{\mathbf{C}}(\mathbf{f})$ amounts to solving

$$\min\{||A^* Dp - \mathbf{f}||_2^2 : p \in \mathbb{R}^K, |p_i|^2 - 1 \leq 0, \forall i = 1, \cdots, K\}. \tag{5.4}$$

To simplify the calculation, $|p_i| \leq 1$ is replaced by $|p_i|^2 \leq 1$.

The Karush-Kuhn-Tucker conditions (see for e.g. [23]) yield the existence of a Lagrange multiplier $\zeta_i \geq 0$ such that for all i, we have

$$(DA(A^* Dp - \mathbf{f}))_i + \zeta_i p_i = 0 \tag{5.5}$$

with either $\zeta_i > 0$, $|p_i| = 1$ or $|p_i| < 1$, $\zeta_i = 0$. In the latter case $(DA(A^* Dp - \mathbf{f}))_i = 0$. Therefore in both cases,

$$\zeta_i = |(DA(A^* Dp - \mathbf{f}))_i|.$$

So we arrive at the following fixed point algorithm: we choose $\tau > 0$ and an initial guess p^0. For any $k \geq 0$,

$$p_i^{k+1} = p_i^k - \tau(DA(A^* Dp^k - \mathbf{f}))_i + |(DA(A^* Dp^k - \mathbf{f}))_i| p_i^{k+1}), \tag{5.6}$$

so that

$$p_i^{k+1} = \frac{p_i^k - \tau(DA(A^* Dp^k - \mathbf{f}))_i}{1 + \tau|(DA(A^* Dp^k - \mathbf{f}))_i|}. \tag{5.7}$$

If the above iteration converges to $\bar{p}$, since $\bar{p}$ is the minimizer of (5.4), $P_{\mathbf{C}}(\mathbf{f}) = A^* D\bar{p}$.

We thus propose the whole algorithm for solving problem 4 :

Algorithm 2 .

(i) Set an initial guess $\mathbf{f}_0$.

(ii) Outer iteration: Iterate (Steps (a)-(c)) on n *until convergence:*

(a)

$$\mathbf{f}_{n+\frac{1}{2}} = P_\Lambda \mathbf{g} + (I - P_\Lambda)\mathbf{f}_n. \tag{5.8}$$

(b) Inner iteration:

Choose appropriate τ *and initial guess* $p^0 = \mathbf{0}$*, iterate on* k *until convergence:*

$$p_i^{k+1} = \frac{p_i^k - \tau(DA(A^*Dp^k - \mathbf{f}_{n+\frac{1}{2}}))_i}{1 + \tau|(DA(A^*Dp^k - \mathbf{f}_{n+\frac{1}{2}}))_i|}. \tag{5.9}$$

(c) Let $\bar{p}$ *be the output of Step (b). We set* $\mathbf{f}_{n+1} = \mathbf{f}_{n+\frac{1}{2}} - A^*D\bar{p}$ *and go back to Step (a).*

We are going to prove the convergence of Algorithm 2 in the next chapter.

Chapter 6

Convergence of Algorithm 2

This chapter is dedicated to showing the convergence of Algorithm 2. We show in Section 6.1 that in the inner iteration, $\lim_{k\to\infty} p^k = \bar{p}$ exists and $A^*D\bar{p} = P_{\mathbf{C}}(\mathbf{f})$. In Section 6.2, we show that (5.1) has a minimizer, which is the convergence limit of the sequence $\{\mathbf{f}_n\}$ in the outer iteration.

6.1 Inner Iteration

We now show the condition for the inner iteration to converge. The prove mimics the one in [6].

Theorem 4 *Let $\lambda_{\max} = \max_{1\le i\le K} \lambda_i$. If $\tau \le 1/\lambda_{\max}^2$, then A^*Dp^k converges to $P_{\mathbf{C}}(\mathbf{f})$ as $k \to \infty$.*

Proof: By induction, it is obvious that for every $k \ge 0$, $|p_i^k| \le 1$ for all i. Fix $k \ge 0$, and let $q = \frac{p^{k+1}-p^k}{-\tau}$. We get

$$\begin{aligned} ||A^*Dp^{k+1} - \mathbf{f}||^2 &= ||A^*Dp^k - \mathbf{f}||^2 - 2\tau\langle A^*Dp^k - \mathbf{f}, A^*Dq\rangle + \tau^2||A^*Dq||^2 \\ &\le ||A^*Dp^k - \mathbf{f}||^2 - \tau(2\langle DA(A^*Dp^k - \mathbf{f}), q\rangle - \lambda_{\max}^2\tau||q||^2), \end{aligned}$$

the last inequality follows from $||A^*Dq|| \leq ||A^*|| ||D|| ||q|| \leq \lambda_{\max}||q||$. Now,

$$\begin{aligned} & 2\langle DA(A^*Dp^k - \mathbf{f}), q\rangle - \lambda_{\max}^2\tau||q||^2 \\ = & \textstyle\sum_{i=1}^K 2q_i(DA(A^*Dp^k - \mathbf{f}))_i - \lambda_{\max}^2\tau|q_i|^2. \end{aligned}$$

From (5.6) and the definition of q,

$$q_i = (DA(A^*Dp^k - \mathbf{f}))_i + \rho_i,$$

with

$$\rho_i = |(DA(A^*Dp^k - \mathbf{f}))_i|p_i^{k+1}.$$

Then for every i,

$$\begin{aligned} & 2q_i(DA(A^*Dp^k - \mathbf{f}))_i - \lambda_{\max}^2\tau|q_i|^2 \\ = & |q_i|^2 + |(DA(A^*Dp^k - \mathbf{f}))_i|^2 - |\rho_i|^2 - \lambda_{\max}^2\tau|q_i|^2 \\ = & (1 - \lambda_{\max}^2\tau)|q_i|^2 + (|(DA(A^*Dp^k - \mathbf{f}))_i|^2 - |\rho_i|^2). \end{aligned}$$

Since $|p_i^{k+1}| \leq 1$, $|\rho_i| \leq |(DA(A^*Dp^k - \mathbf{f}))_i|$. Hence, if $\lambda_{\max}^2\tau \leq 1$, i.e., $\tau \leq 1/\lambda_{\max}^2$, then it reveals that $||A^*Dp^k - \mathbf{f}||$ is decreasing with k.

Furthermore, when $\tau \leq 1/\lambda_{\max}^2$, if $||A^*Dp^{k+1} - \mathbf{f}||^2 = ||A^*Dp^k - \mathbf{f}||^2$, we can deduce that $|\rho_i| = |(DA(A^*Dp^k - \mathbf{f}))_i|$ for every i. Here, the definition of ρ_i implies either $|(DA(A^*Dp^k - \mathbf{f}))_i| = 0$ or $|p_i^{k+1}| = 1$. In either case, (5.7) gives us $p_i^{k+1} = p_i^k$.

Let $m = \lim_{k\to\infty} ||A^*Dp^k - \mathbf{f}||$, and $\bar{p}$ be the limit of a convergence subsequence p^{k_l} of p^k. Let $\hat{p}$ be the limit of p^{k_l+1}, from (5.7), we have

$$\hat{p}_i = \frac{\bar{p}_i - \tau((DA(A^*D\bar{p} - \mathbf{f}))_i}{1 + \tau|(DA(A^*D\bar{p} - \mathbf{f}))_i|}. \tag{6.1}$$

Since $m = ||A^*D\bar{p} - \mathbf{f}|| = ||A^*D\hat{p} - \mathbf{f}||$, by repeating previous calculations, $\bar{p}_i = \hat{p}_i$ for each i, i.e., $\bar{p} = \hat{p}$. Hence

$$(DA(A^*D\bar{p} - \mathbf{f}))_i + |(DA(A^*D\bar{p} - \mathbf{f}))_i|\bar{p}_i = 0,$$

i.e., $\bar{p}$ solves (5.4) and $A^*D\bar{p} = P_{\mathbf{C}}(\mathbf{f})$. Since $\mathbf{C}$ is a convex set, the projection is unique, we deduce that the sequence A^*Dp^k converges to $P_{\mathbf{C}}(\mathbf{f})$. □

We remark that the proof of convergence is independent of the starting point p^0, as long as $|p_i^0| \leq 1$ for all i.

6.2 Outer Iteration

To prove that the outer iteration converges to a minimizer of (5.1), we can apply Theorem 3. We merely have to check that the conditions of Theorem 3 hold.

Lemma 3 *The functions $||DA\mathbf{f}||_1$ and $\frac{1}{2}||P_\Lambda \mathbf{f} - P_\Lambda \mathbf{g}||_2^2$ are both convex and lower semi-continuous. Moreover, $\frac{1}{2}||P_\Lambda \mathbf{f} - P_\Lambda \mathbf{g}||_2^2$ is differentiable with a 1-Lipschitz continuous gradient.*

Proof: Since D, A, P_Λ are linear transforms, and both 1-norm and the square of 2-norm are convex and continuous, it is obvious that $||DA\mathbf{f}||_1$ and $\frac{1}{2}||P_\Lambda \mathbf{f} - P_\Lambda \mathbf{g}||_2^2$ are convex and lower semi-continuous. Note, $\frac{1}{2}||P_\Lambda \mathbf{f} - P_\Lambda \mathbf{g}||_2^2$ is differentiable and its gradient is given by $P_\Lambda(\mathbf{f} - \mathbf{g})$. All that left to prove is that the gradient is 1-Lipschitz continuous:

$$\begin{aligned} ||\nabla(\tfrac{1}{2}||P_\Lambda \mathbf{x} - P_\Lambda \mathbf{g}||_2^2) - \nabla(\tfrac{1}{2}||P_\Lambda \mathbf{y} - P_\Lambda \mathbf{g}||_2^2)||_2 &= ||P_\Lambda(\mathbf{x} - \mathbf{g}) - P_\Lambda(\mathbf{y} - \mathbf{g})||_2 \\ &= ||P_\Lambda \mathbf{x} - P_\Lambda \mathbf{y}||_2 \\ &\leq ||\mathbf{x} - \mathbf{y}||_2. \end{aligned}$$

This means that $\frac{1}{2}||P_\Lambda \mathbf{f} - P_\Lambda \mathbf{g}||_2^2$ has a 1-Lipschitz continuous gradient. □

6.2.1 Existence of minimizer

Next, we want to show the existence of minimizer of $F_1(\mathbf{f}) + F_2(\mathbf{f})$.

Lemma 4 *Let A be a tight frame system that satisfies any one of the following:*

(i) $A_0^*A_0$ does not has 1 as its eigenvalue; or

(ii) 1 is a simple eigenvalue of $A_0^*A_0$, and the corresponding eigenvector $\mathbf{u}$ satisfies $P_\Lambda \mathbf{u} \neq \mathbf{0}$.

Then the minimization problem (5.1) *has at least one minimizer.*

Proof: By Proposition 3.1(i) in [14], or the theorem in section 2.4, it suffices to prove that $F(\mathbf{f}) = F_1(\mathbf{f}) + F_2(\mathbf{f})$ is coercive, i.e. whenever $||\mathbf{f}||_2 \to \infty$, $F(\mathbf{f}) \to \infty$. Let $\lambda_{\min} = \min_{i=N_0+1}^{K} \lambda_i$, where λ_i is given in (3.8), we have

$$\begin{aligned} F(\mathbf{f}) &\geq ||DA\mathbf{f}||_1 \\ &\geq \lambda_{\min} \textstyle\sum_{i=1}^{M} ||A_i\mathbf{f}||_1 \\ &\geq \lambda_{\min} \sqrt{\textstyle\sum_{i=1}^{M} ||A_i\mathbf{f}||_2^2}. \end{aligned} \tag{6.2}$$

Note that for any $\mathbf{f} \in \mathbb{R}^N$, using (3.2), we get

$$\sum_{i=1}^{M} ||A_i\mathbf{f}||_2^2 = \mathbf{f}^*(\sum_{i=1}^{M} A_i^*A_i)\mathbf{f} = \mathbf{f}^*(I - A_0^*A_0)\mathbf{f}.$$

If condition (i) is satisfied, 0 is not an eigenvalue of $I - A_0^*A_0$. Let γ_1 be the smallest eigenvalue of $I - A_0^*A_0$, then

$$\sum_{i=1}^{M} ||A_i\mathbf{f}||_2^2 = \mathbf{f}^*(I - A_0^*A_0)\mathbf{f} \geq \gamma_1 ||\mathbf{f}||_2^2.$$

So, $F(\mathbf{f}) \geq \lambda_{\min}\sqrt{\gamma_1}||\mathbf{f}||_2 \to \infty$ whenever $||\mathbf{f}||_2 \to \infty$.

On the other hand, if condition (ii) is satisfied, let $\mathbf{u}$ be the eigenvector of $A_0^*A_0$ corresponding to the eigenvalue 1, and V be the subspace orthogonal to $\mathbf{u}$, i.e., $\mathbf{V} = \{\mathbf{f} : \mathbf{u}^T\mathbf{f} = 0\}$. Since 1 is a simple eigenvalue of $A_0^*A_0$, the matrix

$I - A_0^* A_0$ has a simple eigenvalue of 0. Furthermore, the corresponding null space is 1-dimensional and is spanned by $\mathbf{u}$. Therefore, an arbitrary $\mathbf{f} \in \mathbb{R}^N$ can be decomposed into $\mathbf{f} = k\mathbf{u} + \mathbf{f_V}$, where $\mathbf{f_V} \in \mathbf{V}$, and

$$\sum_{i=1}^{M} ||A_i \mathbf{f}||_2^2 = \mathbf{f}^*(I - A_0^* A_0)\mathbf{f} \geq \gamma_2 ||\mathbf{f_V}||_2^2, \tag{6.3}$$

where γ_2 be the smallest eigenvalue of $I - A_0^* A_0$ other than 0.

Now, if $||\mathbf{f}||_2$ goes to infinity, at least one of $||\mathbf{f_V}||_2$ and $|k|$ must go to infinity. If $||\mathbf{f_V}||_2 \to \infty$, then by (6.2) and (6.3), $F(\mathbf{f}) \geq \lambda_{\min}\sqrt{\gamma_2}||\mathbf{f_V}||_2 \to \infty$, and the proof is completed. Otherwise, if $||\mathbf{f_V}||_2$ is bounded, then $|k| \to \infty$. Since $P_\Lambda \mathbf{u} \neq 0$, there exist at least one index $i_0 \in \Lambda$ such that $\mathbf{u}(i_0) \neq 0$. This implies that $k\mathbf{u}(i_0)$ and hence $\mathbf{f}(i_0)$ would go to infinity, very far from the given data $\mathbf{g}(i_0)$. Therefore, $||P_\Lambda \mathbf{f} - P_\Lambda \mathbf{g}||_2^2 \to \infty$. Hence, $F(\mathbf{f}) \geq ||P_\Lambda \mathbf{f} - P_\Lambda \mathbf{g}||_2^2 \to \infty$ and we are done. □

Putting the above two lemmas together, and using Theorem 3, we now state the convergence theorem of Algorithm 2.

Theorem 5 *If one of the conditions (i) and (ii) in Lemma 4 holds, then Algorithm* 2 *converges to a minimizer of the minimization problem* (5.1) *for any initial guess.*

Now, we can show the convergence if the example in (2.14) is used.

Corollary 1 *With the tight frame system* $A = M$ *in* (2.14), *Algorithm* 2 *converges to a minimizer of the minimization problem* (5.1) *for any initial guess.*

Proof: We prove by verifying the condition (ii) of Lemma 4. Note that by the results in [34], the matrices $H_0^{(l)}$ for $l = 1, \cdots, L$ can be diagonalized by the discrete cosine transform (DCT) matrix. Moreover, the eigenvalues $\mu_i^{(l)}$

of $H_0^{(l)}$ are given by

$$\mu_i^{(l)} = \cos^{2m}(2^{l-1}\frac{i\pi}{2N}),$$

for $0 \le i \le n-1$.

If $A = M$, then $A_0 = M_0 \equiv \prod_{l=0}^{L-1} H_0^{(L-l)}$. Since the matrices $H_0^{(l)}$ for $l = 1, \cdots, L$ can be diagonalized in DCT by the same matrix, so the eigenvalues of $A_0^* A_0$ are given by

$$\prod_{l=1}^{L} (\mu_i^{(l)})^2 = \prod_{l=1}^{L} \cos^{4m}\left(2^{l-1}\frac{i\pi}{2N}\right),$$

for $0 \le i < n-1$. Note that $\prod_{l=1}^{L} \cos^{4m}\left(2^{l-1}\frac{i\pi}{2N}\right) = 1$ if and only if $i = 0$. Hence, 1 is a simple eigenvalue of $A_0^* A_0$. Also, it can be easily verified that the corresponding eigenvector is $\mathbf{1}$, and $P_\Lambda \mathbf{1} \neq 0$. □

When $\mathbf{f}$ is 2D, the tight frame systems are constructed using the tensor products, and defined in analogous to (2.14). The low-pass filter is $M_0 \bigotimes M_0$. Hence we can get the convergence for Algorithm 2 by arguing similarly to Corollary 1.

Chapter 7

Numerical Results

In this chapter, we compare Algorithm 2, which minimizes (5.1), with (3.12) and (3.6). We use the 2D version of the multiresolution tight framelet systems, with parameters L representing the number of levels and m characterizing the filters (see section 2.2). The diagonal of the scaling matrix D is chosen to be

$$\lambda = c\cdot(\underbrace{0,\cdots,0}_{N},\underbrace{2^{-L/2},\cdots,2^{-L/2}}_{2mN},\cdots,\underbrace{2^{-l/2},\cdots,2^{-l/2}}_{2mN},\cdots,\underbrace{1,\cdots,1}_{2mN})^T, \quad (7.1)$$

where c is a parameter to be determined. We tune c in (7.1) manually until we get the best solution in terms of peak signal-to-noise ratio (PSNR).

We choose the initial guess of the inpainting image to be the cubic interpolation of the observed image $\mathbf{g}$. The iteration is stopped when $||\mathbf{f}_{n+1} - \mathbf{f}_n||_2/||\mathbf{g}||_2 \leq 10^{-4}$. To speed up Algorithm 2, while we set the initial guess, p^0, of the first inner iteration to be $\mathbf{0}$, we set the initial guess of subsequent inner iterations to be the output $\bar{p}$ obtained from the previous inner iteration.

In Figures 7.1, 7.2 and 7.3, we compare the three methods using the real image *peppers*, covered by texts. The image size is 256×256. We use the tight frame system corresponding to $m = 2$ and $L = 4$. Figure 7.1 shows the results when there is no noise, while Figure 7.2 is the results when white Gaussian noise with standard deviation equals to 5 is added, and Figure 7.3 corresponds to the results when white Gaussian noise with standard deviation equals to 10 is present. From these images, the results of the three methods are close, and Algorithm 2 performs better.

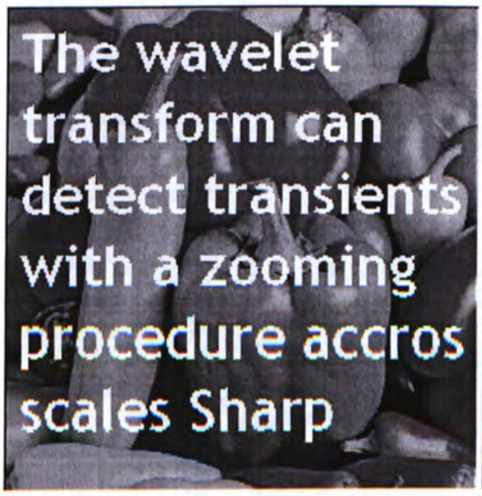
	Algorithm (3.12)	Algorithm (3.6)	Algorithm 2
no noise	33.83.	33.82	33.85
$\sigma = 5$	31.18	31.22	31.31
$\sigma = 10$	27.84	27.81	27.96

Next, we present the results when applying the above methods in zooming. The images are down-sampled, in which only pixels with odd-odd indices are known. The test images are *camera* and *lena*, both of sizes 256×256. We have added white Gaussian noise with standard derivation 5 to both images. The parameters of the tight frame system are also $m = 2$ and $L = 4$. The results for *camera* are shown in Figure 7.4, while the results for *lena* are shown in Figure 7.5.

The PSNRs of the restored images by the three methods are tabulated as follows. Again, the three methods give similar results while Algorithm 2 is slightly better.

	Algorithm (3.12)	Algorithm (3.6)	Algorithm 2
Camera	25.31	25.22	25.31
Lena	28.40	28.41	28.45

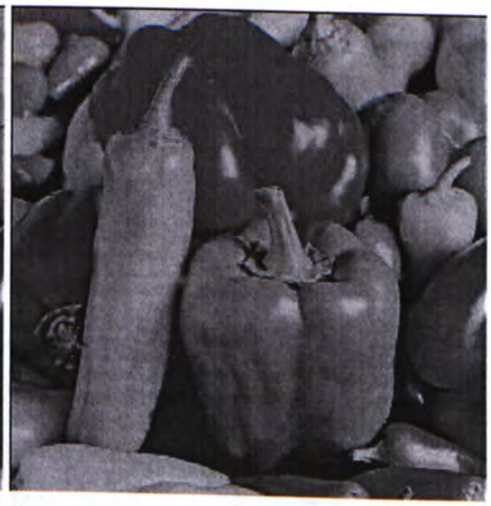
Figure 7.1: The image "Pepper". From left to right: Image covered by texts; the restoration by (3.12) (PSNR=33.83); the restoration by (3.6) (PSNR=33.82); and the restoration by Algorithm 2 (PSNR=33.85).



(a)

(b)

(c)



(d)

Figure 7.2: The image "Pepper". From left to right: Image with additive white Gaussian noise with standard deviation 5, and covered by texts; the restoration by (3.12) (PSNR = 31.18); the restoration by (3.6) (PSNR = 31.22); and the restoration by Algorithm 2 (PSNR = 31.31).

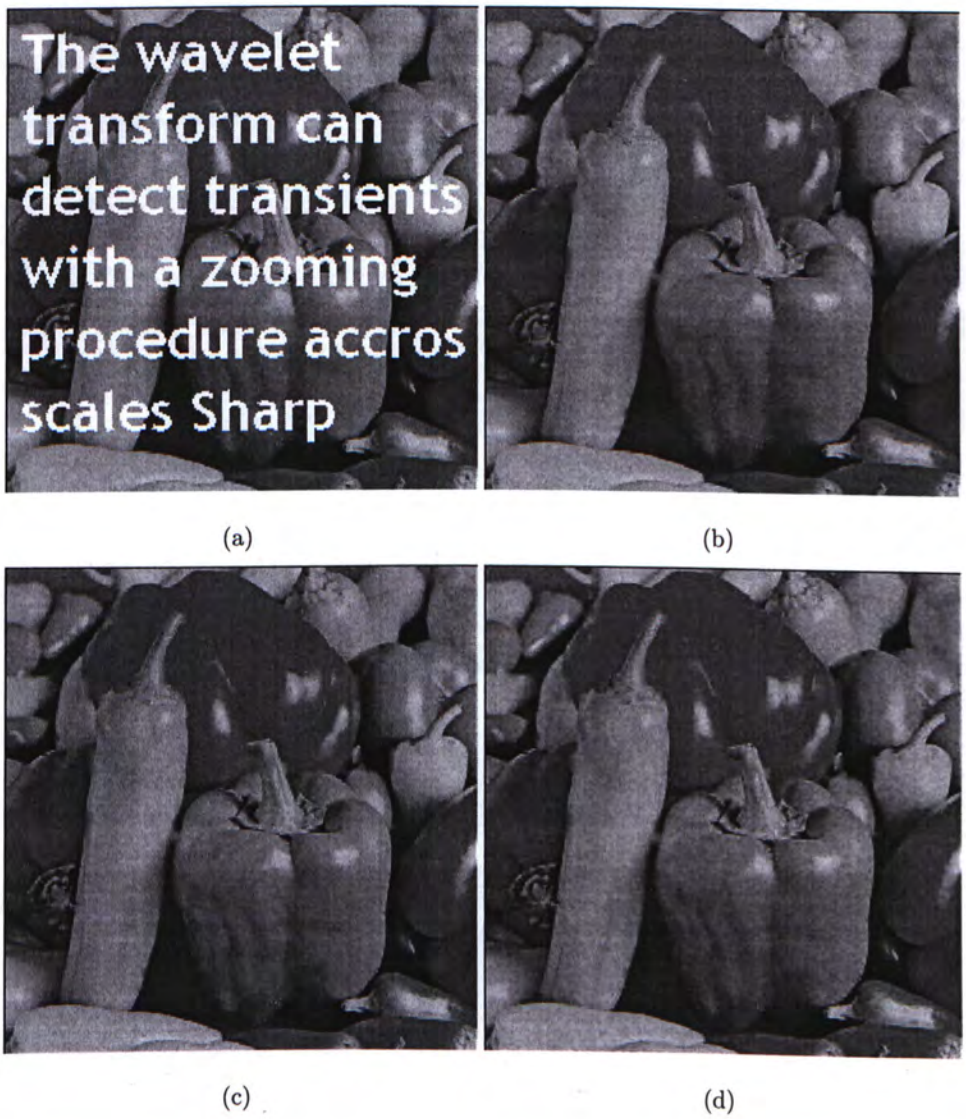
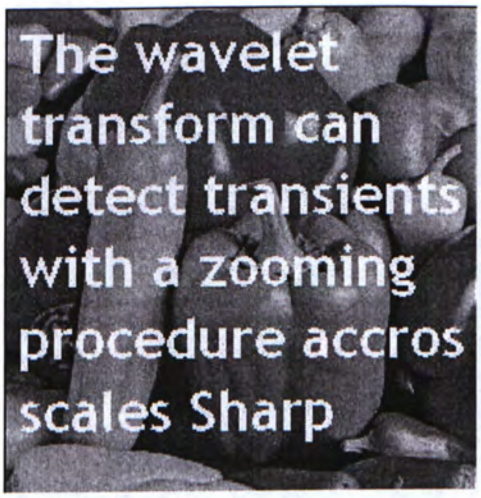


Figure 7.3: The image "Pepper". From left to right: Image with additive white Gaussian noise with standard deviation 10, and covered by texts; the restoration by (3.12) (PSNR = 27.84); the restoration by (3.6) (PSNR = 27.81); and the restoration by Algorithm 2 (PSNR = 27.96).



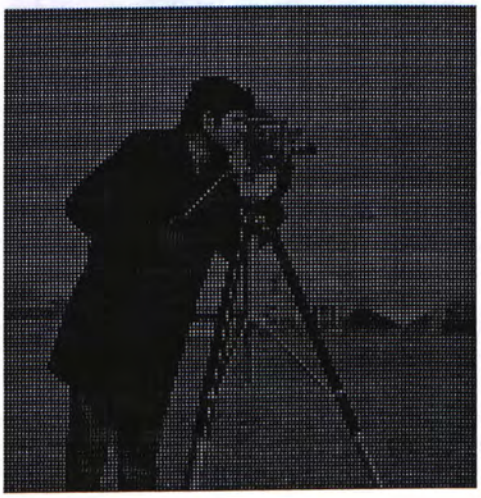
(a)

(b)

(c)

(d)

Figure 7.4: The image "Camera". From left to right: Image with additive white Gaussian noise with standard deviation 5, and only pixels of odd-odd index are observed; the restoration by (3.12) (PSNR = 25.31); the restoration by (3.6) (PSNR = 25.22); and the restoration by Algorithm 2 (PSNR = 25.31).



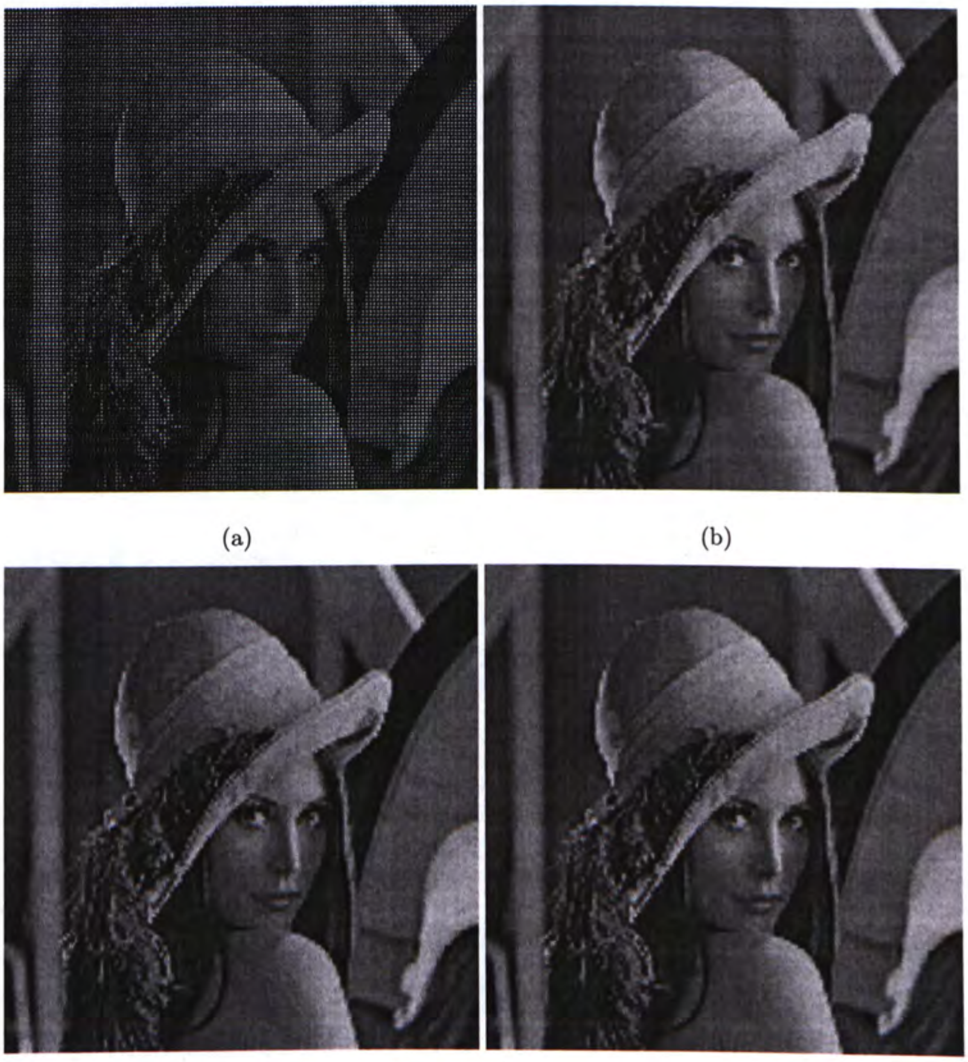
(a)

(b)

(c)

(d)

Figure 7.5: The image "Lena". From left to right: Image with additive white Gaussian noise with standard deviation 5, and only pixels of odd-odd index are observed; the restoration by (3.12) (PSNR = 28.4); the restoration by (3.6) (PSNR = 28.41); and the restoration by Algorithm 2 (PSNR = 28.45).



(a) (b)

(c) (d)

Chapter 8

Conclusion

In this thesis, we propose a new iteration algorithm for image inpainting based on framelet systems. Here, by studying the relationship between several wavelet-based methods as constrained minimization problems, we introduce a new iteration algorithm and prove its convergence. The convergence limit of the new iteration minimizes a functional that is linked to the regularity of the solution. The restored images are slightly better than that of existing methods. Faster convergence algorithms for this minimization functional are yet to be found. This may lead to a further research topic.

Bibliography

[1] S. Armstrong, A. Kokaram, and P. J. W. Rayner. Nonlinear interpolatino of missing data using min-max functions, *IEEE Int. Conf. Nonlinear Signal and Image Processings*, 1997.

[2] C. Ballester, M. Bertalmío, V. Caselles, G. Sapiro, and J. Verdera. Filling-in by joint interpolation of vector fields and grey levels, *IEEE Trans. Image Process.* 10(8), pp. 1200-1211, 2001.

[3] M. Bertalmío, G. Sapiro, V. Caselles and C. Ballester. Image Inpainting, *SIG GRAPH*, 34, pp. 417-424, 2000.

[4] L. Borup, R. Gribonval and M. Nielsen. Bi-framelet systems with few vanishing moments chatacterize Besov spaces, *Appl. Comput. Harmon. Anal.*, 17, pp. 3-28, 2004.

[5] J.F. Cai, R. Chan, and Z. Shen. A Framelet-Based Image Inpainting Algorithm, Research Report 2007-02(342), Department of Mathematics, The Chinese University of Hong Kong, 2007.

[6] A. Chambolle. An Algorithm for Total Variation Minimization and Applications, *J. Mathematical Imaging and Vision*, 20, pp. 89-97, 2004.

[7] R. Chan, L. Shen, and Z. Shen. A Framelet-Based Approach for Image Inpainting, Research Report 2005-04(325), Department of Mathematics, The Chinese University of Hong Kong, 2005.

[8] T.F. Chan, G.H. Golub, adn P. Mulet. A Nonlinear Primal-Dual Method for Total Variation-Based Image Restoration, *SIAM J. Sci. Comp.*, 20(6), pp. 1964-1977, 1999.

[9] T.F. Chan, S.H. Kang and J. Shen. Euler's elastica and curvature based inpaintings, UCLA CAM Report 2001-12, 2001.

[10] T.F. Chan, and J. Shen. Nontexture inpainting by curvature driven diffusions (CDD). *J. Visual Comm. Image Rep.* 12(4), pp. 436-449, 2001.

[11] T.F. Chan, and J. Shen. Mathematical Models for Local Non-Texture Inpainting, *SIAM J. Appl Math* 62, pp. 1019-1043, 2001.

[12] T.F. Chan, and J. Shen. A good image model eases restoration - on the contribution of Rudin-Osher-Fatemi's BV image model, *IMA Technical Report 1829 at www.ima.umn.edu/pub/pub.html*, 2002.

[13] T.F. Chan, and J. Shen. *Image processing and analysis : variational, PDE, wavelet, and stochastic methods*, Society for Industrial and Applied Mathematics, Philadelphia, 2005.

[14] P. L. Combettes and V. R. Wajs. Signal Recovery by Proximal Forward-Backward Splitting, *SIAM J. Multiscale Modeling and Simulation*, 4, pp. 1168-1200, 2005.

[15] I. Daubechies, M. Defrise, and C. De Mol. An Iterative Thersholding Algorithm for Linear Inverse Problems with a Sparsity Constraint, *Communications on Pure and Applied Mathematics*, 57, pp. 1413-1457, 2004.

[16] I. Daubechies, B. Han, A. Ron, and Z. Shen. Framelets: MRA-Based Constructions of Wavelet Frames, *Applied and Computational Harmonic Analysis*, 14, pp. 1-46, 2003.

[17] D.Donoho. De-nosing by Soft-Thresholding, *IEEE Transactions on Information Theory*, 41, pp. 613-627, 1995.

[18] D.Donoho. Compressed Sensing, *IEEE Transactions on Information Theory*, 52, pp. 1289-1306, 2006.

[19] M. Elad, J.-L. Starck, P. Querre, and D. Donoho. Simultaneous Cartoon and Texture Image Inpainting using morphplogical component analysis (MCA), *Applied and Computational Harmonic Analysis*, 19, pp. 340-358, 2005.

[20] S. Esedoglu, and J. Shen, Digital inpainting based on Mumford-Shah-Euler image model, *European Journal of Applied Mathematics*, 13, pp. 353-370, 2002.

[21] M.J. Fadili, and J.-L. Starck. Sparse Representations and Bayesian Image Inpainting, *Proc. SPARS'05*, Vol.I, Rennes, France, 2005.

[22] P. Ferreira. mathematics of multimedia signal processing ii discrete finite frames and signal reconstruction. In J. S. B. (Ed.), editor, *Signal Processing for Multimedia*, pp. 35-54, IOS Press, 1999.

[23] J.-B. Hiriart-Urruty and C. Lemaréchal. *Convex Analysis and Minimization Algorithms, vol. I*, Springer-Verlag, Berlin, 1996.

[24] Y. Hur and A. Ron. CAPlets: wavelet representations without wavelets, preprint, 2005.

[25] K. H. Jung, J. H. Chang, and C. W. Lee. Error concealment technique using data for block-based image coding, *Proc. SPIE*, 2308, pp. 1466-1477, 1994.

[26] A. C. Kokaram, R. D. Morris, W. J. Fitzgerald, and P. J. W. Rayer. Detection of missing data in image sequences I, *IEEE Trans. Image Process*, 11 pp. 1496-1508, 1995.

[27] A. C. Kokaram, R. D. Morris, W. J. Fitzgerald, and P. J. W. Rayer. Detection of missing data in image sequences II, *IEEE Trans. Image Process*, 11 pp.1509-1519, 1995.

[28] W. Kwok, and H. Sun. Multidirectional interpolation for spatial error concealment, *IEEE Trans. Consumer Electronics*, 39, 1993.

[29] S. Mallat. *A Wavlet Tour of Signal Processing, 2nd Edition*, Academic Press, 1999.

[30] S. Masnou, and J. M. Morel. Level-lines based disocclusion, *Proc. 5th IEEE Int. Conf. on Image Process.*, pp. 259-263, Chicago, IL, 1998.

[31] J.-J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien, *C.R. Acad. Sci. Paris Sér. A Math.*, 255, pp. 1897-2899, 1962.

[32] J.-J. Moreau. Proximité et dualité dans un espace hilbertien, *Bull. Soc. Math. France*, 93, pp.273-299, 1965.

[33] D. Mumford, and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems, *Comm. Pure Applied. Math.*, 42, pp. 577-685, 1989.

[34] M. Ng, R. Chan, and W. Tang. A fast algorithm for debluring models with Neumann boundary conditions, *SIAM Journal on Scientific Computing*, 21 pp. 851-866, 2000.

[35] S. Osher, and J. A. Sethian. Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton- Jacobi formulations, *J. Comput. Phys.*, 79, 1988.

[36] A. Ron, and Z. Shen. Affine system in $L_2(\mathbb{R}^d)$: the analysis of the analysis operator, *Journal Func. Anal.*, 148, pp.408-447, 1997.

[37] L. Rudin, and S. Osher. Total Variation Based Image Restoration With Free Local Constraints, *Proceedings of the International Conference on Image Processing*, Vol. I, pp. 31-35, 1994.

[38] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms, *Physica D*, 60, pp. 259-268, 1992.

[39] I.W. Selesnick, R. Van Slyke, and O.G. Guleryuz, "Pixel Recovery via ℓ_1 Minimization in the Wavelet Domain", *Proc. IEEE Intl Conf. on Image Proc.(ICIP2004)*, Vol. III, pp. 1819-1822, Singapore, 2004.

[40] A.N. Tikhonov, and V.Y. Arsenin. *Solutions of Ill-Posed Problems*, Winston and Sons, Washington, D.C., 1977.

[41] Tsai, A. Jr, A. Yezzi, and A. S. Willsky. Curve evolution implementation of the Mumford-Shah functinoal for im age segmentation, denoising, interpolation and magnification. *IEEE Trans. Image Process* 10, pp. 1169-1186, 2001.