Regularization schemes involving self-similarity in imaging inverse problems

Mehran Ebrahimi and Edward R. Vrscay

Department of Applied Mathematics
University of Waterloo
Waterloo, Ontario, Canada, N2L 3G1

E-mail: m2ebrahi@uwaterloo.ca, ervrscay@uwaterloo.ca

Abstract. In this paper we introduce and analyze a set of regularization expressions based on self-similarity properties of images in order to address the classical inverse problem of image denoising and the ill-posed inverse problem of single-frame image zooming.

The regularization expressions introduced are constructed using either the fractal image transform or the newly developed "Nonlocal-means (NL-means) image denoising filter" of Buades *et al.* (2005).

We exploit these regularization terms in a global MAP-based formulation and produce analytical and computational solutions. Analytical comparisions are made with results based on classical methods (e.g., fractal-based denoising and zooming, and NL-means image denoising).

1. Introduction

We consider the linear degradation model[1]

$$\underline{u} = \mathbf{H}\underline{X} + \underline{n}, \tag{1}$$

in which $\mathbf{H}$ is a linear operator, $u \in l^2(\Omega)$ is the $p \times q$-pixel observed image, i.e.,

$$\Omega = [1, \ldots, p] \times [1, \ldots, q], \tag{2}$$

and $n \in l^2(\Omega)$ is additive white, independent Gaussian noise with zero-mean and some known variance, and $X \in l^2(\Psi)$ is the image to be recovered such that

$$\Psi = [1, \ldots, pz] \times [1, \ldots, qz]. \tag{3}$$

Given $\underline{u}$, one is interested in finding an approximation of $\underline{X}$, to be denoted as $\underline{X^\star}$. This is normally performed by incorporating additional *a priori* information about the image, a method of regularization. Given $\underline{u}$, the approximation $\underline{X^\star}$ can be a minimizer (or if certain conditions hold, the unique minimizer) of the expression

$$\underline{X^\star} \;=\; \arg\min_{\underline{X}} \|\mathbf{H}\underline{X} - \underline{u}\|^2 + \lambda \mathbf{\Gamma}(\underline{X}). \tag{4}$$

[1] In this paper, the notation $\underline{u}$ refers to the lexicographic representation of an image u.

This is known as the *maximum a posteriori* (MAP)-based formulation in statistical estimation theory, in which $\mathbf{\Gamma}(\underline{X})$ is a regularization functional. Regularization not only acts as an *algebraic stabilizer* in estimating the solutions of ill-posed inverse problems (i.e., if $\mathbf{H}^T\mathbf{H}$ is singular), but it may also *improve* the solutions of well-posed problems, for example, in case of denoising. Equivalently, in the Bayesian point of view, the probability density function (PDF) of the image, serving as the prior $p(\underline{\mathbf{X}})$, plays the role of regularization. The simplest form of regularization, that of Tikhonov [24], applies a uniform spatial smoothness to the outcome.

In this paper, we assume that the penalty functional has the well-known form,

$$\mathbf{\Gamma}(\underline{X}) = \parallel \mathbf{C}\underline{X} - \mathbf{D} \parallel^2, \tag{5}$$

so that the minimization problem becomes

$$\underline{X^\star} \quad = \quad \arg\min_{\underline{X}} \|\mathbf{H}\underline{X} - \underline{u}\|^2 + \lambda \parallel \mathbf{C}\underline{X} - \mathbf{D} \parallel^2. \tag{6}$$

It can be shown that

$$\underline{X^\star} = (\mathbf{H}^T\mathbf{H} + \lambda\mathbf{C}^T\mathbf{C})^{-1}(\mathbf{H}^T\underline{u} + \lambda\mathbf{C}^T\mathbf{D}), \tag{7}$$

when $\mathbf{H}^T\mathbf{H} + \lambda\mathbf{C}^T\mathbf{C}$ is non-singular.

In this paper, we show how self-similarity properties of the observed data $\underline{u}$ may be used to construct the penalty functional $\mathbf{\Gamma}(\underline{X})$. We consider only two forms of the degradation operator $\mathbf{H}$: (i) for denoising, $\mathbf{H} = I$ the identity operator, and (ii) for the inverse problem of image zooming, $\mathbf{H} = \mathcal{D}$ the down-sampling operator (by a factor of z).

In Section 2, we review the role of self-similarity in inverse problems involving fractal-based and example-based approaches. Because of its relevance to this work, the non-local(NL)-means denoising [5, 6] will be discussed. In Section 3, we present regularization models for image denoising and zooming based upon fractal transforms. In Section 4, a regularization scheme based on NL-means denoising will be introduced. In Section 5, we define a contractive operator associated with the NL-means denoising method, and provide some theoretical results. A number of computational issues and results will be presented in Section 6. Some concluding remarks are presented in Section 7.

2. The role of self-similarity in various imaging inverse problems

2.1. Fractal based methods

The idea of using self-similarities for image coding and compression began with the seminal work of Barnsley and Demko [4] resulting in an intensive research activity in the 1990's – see, for example, [3, 12, 22, 23, 16]. In fractal image coding, one seeks to approximate a target image $\underline{u}$ by the fixed point $\underline{u^\star}$ of a contractive, resolution-independent operator $\mathbf{T}$ called the fractal transform. The essence of the fractal transform is to approximate smaller "range" subblocks of an image with modified copies of (subsampled) larger"domain" subblocks. Because of the complicated mixing nature of fractal transforms, it is not practical to search for an operator T that minimizes the approximation error $\parallel \underline{u} - \underline{u^\star} \parallel$. Instead, one tries to minimize the so-called "collage error" $\parallel \underline{u} - \mathbf{T}\underline{u} \parallel$, essentially the error in approximating the range subblocks with modified copies of the domain subblocks. This greedy algorithm is known as "collage coding." Appendix (A) provides a brief discussion of the important ideas fractal coding and decoding.

More recently, the ability of fractal coding to solve other inverse problems has been investigated. In [17, 18, 1], it has been observed that fractal-based methods have denoising capabilities. Due to the resolution-independent nature of the fractal transform, interpolation algorithms called "fractal zoom" have been also been developed in the literature [12, 16, 22, 23].

A major drawback of traditional fractal-based methods, however, is that their output is quite restricted, i.e., no prior knowledge, extra regularization method or tuning (regularization) parameters can be combined with these methods. Recently, we have examined various possibilities to address this problem [8, 9].

2.2. Self-Similarity in various example-based approaches

A very attractive approach for solving imaging inverse problems is to exploit examples in defining the PDF of the image instead of intuitively defining a regularization term. There are various ways to apply examples in inverse problems as comprehensively described in [11]. In the work of [7, 10, 25] on texture synthesis, and inpainting, examples are taken from the image *itself* and have been used directly in the reconstruction procedure.

Another important example-based approach, nonlocal-means (NL-means) image denoising [5, 6] addresses the denoising problem using examples *from the noisy image itself* at the *same scale* with a Gaussian-type weighting scheme. The authors have demonstrated that their algorithm has the ability to outperform classical denoising methods, including Gaussian smoothing, Wiener filter, TV filter, wavelet thresholding, and anisotropic diffusion. Later in this paper, we shall construct regularization expressions based on the NL-means algorithm. For the benefit of the reader, the NL-means algorithm is briefly reviewed in Appendix (B).

3. Regularization via the fractal-transform operator

3.1. Denoising

In this subsection we consider the following problem of image denoising: Given

$$\underline{u} = \underline{X} + \underline{n}, \tag{8}$$

find an approximation $\underline{X^\star}$ of $\underline{X}$. As mentioned earlier, it was observed [17, 18, 1] that fractal-based methods have denoising capabilities. If we assume that $\mathbf{T}$ is a contractive fractal transform of an image $\underline{u}$, then its unique fixed point $\underline{u^\star}$ is an approximation of $\underline{X}$.

We seek to improve this approximation by emphasizing the influence of the observed data $\underline{u}$ on the reconstruction process. First of all, note that we can write $\underline{u^\star}$, the unique fixed point of the contractive fractal transform $\mathbf{T}$ as

$$\underline{u^\star} = \arg\min_{\underline{X}} \| \underline{X} - \mathbf{T}(\underline{X}) \|^2. \tag{9}$$

Now $\mathbf{T}$ can be represented as follows,

$$\mathbf{T}(\underline{X}) = \mathbf{M}\underline{X} + \mathbf{B}, \tag{10}$$

where the matrices $\mathbf{M}$ and $\mathbf{B}$ contain the "greyscale map" parameters, α_i and β_i, respectively, as well as the domain-range assignments of the transform $\mathbf{T}$. (See Appendix (A) for the definitions of α_i and β_i.)

We consider two constructions of the regularization expression:

(i) Squared collage error,

$$\mathbf{\Gamma}(\underline{X}) = \| \underline{X} - \mathbf{T}(\underline{X}) \|^2 \tag{11}$$

(ii) Squared fixed-point approximation error,

$$\mathbf{\Gamma}(\underline{X}) = \| \underline{X} - \underline{u^\star} \|^2. \tag{12}$$

In the first case, where $\mathbf{\Gamma}$ is the collage error, we define $\mathbf{A} = \mathbf{I} - \mathbf{M}$, in which $\mathbf{I}$ is the identity matrix, so that

$$\mathbf{\Gamma}(\underline{X}) = \| \underline{X} - \mathbf{T}(\underline{X}) \|^2 = \| \mathbf{A}\underline{X} - \mathbf{B} \|^2. \tag{13}$$

Then

$$\begin{aligned} \underline{X^\star} &= \arg\min_{\underline{X}} \|\underline{X} - \underline{u}\|^2 + \lambda \| \underline{X} - \mathbf{T}\underline{X} \|^2 \\ &= \arg\min_{\underline{X}} \|\underline{X} - \underline{u}\|^2 + \lambda \| \mathbf{A}\underline{X} - \mathbf{B} \|^2 \\ &= \arg\min_{\underline{X}} \|\underline{X} - \underline{u}\|^2 + \lambda \| \mathbf{A}(\underline{X} - \underline{u^\star}) \|^2. \end{aligned} \tag{14}$$

It is easy to show that

$$\underline{X^\star} = (\mathbf{I} + \lambda \mathbf{A}^T\mathbf{A})^{-1}(\underline{u} + \lambda \mathbf{A}^T\mathbf{B}). \tag{15}$$

In the second case, where $\mathbf{\Gamma}$ is the squared collage error, we define

$$\underline{Y^\star} = \arg\min_{\underline{X}} \|\underline{X} - \underline{u}\|^2 + \lambda \| \underline{X} - \underline{u^\star} \|^2. \tag{16}$$

It can be easily shown that

$$\underline{Y^\star} = \frac{1}{1+\lambda}\underline{u} + \frac{\lambda}{1+\lambda}\underline{u^\star}. \tag{17}$$

This is simply a weighted average of the measurement $\underline{u}$ and the attractor $\underline{u^\star}$ of the fractal transform $\mathbf{T}$. The following is an asymptotic result for the difference of the minimizers of the above two cases.

Proposition 1.

$$\underline{Y^\star} - \underline{X^\star} = \lambda(\mathbf{I} - \mathbf{A}^T\mathbf{A})(\underline{u^\star} - \underline{u}) + O(\lambda^2) \quad \textit{as } \lambda \to 0. \tag{18}$$

Proof. Note that, as $\lambda \to 0$,

$$\underline{X^\star} = (\mathbf{I} + \lambda \mathbf{A}^T\mathbf{A})^{-1}(\underline{u} + \lambda \mathbf{A}^T\mathbf{B}) = [\mathbf{I} - \lambda \mathbf{A}^T\mathbf{A} + O(\lambda^2)](\underline{u} + \lambda \mathbf{A}^T\mathbf{B}). \tag{19}$$

Therefore,

$$\begin{aligned}
\underline{Y^\star} - \underline{X^\star} &= \frac{1}{1+\lambda}\underline{u} + \frac{\lambda}{1+\lambda}\underline{u^\star} - \underline{X^\star} \\
&= (1-\lambda)\underline{u} + \lambda\underline{u^\star} + O(\lambda^2) - \underline{X^\star} \\
&= (1-\lambda)\underline{u} + \lambda\underline{u^\star} + O(\lambda^2) - [\mathbf{I} - \lambda\mathbf{A}^T\mathbf{A} + O(\lambda^2)](\underline{u} + \lambda\mathbf{A}^T\mathbf{B}) \\
&= \underline{u} - \lambda\underline{u} + \lambda\underline{u^\star} - \underline{u} + \lambda\mathbf{A}^T\mathbf{A}\underline{u} - \lambda\mathbf{A}^T\mathbf{B} + O(\lambda^2) \\
&= \lambda(\underline{u^\star} - \underline{u} + \mathbf{A}^T(\mathbf{A}\underline{u} - \mathbf{B})) + O(\lambda^2) \\
&= \lambda(\underline{u^\star} - \underline{u} + \mathbf{A}^T(\mathbf{A}\underline{u} - \mathbf{B} - (\mathbf{A}\underline{u^\star} - \mathbf{B}))) + O(\lambda^2) \\
&= \lambda(\underline{u^\star} - \underline{u} - \mathbf{A}^T\mathbf{A}(\underline{u^\star} - \underline{u})) + O(\lambda^2) \\
&= \lambda(\mathbf{I} - \mathbf{A}^T\mathbf{A})(\underline{u^\star} - \underline{u}) + O(\lambda^2).
\end{aligned} \tag{20}$$

□

3.2. Zooming

In this subsection, we consider the following problem of image zooming: Given the observed image data $\underline{u}$, where

$$\underline{u} = \mathcal{D}\underline{X} + \underline{n} \tag{21}$$

in which $\mathcal{D}$ is the downsampling by a factor of z, find an approximation of $\underline{X}$.

In the fractal coding literature, the normal procedure of zooming an image by a factor of z is to find a fractal transform $\mathbf{T}$ for $\underline{u}$, once again by minimizing $\| \mathbf{T}\underline{u} - \underline{u} \|$. One then "zooms" this attractor by applying the fractal transform $\mathbf{T}$ (actually the operator $\mathbf{T}_z$ induced by $\mathbf{T}$) to an image that is z-times larger in each direction to produce the attractor $\underline{u_z^\star}$.

Here, we shall use the zoomed fixed-point error to define the regularization expression, i.e., $\mathbf{\Gamma}(\underline{X}) = \|\underline{X} - \underline{u_z^\star}\|^2$ so that the minimization problem becomes

$$\underline{Y_z^\star} = \arg\min_{X} \|\mathcal{D}\underline{X} - \underline{u}\|^2 + \lambda\|\underline{X} - \underline{u_z^\star}\|^2. \tag{22}$$

Notice that the above functional balances the consistency with original data, via the term $\|\mathcal{D}\underline{X} - \underline{u}\|^2$, along with the self-similarity constraints encoded in $\underline{u_z^\star}$. consistency with the original data It can be shown that the solution of this minimization problem is

$$\underline{Y_z^\star} = (\mathcal{D}^T\mathcal{D} + \lambda\mathbf{I})^{-1}(\mathcal{D}^T\underline{u} + \lambda\underline{u_z^\star}). \tag{23}$$

Some computational results are presented in Section 6.

4. A regularization scheme based on NL-means denoising

In this section we return to the denoising problem

$$\underline{u} = \underline{X} + \underline{n} \tag{24}$$

In Appendix (B) the NL-means algorithm of Buades *et al.* is briefly described, including the determination of a weight matrix $\mathbf{W}$, for some smoothness parameter h, so that $\mathbf{W}\underline{u}$ is the denoised copy of $\underline{u}$. In what follows, we make use of this matrix to improve the results.

Once again, we consider two forms for the regularization term,

(i) The collage error

$$\mathbf{\Gamma}(\underline{X}) = \| \underline{X} - \mathbf{W}\underline{X} \|^2. \tag{25}$$

(ii) The explicit approximation error

$$\mathbf{\Gamma}(\underline{X}) = \| \underline{X} - \mathbf{W}\underline{u} \|^2. \tag{26}$$

In the first case, define

$$\underline{X_m^\star} \;=\; \arg\min_{\underline{X}} \| \underline{X} - \underline{u} \|^2 + \lambda\| \underline{X} - \mathbf{W}\underline{X} \|^2, \tag{27}$$

with solution

$$\underline{X_m^\star} = [\mathbf{I} + \lambda(\mathbf{W} - \mathbf{I})^T(\mathbf{W} - \mathbf{I})]^{-1}(\underline{u}). \tag{28}$$

Based on the properties of the matrix $\mathbf{W}$, it can be shown that $[I + \lambda(\mathbf{W} - \mathbf{I})^T(\mathbf{W} - \mathbf{I})]$ is non-singular if $\lambda > 0$.

In the second case, define

$$\underline{Y_m^\star} \;=\; \arg\min_{\underline{X}} \| \underline{X} - \underline{u} \|^2 + \lambda\| \underline{X} - \mathbf{W}\underline{u} \|^2. \tag{29}$$

The solution is

$$\underline{Y_m^\star} = \frac{1}{1+\lambda}\underline{u} + \frac{\lambda}{1+\lambda}\mathbf{W}\underline{u}. \tag{30}$$

The minimizer $\underline{Y_m^\star}$ obtained in the second case is exactly the one proposed in [5, 6] as a way of achieving superior results by taking a weighted average of the output of the NL-means denoising and the original image. Below is an asymptotic result on the difference of the solutions of the above two cases.

Proposition 2.

$$\underline{Y_m^\star} - \underline{X_m^\star} = \lambda\mathbf{W}^T(\mathbf{W} - \mathbf{I})\underline{u} + O(\lambda^2) \quad \textit{as } \lambda \to 0. \tag{31}$$

Proof. Note that as $\lambda \to 0$, $\underline{Y_m^\star} = \lambda \mathbf{W}\underline{u} + (1-\lambda)\underline{u} + O(\lambda^2)$. Also,

$$\begin{aligned} \underline{X_m^\star} &= [\mathbf{I} + \lambda(\mathbf{W}-\mathbf{I})^T(\mathbf{W}-\mathbf{I})]^{-1}\underline{u} \\ &= [\mathbf{I} - \lambda(\mathbf{W}-\mathbf{I})^T(\mathbf{W}-\mathbf{I})]\underline{u} + O(\lambda^2). \end{aligned} \tag{32}$$

Hence,

$$\begin{aligned} \underline{Y_m^\star} - \underline{X_m^\star} &= \lambda \mathbf{W}\underline{u} + (1-\lambda)\underline{u} - [\mathbf{I} - \lambda(\mathbf{W}-\mathbf{I})^T(\mathbf{W}-\mathbf{I})]\underline{u} + O(\lambda^2) \\ &= \lambda \mathbf{W}\underline{u} + \underline{u} - \lambda\underline{u} - \underline{u} + \lambda \mathbf{W}^T\mathbf{W}\underline{u} - \lambda \mathbf{W}\underline{u} - \lambda \mathbf{W}^T\underline{u} + \lambda\underline{u} + O(\lambda^2) \\ &= \lambda \mathbf{W}^T\mathbf{W}\underline{u} - \lambda \mathbf{W}^T\underline{u} + O(\lambda^2) \\ &= \lambda \mathbf{W}^T(\mathbf{W}-\mathbf{I})\underline{u} + O(\lambda^2). \end{aligned} \tag{33}$$

□

5. A contractive operator associated with NL-means denoising

It is interesting to investigate the consequence of iterating the NL-means denoising operator $\mathbf{W}\underline{X}$ on images. In what follows, we show that $\mathbf{W}$ is a not a projection operator, i.e, $\mathbf{W}^2\underline{X}$ is not neccessarily equal to $\mathbf{W}\underline{X}$ in general.

Proposition 3. *In general,* $\mathbf{W}$ *is not a projection operator, in which case* $\lim_{k\to\infty} \mathbf{W}^k\underline{X}$ *is a constant-valued image.*

Proof. Note that W is a square matrix whose rows consists of nonnegative real numbers, with each row summing to 1, i.e., W a right stochastic matrix. The Perron-Frobenius theorem for the right stochastic matrix W ensures that there exists a stationary probability vector π such that

$$\pi \mathbf{W} = \pi. \tag{34}$$

The jth element of the vector π may be computed by taking the limit

$$\lim_{k\to\infty} (\mathbf{W}^k)_{(i,j)} = \pi_j. \tag{35}$$

Hence, independent of i,

$$\begin{aligned} \forall i \quad (\lim_{k\to\infty} \mathbf{W}^k\underline{X})_i &= \lim_{k\to\infty} \sum_j (\mathbf{W}^k)_{(i,j)}\underline{X}_j \\ &= \sum_j \lim_{k\to\infty} (\mathbf{W}^k)_{(i,j)}\underline{X}_j \\ &= \sum_j \pi_j \underline{X}_j \\ &= \pi\underline{X} = c. \end{aligned} \tag{36}$$

Therefore,

$$\lim_{k\to\infty} \mathbf{W}^k\underline{X} = (c, c, \ldots, c)^T \tag{37}$$

is a constant-valued image. If $\mathbf{W}$ is a projection operator then

$$\mathbf{W}\underline{X} = \mathbf{W}^n\underline{X} = \lim_{k\to\infty} \mathbf{W}^k\underline{X} = (c, c, \ldots, c)^T, \text{ for any } n \geq 1, \tag{38}$$

which does not hold in general. Hence, $\mathbf{W}$ is not a projection operator. □

We now define a contracting denoising operator, $\mathbf{S}$, which acts on $\underline{X}$ by producing a linear combination of the original image, $\underline{u}$, and $\mathbf{W}\underline{X}$, which is a denoised copy of $\underline{X}$ under $\mathbf{W}$:

$$\mathbf{S}(\underline{X}) = (1-\eta)\underline{u} + \eta\mathbf{W}\underline{X}. \tag{39}$$

The following proposition suggests that under certain conditions $\mathbf{S}$ is a contraction, implying the existence of a unique fixed point $\underline{X}_s^\star = \mathbf{S}(\underline{X}_s^\star)$.

Proposition 4. *$\mathbf{S}$ is a contraction on $l^2(\Psi)$ if $|\eta|\|\mathbf{W}\| < 1$, in which case its unique fixed point is given by* [2]

$$\underline{X}_s^\star = (1-\eta)(\mathbf{I}-\eta\mathbf{W})^{-1}\underline{u}. \tag{40}$$

Proof.

$$\|\mathbf{S}(\underline{X}_1) - \mathbf{S}(\underline{X}_2)\| = |\eta|\|\mathbf{W}(\underline{X}_1 - \underline{X}_2)\| \leq |\eta|\|\mathbf{W}\| \cdot \|(\underline{X}_1 - \underline{X}_2)\|. \tag{41}$$

It follows that $\mathbf{S}$ is a contraction if $|\eta|\|\mathbf{W}\| < 1$. Its fixed point $\underline{X}_s^\star = \mathbf{S}(\underline{X}_s^\star)$ must satisfy the equation

$$(1-\eta)\underline{u} + \eta\mathbf{W}\underline{X}_s^\star = \underline{X}_s^\star. \tag{42}$$

Solving for $\underline{X}_s^\star$ yields the result

$$\underline{X}_s^\star = (1-\eta)(\mathbf{I}-\eta\mathbf{W})^{-1}\underline{u}. \tag{43}$$

□

It is now interesting to consider the difference between $\underline{X}_s^\star$ and

$$\mathbf{S}(\underline{u}) = \eta\mathbf{W}\underline{u} + (1-\eta)\underline{u}. \tag{44}$$

Note that $\mathbf{S}(\underline{u})$ is the same as $\underline{Y}_m^\star$ defined previously in Eq. (30) if $\eta = \frac{\lambda}{1+\lambda}$.

Proposition 5.

$$\mathbf{S}(\underline{u}) - \underline{X}_s^\star = \eta\Big(\sum_{n=1}^{\infty}\eta^n\mathbf{W}^n\Big)(\mathbf{I}-\mathbf{W})\underline{u}. \tag{45}$$

Proof.

$$\begin{aligned}
\mathbf{S}(\underline{u}) - \underline{X}_s^\star &= \Big(\eta\mathbf{W}\underline{u} + (1-\eta)\underline{u}\Big) - (1-\eta)(\mathbf{I}-\eta\mathbf{W})^{-1}\underline{u} \\
&= \eta\mathbf{W}\underline{u} + (1-\eta)\underline{u} - (1-\eta)\underline{u} - (1-\eta)\Big(\sum_{n=1}^{\infty}\eta^n\mathbf{W}^n\Big)\underline{u} \\
&= \eta\mathbf{W}\underline{u} - \eta\mathbf{W}\underline{u} - \sum_{n=2}^{\infty}\eta^n\mathbf{W}^n\underline{u} + \sum_{n=1}^{\infty}\eta^{n+1}\mathbf{W}^n\underline{u} \\
&= -\sum_{n=2}^{\infty}\eta^n\mathbf{W}^n\underline{u} + \sum_{n=1}^{\infty}\eta^{n+1}\mathbf{W}^n\underline{u} \\
&= \Big(\sum_{n=1}^{\infty}\eta^{n+1}(-\mathbf{W}^{n+1} + \mathbf{W}^n)\Big)\underline{u} \\
&= \eta\Big(\sum_{n=1}^{\infty}\eta^n\mathbf{W}^n\Big)(\mathbf{I}-\mathbf{W})\underline{u}.
\end{aligned} \tag{46}$$

□

[2] Elements of $l^2(\Psi)$ should be represented in the lexicographic representation for this statement to be meaningful.

6. Computational considerations and results

In this section we present some computational results for our fractal-based regularizations as applied to the denoising and zooming problems outlined in Section 3.

In Figure 1 are shown some results obtained by the image denoising method (i) of Section 3.1 where the regularization function $\mathbf{\Gamma}$ is the squared fixed-point approximation error, $\| \underline{X} - \underline{u^\star} \|^2$, cf. Eq. (16).

The "input" or observed noisy image was the 256×256-pixel image $\underline{u}$, at the upper left. This noisy image was fractally coded using domain and range blocks of size 16×16 and 8×8 pixels, respectively. The attractor $\underline{u^\star}$ of the fractal transform is shown at the bottom right. Recall from Eq. (17) that the solution $\underline{Y^\star}$ to the minimization problem is given by a weighted average of $\underline{u}$ and $\underline{u^\star}$. Various weighted averages are shown in the figure, from $\lambda = 0$, corresponding to the original noisy image $\underline{u}$, to the limit $\lambda \to \infty$, corresponding to the fractal attractor $\underline{u^\star}$.

We have observed that, for various test images employed in our study, the denoising method (ii) using squared collage-error regularization function, $\mathbf{\Gamma}(\underline{X}) = \| \underline{X} - \mathbf{T}(\underline{X}) \|^2$, yields results $\underline{X^\star}$ (cf. Eq. (15)) that do not differ visually from $\underline{Y^\star}$. However, computing $\underline{X^\star}$ is much more computationally intensive.

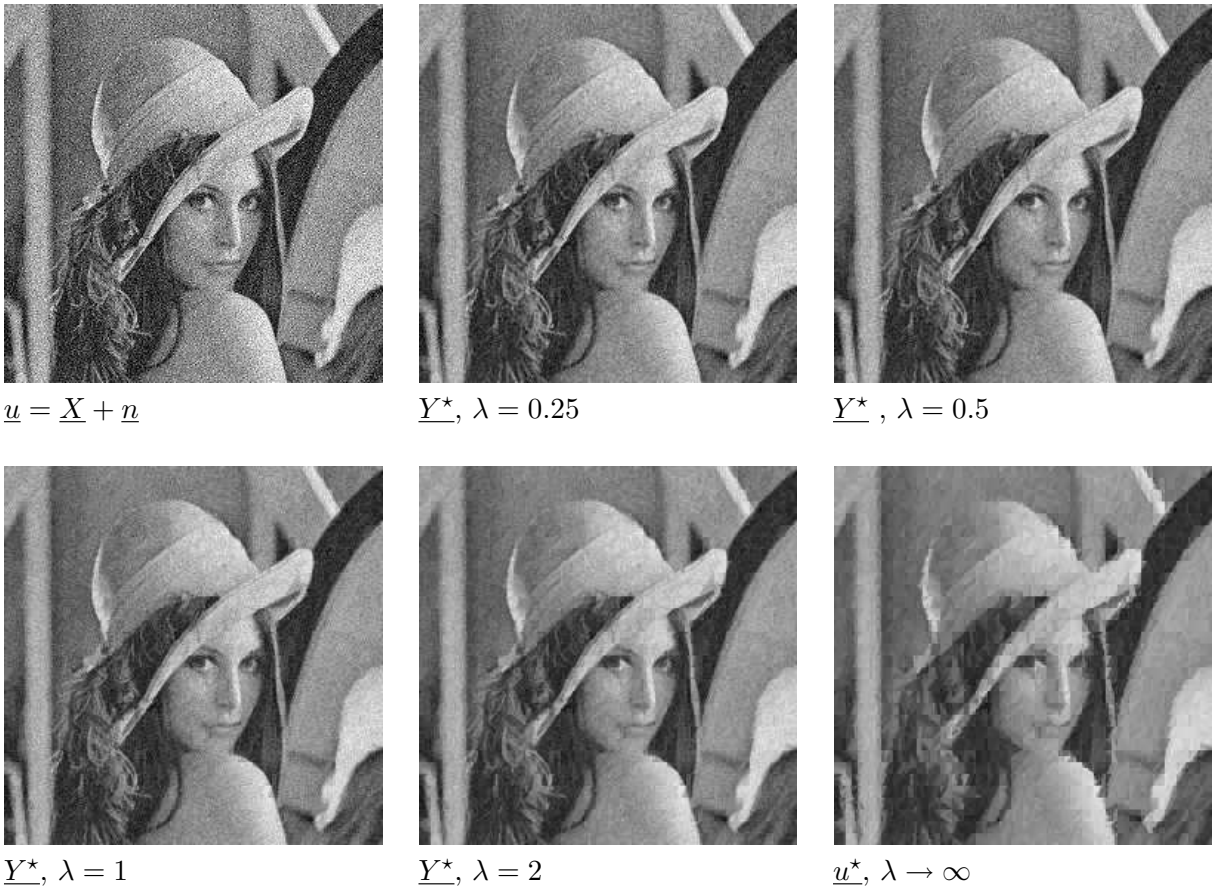

Figure 1. Image denoising with fractal-based regularization

In Figure 2 are presented some results obtained by the image zooming method of Section 3.2. The 128×128-pixel input image $\underline{u}$ is shown at the top. The goal was to zoom it by a factor of 2. The first column of this Figure shows the result obtained by simple pixel-replication, for which there is a great deal of visible blockiness. The input image $\underline{u}$ was then fractally coded

Figure 2. Image zooming with fractal-based regularization

using 8×8-pixel domain blocks and 4×4-pixel range blocks. The resulting fractal transform $\mathbf{T}$ was then applied an initial seed image of size 256×256 to obtain the 256×256-pixel attractor $\underline{u_z^\star}$. In the third column $\underline{Y_z^\star}$ is shown for the fixed value of $\lambda = 1$. We have used the conjugate gradient method to approximate $\underline{Y_z^\star}$ in this image.

Due to the fact that the matrix $\overline{W}$ in the NL-means regularization schemes outlined in Section 4 and 5 are non-sparse and very large, the computation of the $\underline{Y_m^\star}$ and $\underline{X_s^\star}$ was not feasible.

7. Conclusions

In this paper, we have introduced a set of regularization functions based on the self-similarity of images for the purpose of image denoising and image zooming. Thse regularization functions are based on existing schemes that exploit the self-similarity of images, namely fractal image coding and nonlocal-means image denoising. Some analytic asymptotic results have also been presented. Some computational results of image zooming and denoising using the fractal-based regularization schemes have also been presented.

By no means do we believe that the schemes introduced in this paper will replace existing efficient methods of denoising and zooming. Our method, however, introduces the possibility of using self-similarity-based priors in some imaging problems that may well be combined with other existing prior information about the image, possibly producing improved results. We also believe that the ideas introduced here could be further investigated for the case of more general degradation operators.

Appendix (A): Basics of fractal image encoding and decoding

More details on fractal image coding can be found in many places [3, 22, 17]. Fractal image coding seeks to approximate an image by a union of spatially-contracted and greyscale-modified copies of subblocks of itself.

If we let the image of interest be represented by an image function $u(x, y)$, denoted by $\underline{u}$ in the lexicographic representation, the result of the coding procedure is a contractive mapping $\mathbf{T}$, the so-called *fractal transform* operator. The fixed point $\underline{u^\star}$ of $\mathbf{T}$ provides an approximation to $\underline{u}$. In other words,

$$\underline{u} \cong \underline{u^\star} = \mathbf{T}\underline{u^\star}. \tag{47}$$

To obtain $\mathbf{T}$, the image is first partitioned (e.g., uniform, quadtree) into a set of nonoverlapping range blocks C_i. For each range block C_i, one searches for a larger domain block $P_{J(i)}$ (from an appropriate "domain pool" $\mathcal{P}$ that is often common to all range blocks of the same size) such that $\underline{u}\Big|_{C_i}$ (by this notation we mean, the block C_i of $\underline{u}$) is approximated by a modified copy of $\underline{u}\Big|_{P_{J(i)}}$, i.e.,

$$\underline{u}\Big|_{C_i} \cong \phi_i\Big(\underline{u}\Big|_{P_{J(i)}}\Big) = \phi_i\Big(\underline{u}\Big|_{w_i^{-1}(C_i)}\Big), \tag{48}$$

where $\phi_i : \mathbf{R} \to \mathbf{R}$ is a greyscale map that operates on pixel intensities and w_i denotes the 1-1 contraction/decimation that maps pixels of $P_{J(i)}$ onto pixels of C_i. The *fractal code* defining $\mathbf{T}$ consists of the maps ϕ_i as well as the domain-range assignments determined during the coding procedure. In practice, greyscale maps are assumed to be affine, i.e.,

$$\phi_i(t) = \alpha_i t + \beta_i. \tag{49}$$

For a given domain-range block pair $P_{J(i)}/C_i$, the optimal value of the α and β parameters is usually accomplished by means of least-squares fitting, given $\underline{u}$, for each i

$$\min_{J(i),\alpha_i,\beta_i} \left\| \underline{u}\Big|_{C_i} - \Big\{\alpha_i \mathcal{D}\Big(\underline{u}\Big|_{P_{J(i)}}\Big) + \beta_i\Big\} \right\|_2 \tag{50}$$

At the decoding stage, given a contractive fractal transform $\mathbf{T}$, we may generate its fixed point $\underline{u}^\star$ by iteration $\underline{u_{n+1}} = \mathbf{T}(\underline{u_n})$, starting with an arbitrary image $\underline{u_0}$.

$$\underline{u_{n+1}}\Big|_{C_i} = \Big(\mathbf{T}\underline{u_n}\Big)\Big|_{C_i} = \alpha_i \mathcal{D}\Big(\underline{u_n}\Big|_{P_{J(i)}}\Big) + \beta_i. \tag{51}$$

in which $\mathcal{D}$ in this equation is the down-sampling operator. Banach's contraction mapping theorem guarantees that the sequence of images $\underline{u_n}$ converges to $\underline{u}^\star$, if $|\alpha_i| < 1$ for any i. Some recent investigations [1, 2] have shown that images generally possess a great deal of local (affine) self-similarity: Given a subimage $\underline{u}|_{R_i}$ there are often a good number of domain blocks D_j whose subimages $\underline{u}|_{D_j}$ approximate it as well as the "best" domain block. This feature, which never seems to have been quantified previously, accounts for the rather small degradations that are experienced when the size of the *domain pools* – the domain blocks D_j to be examined – is decreased.

Appendix (B): Non-local-means image denoising [5, 6]
Consider the following image denoising problem

$$u = X + n$$

For any $(x, y) \in \Omega$ define the denoising formula

$$NL\big[u(x,y)\big] = \frac{1}{C(x,y)} \sum_{(p,q)\in\Omega} w(x,y,p,q)u(p,q), \qquad \text{such that} \tag{52}$$

$$w(x,y,p,q) = \exp\Big(-\frac{\parallel u(\mathcal{N}^d\{(x,y)\}) - u(\mathcal{N}^d\{(p,q)\}) \parallel_{2,a}^2}{h^2}\Big), \text{ and}$$
$$C(x,y) = \sum_{(p,q)\in\Omega} w(x,y,p,q) \tag{53}$$

where the expressions $\mathcal{N}^d\{\dots\}$ and $\parallel . \parallel_{2,a}$ are defined in the following way.

Neighborhoods: For any point in the domain of observation $(x, y) \in \Omega$, define

$$\mathcal{N}^d\big\{(x,y)\big\} = \Big\{(x+i, y+j) \mid (i,j) \in \mathbf{Z}^2, \max\{|i|,|j|\} \le d\Big\}. \tag{54}$$

Gaussian-semi-norm: For an image I the Gaussian-weighted-semi-norm $\parallel . \parallel_{2,a}$ is defined in terms of the l^2 norm as

$$\parallel I \parallel_{2,a} = \parallel G_a \star I \parallel_2 \tag{55}$$

in which G_a is a two-dimensional Gaussian kernel of standard deviation a.

The idea of the NL-means algorithm is that given a discrete noisy image u, the estimated denoised value $NL\big(u(x,y)\big)$ is computed as a weighted average of all the pixels in the image, $u(p, q)$, where the weights $w(x, y, p, q)$ depend on the similarity of neighborhoods of the pixels x and y, and w is a decreasing function of the weighted Euclidean distance of the neighborhoods. The parameter h characterizes the degree of filtering. It controls the decay of the exponential function and therefore the decay of the weights as a function of the Euclidean distances. The NL-means algorithm not only compares the grey level in a single point but the geometrical configuration in a whole neighborhood [5, 6]. When u is represented in the lexicographic order as $\underline{u}$, the output of the NL-means denoising can be written as $\mathbf{W}\underline{u}$, where $\mathbf{W}$ is a right stochastic matrix containing the associated non-negative weights.

Acknowledgments
The authors acknowledge MITACS for a graduate student conference travel support awarded to ME to present this work at AIP 2007. ME also gratefully acknowledges support from an Ontario Graduate Scholarship as well as from the University of Waterloo. ERV gratefully acknowledges research support from a Natural Sciences and Engineering Research Council (NSERC) Discovery Grant.

References

[1] Alexander S K 2005 *Multiscale methods in image modelling and image processing, Ph.D. Thesis* (Dept. of Applied Mathematics, University of Waterloo)
[2] Alexander S K, Kovacic S and Vrscay E R 2007 A model for image self-similarity and the possible use of mutual information *Proc. of EUSIPCO* 975-979
[3] Barnsley M F 1988 *Fractals Everywhere* (New York: Academic Press)
[4] Barnsley M F and Demko S 1985 *Proc. Roy. Soc. London* Iterated function systems and the global construction of fractals **A399** 243-275
[5] Buades A, Coll B and Morel J M 2005 *IEEE Int. Conf. on Comp. Vision and Patt. Recog. (CVPR) San-Diego California* A nonlocal algorithm for image denoising **2** 6065
[6] Buades A, Coll B and Morel J M 2005 *SIAM Journal on Multiscale Modeling and Simulation (MMS)* A review of image denoising algorithms, with a new one **4(2)** 490-530
[7] Criminisi A, Perez P and Toyama K 2004 *IEEE Trans. on Image Proc.* Region filling and object removal by exemplar-based image inpainting **13(9)** 1200-1212
[8] Ebrahimi M and Vrscay E R 2006 *Lecture Notes in Computer Science, Volume 4141, Book: Image analysis and Recognition* Fractal image coding as projections onto convex sets (Berlin/Heidelberg: Springer) p 493-506
[9] Ebrahimi M and Vrscay E R 2006 *Proc. CCECE '06, Ottawa, Canada* Regularized fractal image decoding p 1933-1938
[10] Efros A A and Leung T K 1999 *IEEE Int. Conf. on Computer Vision (ICCV), Corfu, Greece* Texture synthesis by non-parametric sampling p 1033-1038
[11] Elad M and Datsenko D 2007 *The Computer Journal* Example-based regularization deployed to super-resolution reconstruction of a Single Image **50(4)** 1-16
[12] Fisher Y 1995 *Fractal Image Compression, Theory and Application* (New York: Springer-Verlag)
[13] Forte B and Vrscay E R 1998 *Theory of generalized fractal transforms in Fractal Image Encoding and Analysism Y. Fisher, Ed.* (New York: Springer-Verlag)
[14] Freeman W T, Pasztor E C and Carmichael O T 2000 *Int. Journal Of Computer Vision* Learning low-level vision **40(1)** 25-47
[15] Freeman W T, Jones T R and Pasztor E C 2002 *IEEE Comp. Graphics And Appl.* Example-based super-resolution **22(2)** 56-65
[16] Gharavi-Al. M, DeNardo R, Tenda Y and Huang T S 1997 *Proc. of SPIE Visual Commun. and Image Proc., San Jose, CA, USA* Resolution enhancement of images using fractal coding **3024** 1089-1100
[17] Ghazel M, Freeman G and Vrscay E R 2003 *IEEE Trans. Image Proc.* Fractal image denoising **12(12)** 1560-1578
[18] Ghazel M, Freeman G and Vrscay E R 2006 *IEEE Trans. Image Proc.* Fractal-wavelet image denoising revisited **15** 2669-2675
[19] Haber E and Tenorio L 2003 *Inverse Problems* Learning regularization functionals **19** 611-626
[20] Ho H and Cham W 1997 *IEEE Int. Conf. on Acoustics, Speech, and Signal Proc. (ICASSP'97)* Attractor image coding using lapped partitioned iterated function systems **4** p 2917
[21] Latouche G and Ramaswami V 1999 *Introduction to Matrix Analytic Methods in Stochastic Modelling* (Philadelphia: ASA and SIAM)
[22] Lu N 1997 *Fractal Imaging* (NY: Academic Press)
[23] Polidori E and Dugelay J L 1997 *Fractals (Suppl. Issue)* Zooming using iterated function systems **5** 111-123
[24] Tikhonov A N and Arsenin V A 1977 *Solution of Ill-posed Problems* (Washington: Winston & Sons)
[25] Wei L Y and Levoy M 2000 *Proc. of SIGGRAPH, New Oleans, Louisiana* Fast texture synthesis using tree-structured vector quantization p 479-488
[26] Zhu S C and Mumford D 1997 *IEEE Trans. on Patt. Analysis and Machine Intel.* Prior learning and Gibbs reaction-diffusion **19(11)** 1236-1250