Theory and Computation of Variational Image Deblurring

Tony F. Chan

Department of Mathematics, UCLA
Los Angeles, CA 90095, USA
E-mail: TonyC@college.ucla.edu

Jianhong Shen

School of Mathematics
University of Minnesota
Minneapolis, MN 55455, USA
E-mail: jhshen@math.umn.edu

To recover a sharp image from its blurry observation is the problem known as *image deblurring*. It frequently arises in imaging sciences and technologies, including optical, medical, and astronomical applications, and is crucial for allowing to detect important features and patterns such as those of a distant planet or some microscopic tissue.

Mathematically, image deblurring is intimately connected to backward diffusion processes (e.g., inverting the heat equation), which are notoriously unstable. As inverse problem solvers, deblurring models therefore crucially depend upon proper regularizers or conditioners that help secure stability, often at the necessary cost of losing certain high-frequency details in the original images. Such regularization techniques can ensure the existence, uniqueness, or stability of deblurred images.

The present work follows closely the general framework described in our recent monograph [18], but also contains more updated views and approaches to image deblurring, including, e.g., more discussion on stochastic signals, the Bayesian/Tikhonov approach to Wiener filtering, and the iterated-shrinkage algorithm of Daubechies et al. [30,31] for wavelet-based deblurring. The work thus contributes to the development of generic, systematic, and unified frameworks in contemporary image processing.

1. Mathematical Models of Blurs

Throughout the current work, an image u is identified with a Lebesgue measurable real function on an open two-dimensional (2D) regular domain Ω. A general point $\boldsymbol{x} = (x_1, x_2) \in \Omega$ shall also be called a *pixel* as in digital image processing. The framework herein applies readily to color images for which u could be considered an RGB-vectorial function.

1.1. *Linear Blurs*

Deblurring is to undo the blurring process applied to a sharp and clear image earlier, and is thus an inverse problem. We hence start with the description of the forward problem - mathematical models of blurring.

In most applications, blurs are introduced by three different types of physical factors: optical, mechanical, or medium-induced, which could lead to familiar out-of-focus blurs, motion blurs, or atmospheric blurs respectively. We refer the reader to [18] for a more detailed account on the associated physical processes. Figures 1 and 2 show two real blur examples directly taken by a digital camera under different circumstances.

Fig. 1. A real example of an out-of-focus blur. Left: the clear image; Right: the out-of-focus image taken by a digital camera that focuses on a point closer than the scene.

Mathematically, blurring can be either *linear* or *nonlinear*. The latter is more challenging to invert due to the scarcity of proper nonlinear models. The current work shall mainly focus on linear deblurring problems.

A general linear blur $u_0 = K[u]$ is defined by a linear operator K. In most applications noise is unavoidable and a real observation is thus often modelled by

$$u_0 = K[u] + n,$$

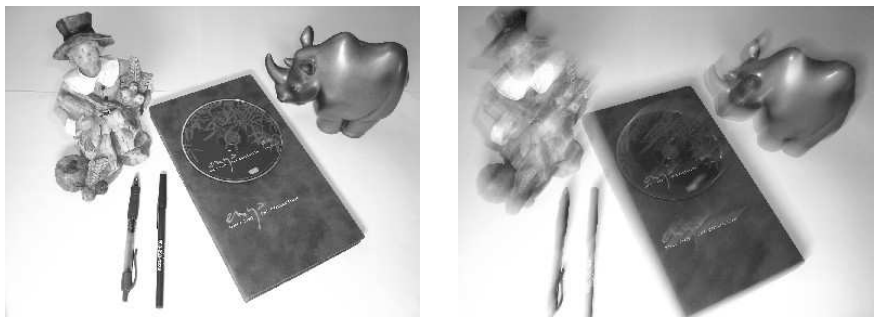

Fig. 2. A real example of a motion blur. Left: the clear image; Right: the motion-blurred image taken by a camera that experiences a rapid jitter during the exposure.

provided that the noise n is additive. (Multiplicative noises can be handled similarly.)

Among all linear blurs, the most frequently encountered type is *shift-invariant*. A linear blur K is said to be shift-invariant if for any shift $\boldsymbol{a} \in \mathbb{R}^2$,

$$u_0(\boldsymbol{x}) = K[u(\boldsymbol{x})] \quad \text{implies that} \quad u_0(\boldsymbol{x} - \boldsymbol{a}) = K[u(\boldsymbol{x} - \boldsymbol{a})].$$

It is well known in signal processing as well as system theory [56] that a shift-invariant linear operator must be in the form of convolution:

$$K[u] = k * u(\boldsymbol{x}) = \int_{\mathbb{R}^2} k(\boldsymbol{x} - \boldsymbol{y}) u(\boldsymbol{y}) d\boldsymbol{y}, \tag{1}$$

for some suitable kernel function $k(\boldsymbol{x})$, or the *point spread function* (PSF).

At any fixed pixel $\boldsymbol{x} \in \Omega$, a general linear blur K induces a linear functional on u, or a generalized function $L_{\boldsymbol{x}} : u \to K[u](\boldsymbol{x})$. Denote it symbolically by $k(\boldsymbol{x}, \cdot)$ so that as in distribution theory [68], one has

$$L_{\boldsymbol{x}}[u] = \langle k(\boldsymbol{x}, \cdot), u(\cdot) \rangle.$$

Suppose that the distribution $k(\boldsymbol{x}, \cdot)$ is actually an ordinary measurable function in $L^1(\Omega)$. Then the linear blur becomes ordinary integrals:

$$u_0(\boldsymbol{x}) = \int_{\Omega} k(\boldsymbol{x}, \boldsymbol{y}) u(\boldsymbol{y}) d\boldsymbol{y}, \qquad \boldsymbol{x} \in \Omega.$$

Herein we shall assume that the image u belongs to $L^p(\Omega)$ with $p \in [1, +\infty]$, and that K is a bounded linear operator from $L^p(\Omega)$ to $L^q(\Omega)$ with some $q \in [1, +\infty]$. As a result, the adjoint K^* is defined from $(L^q)^*$ to $(L^p)^*$, the dual spaces. (One must be aware, however, that $(L^\infty)^* \neq L^1$ [48].)

1.2. *The DC-Condition*

The most outstanding characteristic of a blur operator is the DC-condition:

$$K[1] = 1, \qquad \text{treating} \quad 1 \in L^\infty(\Omega). \tag{2}$$

In classical signal processing [56], DC stands for *direct current* since the Fourier transform of a constant contains no oscillatory frequencies. By duality, $\langle K[u], v\rangle = \langle u, K^*[v]\rangle$, and the DC-condition on K amounts to the *mean-preserving condition* on K^*:

$$\langle K^*[v]\rangle = \langle v\rangle, \quad \text{by setting } u = 1; \quad \text{or} \quad \int_\Omega K^*[v](\boldsymbol{x})d\boldsymbol{x} = \int_\Omega v(\boldsymbol{x})d\boldsymbol{x}, \tag{3}$$

if both v and $K^*[v]$ belong to $L^1(\Omega)$.

In terms of information theory [27], the DC condition implies that constant signals are *invariant* under blurring. In particular, blurs cannot *generate* ripples from flat signals, and thus can never *create* information.

When the blur is shift-invariant with a PSF k, the DC-condition requires

$$\int_{\mathbb{R}^2} k(\boldsymbol{x})d\boldsymbol{x} = 1, \quad \text{or in terms of its Fourier transform,} \quad K(\boldsymbol{\omega} = 0) = 1,$$

since the adjoint is also shift-invariant with PSF $k(-x)$. Moreover, a more convincing blur operator has to be *lowpass* [56,67], i.e., $K(\boldsymbol{\omega})$ must decay rapidly at high frequencies.

1.3. *Nonlinear Blurs*

Blurs could be nonlinear, though linear models prevail in the literature. Consider for example the following nonlinear diffusion model:

$$v_t = \nabla \cdot \left[\frac{1}{\sqrt{1 + |\nabla v|^2}}\nabla v\right], \qquad v\big|_{t=0} = u(\boldsymbol{x}). \tag{4}$$

Let the solution be denoted by $v(\boldsymbol{x}, t)$. For any fixed finite time $T > 0$, define a nonlinear operator $K = K_T$ by: $u_0 = K[u] = v(\boldsymbol{x}, T)$. Nonlinearity is evident since for example $K[\lambda u] \neq \lambda K[u]$ for general u and $\lambda \neq 0$. But the operator K apparently satisfies the DC-condition. Furthermore, (4) is the gradient descent equation of the minimum surface energy

$$E[v] = \int_{\mathbb{R}^2} \sqrt{1 + |\nabla v|^2}d\boldsymbol{x}.$$

As a result, the above nonlinear diffusion model always smoothens out any rough initial surfaces. In particular, small scale features and oscillations of u

must be wiped out in $u_0 = K[u]$, making u_0 a visually blurred and mollified version of the original image u. Notice remarkably that the nonlinear blur is in fact shift-invariant.

2. Illposedness of Deblurring

The illposedness of deblurring could be readily understood in four intriguing aspects. Understanding the root and nature of illposedness helps one design good deblurring models. The following four viewpoints are in some sense the four different facets of a same phenomenon, and hence must not be taken individually.

A. Deblurring is Inverting Lowpass Filtering. In the Fourier domain, a blur operator is often *lowpass* so that high frequency details are compressed by vanishing multipliers. As a result, to deblur a blurry image, one has to multiply approximately the reciprocals of the vanishing multipliers, which is conceivably unstable to noises or other high-frequency perturbations in the image data.

B. Deblurring is Backward Diffusion. By the canonical PDE theory, to blur an image with a Gaussian kernel amounts to running the heat diffusion equation for some finite duration with the given image as the initial data. Therefore, to deblur is naturally equivalent to inverting the diffusion process, which is notoriously unstable.

Stochastically, diffusion corresponds to the Brownian motions of an initial ensemble of particles. Thus to deblur or to de-diffuse amounts to reversing an irreversible random spreading process, which is physically illposed.

C. Deblurring is Entropy Decreasing. The goal of deblurring is to reconstruct the detailed image features from a mollified blurry image. Thus from the standpoint of statistical mechanics, deblurring is a process to increase (Shannon) information, or equivalently, to *decrease entropy*. According to the second law of statistical mechanics [41], deblurring thus could never occur naturally and extra work has to be done to the system.

D. Deblurring is Inverting Compact Operators. In terms of abstract functional analysis, a blurring process is typically a *compact* operator. A compact operator is one that maps any bounded set (according to the

associated Hilbert or Banach norms) to a much better behaved set which is precompact. To achieve this goal, intuitively speaking, a compact operator has to mix spatial information or introduce many coherent structures, which is often realized essentially by dimensionality reduction based on vanishing eigenvalues or singular values. Therefore to invert a compact operator is again equivalent to de-correlating spatial coherence or reconstructing the formerly suppressed dimensions (during the blurring process) of features and information, which is unstable.

This illustration can be further vivified via finite-dimensional linear algebra [65,66]. Looking for an unknown vector $\boldsymbol{u}$ of dimension much *higher* than its observation $\boldsymbol{b}$ for the matrix-vector equation $A\boldsymbol{u} = \boldsymbol{b}$ often has either no solution or infinitely many. Any *unique* meaningful solution has to be defined in some proper way.

3. Tikhonov and Bayesian Regularization

From the above discussion, proper regularization techniques have to be sought after in order to alleviate the illposedness of the deblurring process.

Two universal regularization approaches, which are essentially reciprocal in the two dual worlds of deterministic and stochastic methodologies, are Tikhonov regularization [69] and the Bayesian inference theory [45]. Their intimate connection has been explained in, for example, Mumford [53], and Chan, Shen, and Vese [20].

In essence, both approaches introduce some *prior* knowledge about the target images u to be reconstructed. In the Bayesian framework, it is to introduce some proper probability distribution over all possible image candidates, and necessary bias (i.e., regularization) is encouraged to favor more likely ones. In the Tikhonov setting, the prior knowledge is often reflected through some properly designed "energy" formulations, e.g., a quadratic energy like $a\|u\|^2$ under some proper functional norm.

We now introduce the most general framework of Baysian-Tikhonov regularization for deblurring. Consider the blur model

$$u_0(x) = K[u](x) + n(x), \quad x = (x_1, x_2) \in \mathbb{R}^2,$$

with a general blur operator K and additive white noise n.

First, assume that blur process K is either known explicitly or estimated in advance [18]. As an estimation problem, deblurring can be carried out by the Bayesian principle or MAP (maximum a posteriori probability):

$$\hat{u} = \operatorname{argmax} \operatorname{Prob}(u \mid u_0, K),$$

or equivalently, in terms of the logarithmic likelihood or Gibbs' ensemble formula $E[\cdot] = -\log p(\cdot) +$ *a constant or fixed free energy* [24],

$$\hat{u} = \operatorname{argmin} E[u \mid u_0, K].$$

The Bayesian formula with a *known* blur K is given by

$$\text{Prob}(u \mid u_0, K) = \text{Prob}(u \mid K)\text{Prob}(u_0 \mid u, K)/\text{Prob}(u_0 \mid K).$$

Given an image observation u_0, the denominator is simply a fixed probability normalization constant. Thus effectively one seeks an estimator $\hat{u}$ to minimize the product of the *prior* model $\text{Prob}(u \mid K)$ and the data (or fidelity) model $\text{Prob}(u_0 \mid u, K)$. Since ideal images and blurs are often independent, one has $\text{Prob}(u \mid K) = \text{Prob}(u)$. Therefore in terms of the energy formulation, one attempts to minimize the *posterior energy*

$$E[u \mid u_0, K] = E[u] + E[u_0 \mid u, K]. \tag{5}$$

In the setting of Tikhonov regularization, the prior energy $E[u]$ is virtually a regularizer for the data fitting model $E[u_0 \mid u, K]$. Functionally, $E[u]$ can be specified by a suitable norm or semi-norm in some proper function space such as the BV or Besov spaces, which will be discussed later.

For *blind* deblurring when the kernel K is unknown, the Bayesian formula becomes

$$\max_{u,K} \text{Prob}(u, K \mid u_0) = \text{Prob}(u_0 \mid u, K)\text{Prob}(u, K)/p(u_0).$$

In most applications, the blur mechanism K is uncorrelated to the image content u (e.g., in astronomical imaging, atmospheric turbulence activities K are not influenced by the ideal image observation u of the stars and galaxies many lightyears away). Then one has

$$\text{Prob}(u, K) = \text{Prob}(u)\text{Prob}(K),$$

and the posterior energy takes the form of

$$E[u, K \mid u_0] = E[u_0 \mid u, K] + E[u] + E[K], \tag{6}$$

up to a fixed additive constant (corresponding to the free energy under the given parameters in the models).

In both models (5) and (6) for non-blind and blind deblurring, the data generation model $E[u_0 \mid u, K]$ is often readily expressible via squared fitting error for Gaussian white noise. Thus the key to effective deblurring relies upon the proper proposals on the prior knowledge for the target image u, as well as the blur process K in the blind scenario.

4. Optimal Wiener Filtering for Non-Blind Deblurring

From now on, instead of the black-faced symbols $\boldsymbol{x}$ and $\boldsymbol{\omega}$, a general pixel will be denoted by $x = (x_1, x_2)$ and its frequency dual variable by $\omega = (\omega_1, \omega_2)$. Due to the stochastic nature of Wiener filtering, we shall begin with a brief introduction to 2-D stochastic signals.

4.1. *2-D Stochastic Spatial Signals*

Consider only *real* stochastic images defined on the domains of either $\mathbb{R}^2$ for analog images or the lattice $\mathbb{Z}^2$ for digital ones.

A stochastic image $\boldsymbol{u}(x)$ is said to be *homogeneous* if any of its finite marginal distributions carries no spatial memory, or equivalently, is translation invariant:

$$P_{x+z,\cdots,y+z}(u,\cdots,v) \equiv P_{x,\cdots,y}(u,\cdots,v), \quad \forall z = (z_1, z_2),$$

where the marginal probability is defined by

$$P_{x,\cdots,y}(u,\cdots,v)du\cdots dv = \text{Prob}(\boldsymbol{u}(x) \in [u, u+du], \cdots, \boldsymbol{u}(y) \in [v, v+dv]).$$

Familiar sources for homogenous images include Gibbs' random fields with translation invariant potentials, or Markov random fields with translation invariant graph structures and local conditionals [6,18,40]. Homogeneity is appropriate for modelling certain ideal single-species textures such as sandy beaches or grasslands, which are more or less uniform.

More generally, a stochastic signal $\boldsymbol{u}$ is said to be wide-sense homogeneous (WSH), if its two-point auto-correlation function

$$R_{\boldsymbol{uu}}(x, y) = \text{E}[\boldsymbol{u}(x)\boldsymbol{u}(y)],$$

is translation invariant: for any relocation z,

$$R_{\boldsymbol{uu}}(x+z, y+z) = R_{\boldsymbol{uu}}(x, y).$$

Thus if $\boldsymbol{u}$ is WSH, its auto-correlation function is essentially a single-pixel function: $R_{\boldsymbol{uu}}(x-y) = R_{\boldsymbol{uu}}(x, y)$. Let $\omega = (\omega_1, \omega_2)$ denote the spatial frequency variable. Then the power spectral density $S_{\boldsymbol{uu}}(\omega)$ is defined to be the Fourier transform of $R_{\boldsymbol{uu}}(x)$:

$$S_{\boldsymbol{uu}}(\omega) = \int_{\mathbb{R}^2} R_{\boldsymbol{uu}}(x) e^{-ix\omega} dx.$$

A WSH image $\boldsymbol{n}(x)$ is said to be *white noise*, if $S_{\boldsymbol{nn}}(\omega) \equiv \sigma^2$, or equivalently, its auto-correlation function $R_{\boldsymbol{nn}}(x)$ is a constant multiple of Dirac's delta signal $\delta(x)$.

Two WSH images $\boldsymbol{u}$ and $\boldsymbol{v}$ are said to be *cross*-WSH if their cross-correlation function is translation invariant as well.

$$R_{\boldsymbol{uv}}(x,y) = \mathrm{E}[\boldsymbol{u}(x)\boldsymbol{v}(y)] = R_{\boldsymbol{uv}}(x-y). \tag{7}$$

Define the cross-WSH set of a given WSH image $\boldsymbol{u}$ to be

$$\Lambda_{\boldsymbol{u}} = \{\boldsymbol{v} \mid \boldsymbol{v} \text{ is cross-WSH to } \boldsymbol{u}\ \}. \tag{8}$$

Then $\boldsymbol{u} \in \Lambda_{\boldsymbol{u}}$. Furthermore, we have the following list of straightforward but useful properties.

Theorem 1: Suppose $\boldsymbol{u}$ is WSH and $\Lambda_{\boldsymbol{u}}$ its cross-WSH set. Then $\Lambda_{\boldsymbol{u}}$ is a linear space which is closed under *spatial relocation*:

$$\boldsymbol{v}(\cdot) \in \Lambda_{\boldsymbol{u}} \Rightarrow \boldsymbol{v}(\cdot + z) \in \Lambda_{\boldsymbol{u}},$$

for any relocation $z \in \mathbb{R}^2$, as well as under *linear filtering*:

$$\boldsymbol{v} \in \Lambda_{\boldsymbol{u}} \Rightarrow h * \boldsymbol{v} \in \Lambda_{\boldsymbol{u}},$$

for any filter $h = h(x)$. Let $H(\omega)$ denote the impulse response of h. Then,

$$R_{h*\boldsymbol{v},\boldsymbol{u}}(x) = h * R_{\boldsymbol{vu}}(x), \quad \text{and} \quad S_{h*\boldsymbol{v},\boldsymbol{u}}(\omega) = H(\omega)S_{\boldsymbol{vu}}(\omega).$$

4.2. *Stochastic Signals as Random Generalized Functions*

Another intriguing approach to stochastic signals is to treat a stochastic signal as a *random* generalized function.

Recall that a generalized function F, or a Schwartz distribution, is a linear functional on the test function space $\mathcal{D} = C_0^\infty(\mathbb{R}^2)$, so that for any test function $\phi \in \mathcal{D}$,

$$\text{the determinisitc values } \langle F, \phi\rangle \text{ are linear in } \phi.$$

A 2D stochastic field $\boldsymbol{u}$ on $\mathbb{R}^2$ can be treated as a *random* generalized function so that for any test function $\phi \in \mathcal{D}$, the value

$$U_\phi = \langle \boldsymbol{u}, \phi\rangle \text{ is a random variable, and } U_{a\phi+b\psi} = aU_\phi + bU_\psi.$$

The mean field of $\boldsymbol{u}$ is an *ordinary* generalized function $m_{\boldsymbol{u}}$ such that

$$\mathrm{E}(U_\phi) = \langle m_{\boldsymbol{u}}, \phi\rangle, \qquad \text{for any } \phi \in \mathcal{D}.$$

If $m_{\boldsymbol{u}} = 0$, $\boldsymbol{u}$ is said to have zero means. Two random fields $\boldsymbol{u}$ and $\boldsymbol{v}$ are said to be *equivalent* if U_ϕ and V_ϕ share the same probability distribution for any test function $\phi \in \mathcal{D}$.

For any shifting operator $S_z : \phi(x) \to \phi(x-z)$ with $z \in \mathbb{R}^2$, the shifted random field $S_z\boldsymbol{u}$ is defined by the dual formula

$$\langle S_z\boldsymbol{u}, \phi\rangle = \langle \boldsymbol{u}, S_{-z}\phi\rangle, \qquad \phi \in \mathcal{D}.$$

$\boldsymbol{u}$ is said to be *shift invariant* if $\boldsymbol{u}$ is equivalent to $S_z\boldsymbol{u}$ for any shift z. If a test function ϕ is interpreted as a measurement sensor, then a random field is shift invariant if and only if *no* statistical difference can be detected from the measurements when a sensor is moved from one location to another.

A random field $\boldsymbol{u}$ of zero means is said to be wide-sense homogeneous (WSH) if there exists some locally integrable function $R_{\boldsymbol{uu}}(x)$ such that for any two test functions ϕ and ψ, one has

$$\mathrm{E}(U_\phi U_\psi) = \langle \phi, R_{\boldsymbol{uu}} * \psi\rangle.$$

Similarly, two random fields $\boldsymbol{u}$ and $\boldsymbol{v}$ of zero means are said to be cross-WSH if there exists some locally integrable function $R_{\boldsymbol{uv}}(x)$ such that for any test functions ϕ and ψ,

$$\mathrm{E}(U_\phi V_\psi) = \langle \phi, R_{\boldsymbol{uv}} * \psi\rangle.$$

$R_{\boldsymbol{uu}}$ and $R_{\boldsymbol{uv}}$ are called the auto-correlation and cross-correlation functions, and are apparently unique if in existence. If one *formerly* takes Dirac's delta functions as test functions, it is easy to verify the consistency between the current functional definitions and the pointwise definitions in the preceding subsection.

The reader can familiarize the above theory with the help of the following example of *random* harmonic waves in 1D:

$$\boldsymbol{u}(x) = A\cos(x+B), \qquad x \in \mathbb{R},$$

where A and B are independent random variables with B *uniformly* distributed over $[0, 2\pi)$, and A *exponentially* distributed on $(0, \infty)$. Then it is easy to show, for example, that $\boldsymbol{u}$ must be homogenous (or shift invariant).

4.3. *Filtering-Based Deblurring*

Assume that the blur is shift invariant: $u_0 = k * u + n$. Filtering-based deblurring is to produce an estimator $\hat{u}$ of the ideal image u via a linear filtering scheme:

$$\hat{u} = \hat{u}_w = w * u_0, \quad \text{with a suitable filter } w.$$

Without noise, the ideal filter would be directly given by $W(\omega) = \frac{1}{K(\omega)}$, in the Fourier domain, so that $\hat{u} = w * u_0 \equiv u$ for any clear image u and

perfect reconstruction is reached! However, it is a rather unnerving formula since a typical blur k is often *lowpass* and $K(\omega)$ decays rapidly at high frequencies. Such a naive filter therefore exaggerates any high-frequency errors or perturbations.

To alleviate such unwanted instability, in the noise-free case one rewrites the naive filter $W = 1/K$ to

$$W(\omega) = \frac{K^*(\omega)}{K(\omega)K^*(\omega)} = \frac{K^*}{|K|^2}, \quad \text{where } ^* \text{ denotes complex conjugacy.}$$

The vanishing denominator at high frequencies can be guarded away from zero by incorporating a positive factor $r = r(\omega)$:

$$W \to W_r = \frac{K^*}{|K|^2 + r}. \tag{9}$$

The resultant deblurred image $\hat{u}_r$ is then given by

$$\hat{u}_r = w_r * k * u,$$

or in the Fourier domain, the composite effect of the blurring and deblurring processes is achieved by the multiplier

$$W_r(\omega)K(\omega) = \frac{|K(\omega)|^2}{|K(\omega)|^2 + r(\omega)}. \tag{10}$$

The restoration indeed well approximates the identity operator on low frequencies where $r \ll |K|^2$ since K is lowpass. High frequencies are however suppressed since K almost vanishes and $|K|^2 \ll r$. Thus the regularizer r plays a soft cutoff role.

The reader should pay constant attention to the frequent emergency of such an r-factor henceforth. It embodies a universal quantity that is critical for any deblurring problem.

The question is how to choose wisely an optimal regularizer r. A uniform constant is a reasonable guess but lacks clear theoretical backup. What Wiener discovered was that r should be related to the signal-to-noise ratio in the observation u_0, which will be explained in the next subsection.

It is also the right spot to reiterate the earlier analogy drawn from the finite linear algebra of solving $A\boldsymbol{u} = \boldsymbol{b}$. Recall that the least square solution [65,66] is given by the *normal equation*:

$$A^*A\hat{\boldsymbol{u}} = A^*\boldsymbol{b}, \text{ or } \hat{\boldsymbol{u}} = (A^*A)^{-1}A^*A\boldsymbol{u}.$$

Thus the "filter" in the least-square solution is given by

$$W = (A^*A)^{-1}A^*.$$

When the linear operator A "mixes" information too intensely so that

$$\text{rank}\,(A) = \dim(\text{Range}(A)) \ll \#\text{ of rows of } A,$$

A^*A becomes singular (or almost singular if the approximate rank only counts the nonnegligible singular values of A). Then the inversion of $(A^*A)^{-1}$ still remains illposed or unstable. In linear algebra, the filter is then regularized to

$$W_r = (A^*A + rI)^{-1}A^* \tag{11}$$

for some positive small parameter $r > 0$, where I is the identity matrix. The resultant estimator corresponds to the regularized least-square problem:

$$\hat{\boldsymbol{u}}_r = \text{argmin}_{\boldsymbol{u}} \|A\boldsymbol{u} - \boldsymbol{b}\|^2 + r\|\boldsymbol{u}\|^2.$$

Notice the characteristic similarity between (9) and (11).

4.4. *Optimal Wiener Filtering*

Wiener's filter w is to minimize the mean squared estimation error e_w defined by $e_w(x) = \hat{u}_w(x) - u(x)$. That is,

$$w = \text{argmin}_h \text{E}(e_h^2) = \text{argmin}_h \text{E}(h * u_0(x) - u(x))^2. \tag{12}$$

Notice that Wiener's filter is independent of the particular pixel x used in the above definition since e_h is easily seen to be WSH for any fixed real filter $h = h(x)$, provided that u the ideal image and n the noise are independent and WSH.

Variation on the optimal Wiener filter: $w \to w + \delta h$ gives the "equilibrium" equation

$$\text{E}[(w * u_0(x) - u(x))\ (\delta h * u_0(x))] = 0.$$

Taking localized small variation $\delta h(x) = \varepsilon\delta(x - a)$ for some $\varepsilon \ll 1$ at any site a, one can rewrite the equation to $\text{E}[(w * u_0(x) - u(x))u_0(x - a)] = 0$. Since a is arbitrary, it is equivalent to

$$\text{E}[(w * u_0(x) - u(x))u_0(y)] = 0, \quad \forall x, y \in \Omega, \tag{13}$$

the one known as the *orthogonal condition* for Wiener's filter.

By Theorem 1, in terms of the correlation functions, one has

$$w * R_{u_0u_0}(z) = R_{uu_0}(z), \qquad z \in \mathbb{R}^2.$$

The optimal Wiener filter is thus given by $W(\omega) = S_{uu_0}(\omega)/S_{u_0u_0}(\omega)$, expressed in terms of the power spectral densities. For the blur model: $u_0 = k * u + n$, one has, according to Theorem 1,

$$S_{uu_0} = K^*(\omega)S_{uu}(\omega), \text{ and } S_{u_0u_0} = |K(\omega)|^2 S_{uu}(\omega) + S_{nn}(\omega).$$

Therefore, we have established the following theorem.

Theorem 2: (Wiener Filter for Deblurring) The optimal Wiener filter is given by, in the Fourier domain,

$$W(\omega) = \frac{K^* S_{uu}}{|K|^2 S_{uu} + S_{nn}} = \frac{K^*}{|K|^2 + r_w}, \tag{14}$$

where the regularizer $r_w = S_{nn}/S_{uu}$ is the squared noise-to-signal ratio.

For a Gaussian white noise with variance σ^2, one has $S_{nn}(\omega) \equiv \sigma^2$. Since S_{uu} is often bounded, the Wiener regularizer r_w is therefore well bounded above zero.

We refer the reader to, e.g., [43,47] for further improvement of the above classical Wiener filters, especially on relaxing the stochastic assumptions on the signals and the conditions on the blur model.

4.5. *Connection to the Bayesian/Tikhonov Method*

We now show that Wiener filtering is intimately connected to the general framework of Bayesian or Tikhonov regularization laid out in the preceding section.

Take the quadratic data-fitting model

$$E[u_0 \mid u, k] = \lambda \|k * u - u_0\|^2 = \lambda \int_{\mathbb{R}^2} (k * u - u_0)^2 dx$$

for additive Gaussian white noise, where λ is inversely proportional to the noise variance σ^2.

For the prior model $E[u]$, assume the ideal image u belongs to the fractional-Sobolev space $H^\gamma(\mathbb{R}^2)$. Formally, this means that $u \in L^2(\mathbb{R}^2)$ and its fractional gradient $\nabla^\gamma u \in L^2(\mathbb{R}^2)$. More rigorously, the norm is properly defined in the Fourier domain by:

$$\|u\|_\gamma^2 = \int_{\mathbb{R}^2} (1 + |\omega|^2)^\gamma |U(\omega)|^2 d\omega,$$

where $U(\omega)$ denotes the Fourier transform of $u(x)$. Define

$$r(\omega) = \frac{(1 + |\omega|^2)^\gamma}{\lambda} = \frac{\lambda^{-1}}{(1 + |\omega|^2)^{-\gamma}}. \tag{15}$$

Notice that $r(\omega)$ can indeed be considered as the squared noise-to-signal ratio as in the Wiener filter since λ^{-1} is proportional to the noise variance and the denominator is proportional to the squared signal strength. (More precisely, the noise n has been assumed in L^2 and its power spectral density is the ideal variance σ^2 modulated by some decay factor $|\omega|^{-2\alpha}$, which is however shared by the signal and cancelled out. This makes the above r a more authentic squared noise-to-signal ratio.)

Notice that the power-law decay in (15) is very common in stochastic signal analysis and processing.

One is thus led to the following posterior energy for deblurring:

$$E[u \mid u_0, k] = \lambda \|k * u - u_0\|^2 + \|u\|_\gamma^2.$$

In the Fourier domain, it is equivalent to the energy

$$E[U \mid U_0, K] = \int_\Omega |K(\omega)U(\omega) - U_0(\omega)|^2 \, d\omega + \int_\Omega r(\omega)|U(\omega)|^2 d\omega. \quad (16)$$

Performing variation on U, one has the equilibrium equation for the optimal estimator:

$$K^*(KU - U_0) + rU = 0, \text{ or } U(\omega) = \frac{K^*(\omega)}{|K|^2(\omega) + r(\omega)} U_0(\omega),$$

with r being the squared noise-to-signal ratio. This could be considered as the deterministic version of Wiener filtering. To our best knowledge, such an explicit connection has never been made before in the literature.

More generally, with $r(\omega)$ already computed for Wiener filtering (14), one can substitute it into the Bayesian/Tikhonov formulation (16), and arrive at the precise posterior energy form for *deterministic* Wiener filtering.

An interesting case occurs if $r(\omega) = (1+|\omega|^2)^{-\mu}$ for some notable $\mu > 0$, which corresponds to the scenario when the target image signal u is highly oscillatory, or is functionally a generalized function instead of L^2 [64].

5. Deblurring Blurred BV Images

One of the most powerful deterministic image prior model is the space of functions of bounded variations $\mathrm{BV}(\Omega)$, first introduced into image processing by Rudin, Osher, and Fatemi [61]. In this section, we discuss the theory and computation of deblurring BV images.

5.1. *TV Deblurring by Rudin, Osher, and Fatemi*

The total variation (TV) of a BV image u is conventionally denoted by $\int_\Omega |Du|$ or $|Du|(\Omega)$ [38,42]. When the image u is smooth so that its gradient

∇u belongs to L^1, the TV is simply the ordinary L^1 integral

$$\int_\Omega |Du| = \int_\Omega |\nabla u| dx,$$

in the sense of Sobolev norm. For a more generic BV image u that has discontinuous jumps, the TV $|Du|$ is in fact a Radon measure so that for any open set $Q \subset \Omega$,

$$|Du|(Q) = \sup_{\boldsymbol{g} \in C_0^1(Q, B^2)} \int_Q u(\nabla \cdot \boldsymbol{g}) dx,$$

where $B^2 \subset \mathbb{R}^2$ denotes the unit open disk centered at the origin and $\boldsymbol{g} = (g_1, g_2)$ is vectorial. For more introduction to BV images and their applications in image analysis and processing, we refer the reader to our new monograph [18], the more mathematically oriented monographs [38,42], as well as numerous existent works, e.g., [1,7,9,12,16,17,50,57,58,62,63,70,71,72].

In one adopts the TV measure for image regularization: $E[u] = \alpha \int_\Omega |Du|$, the posterior energy for Bayesian/Tikhonov deblurring then takes the form of

$$\begin{aligned} E[u \mid u_0, k] &= E[u] + E[u_0 \mid u, k] \\ &= \alpha \int_\Omega |Du| + \frac{\lambda}{2} \int_\Omega (k * u - u_0)^2 dx, \end{aligned} \tag{17}$$

with $x = (x_1, x_2) \in \Omega = \mathbb{R}^2$ and two suitable positive weights α and λ. This was the restoration model originally proposed and computed by Rudin-Osher-Fatemi [60,61], and later further studied by many others [9,72,73].

Notice that as far as energy minimization is concerned, only the ratio $r = \alpha/\lambda$ contributes to the solution process. As for parametric estimation in statistics, one could also treat r as an unknown as well, and expand the energy to $E[u, r \mid u_0, k]$ by absorbing some prior knowledge $E[r]$ on r.

The previous discussion on the optimal Wiener filtering (14) seems to suggest that the ratio $r = \alpha/\lambda$ is in the same dimension of the noise-to-signal ratio r_w. In particular, r should be proportional to the variance σ^2 of the noise, which is natural since by the Bayesian rationale for least square fidelities, one indeed has $\lambda = O(1/\sigma^2)$.

5.2. *Dealing With Bounded Image Domains*

In model (17), it has been conveniently assumed that the image domain Ω is the entire plane $\mathbb{R}^2$ to facilitate shift invariance. In real applications,

however, Ω is often a bounded disk or square for which the blur

$$K[u] = k * u(x) = \int_{\mathbb{R}^2} k(x-y)u(y)dy, \qquad x \in \Omega$$

needs to be properly redefined.

First, one can remodify the blur to a *shift-variant* PSF given by

$$k(x,y) = \frac{k(x-y)}{\int_\Omega k(x-z)dz}, \quad \forall x, y \in \Omega. \tag{18}$$

We assume that the original PSF $k(x)$ is nonnegative, and $x = (0,0)$ belongs to the support of the measure $d\mu(x) = k(x)dx$. That is, the integral of $k(x)$ on any neighborhood of $(0,0)$ is positive. Then the denominator in (18) is always positive. It is also easy to see that the DC-condition $K[1] = 1$ still holds after the modification.

An alternative way is to first extrapolate u beyond Ω. Let

$$Q : u\big|_\Omega \to \tilde{u} = Q[u]\big|_{\mathbb{R}^2},$$

be a suitable linear extrapolation operator which extends u on Ω onto the entire plane. (Functionally, Q could be some linear operator from, e.g., $W^{1,\infty}(\Omega)$ to $W^{1,\infty}(\mathbb{R}^2)$.) Then the blur is modified to

$$K[u](x) = k * \tilde{u}(x) = k * Q[u](x), \quad \forall x \in \Omega, \tag{19}$$

or equivalently, $K = 1_\Omega \cdot (k * Q)$ with a multiplier $1_\Omega(x)$.

The DC-condition is satisfied if and only if $k * Q[1] \equiv 1$ when restricted in Ω. In particular, the natural condition $Q[1] \equiv 1$ would suffice since k satisfies the DC-condition on $\mathbb{R}^2$.

If Q is represented by some kernel $g(x,y)$ with $y \in \Omega, x \in \mathbb{R}^2$. Then the modified K is represented by

$$k(x,y) = \int_{\mathbb{R}^2} k(x-z)g(z,y)dz, \quad x, y \in \Omega.$$

Therefore the DC-condition is satisfied when g and k meet the following compatibility condition

$$\int_\Omega \int_{\mathbb{R}^2} k(x-z)g(z,y)dzdy \equiv 1, \quad \forall x \in \Omega. \tag{20}$$

Finally, another less traditional approach to handling bounded domains can be based on the inpainting technique [4,5,11,13,14,15,16,17,19,36]. Suppose that $k(x)$ is compactly supported on a disk $B_\rho(0) = \{x \in \mathbb{R}^2 : |x| < \rho\}$, and the ρ-neighborhood of Ω is defined by

$$\Omega_\rho = \{x \in \mathbb{R}^2 \mid \mathrm{dist}(x, \Omega) < \rho\}.$$

Assume also the ideal image $u \in \mathrm{BV}(\Omega)$. Then, instead of the original model (17), one can attempt to minimize the modified version –

$$E[u \mid u_0, k, \rho] = \alpha \int_{\Omega_\rho} |Du| + \frac{\lambda}{2} \int_\Omega (k * u - u_0)^2 dx. \tag{21}$$

The convolution inside the fidelity term no longer stirs up any problem.

In summary, both the restricted-kernel method (18) and the image-extrapolation method (19) lead to a *shift-variant* blur K with kernel $k(x, y)$, and the deblurring model for BV images becomes

$$\min_u E_{\mathrm{TV}}[u \mid u_0, K] = \alpha \int_\Omega |Du| + \frac{\lambda}{2} \int_\Omega (K[u] - u_0)^2 dx. \tag{22}$$

Next we briefly discuss the solutions to this model. More details can be found, for example, in [1,9,18].

5.3. *Existence and Uniqueness*

Following the preceding preparation, the image domain Ω can be assumed bounded and Lipschitz in $\mathbb{R}^2$. In addition, we assume that (i) the ideal image $u \in \mathrm{BV}(\Omega)$, (ii) the blurry and noisy observation $u_0 \in L^2(\Omega)$, and (iii) the linear blur $K : L^1(\Omega) \to L^2(\Omega)$ is bounded, injective, and satisfies the DC-condition: $K[1] \equiv 1$.

Condition (i) and (ii) are necessary for (22) to be well defined. Injectivity in (iii) is also necessary for the uniqueness of optimal deblurring.

The proof for the following theorem can be found in, e.g., [9,18].

Theorem 3: (Existence and Uniqueness of BV Deblurring) Under the preceding three conditions, the optimal deblurring $u_* = \mathrm{argmin} E[u \mid u_0, K]$ for model (22) exists and is unique.

Furthermore, the unique minimizer must satisfy the mean constraint.

Corollary 4: *(The Mean Constraint) The unique minimizer u_* must automatically satisfy the mean constraint $\langle K[u_*] \rangle = \langle u_0 \rangle$.*

Stochastically, this is a natural inference from the blur model

$$u_0 = K[u] + n,$$

since the noise has zero means. Deterministically, this fact has to be proven from the deblurring model.

Proof: For the unique minimizer u_*, define for any $c \in \mathbb{R}$

$$e(c) = E[u_* - c \mid u_0, K].$$

Then $c_* = \operatorname{argmin} e(c)$ has to minimize

$$\int_\Omega (K[u_*] - u_0 - c)^2 dx, \qquad \text{since } K[c] = c.$$

As a result, the unique minimizer $c_* = \langle K[u_*] - u_0 \rangle$. On the other hand c_* has to be zero since $u_* - c_* = u_*$ due to uniqueness. Therefore,

$$\langle K[u_*] \rangle = \langle u_0 \rangle,$$

which establishes the assertion. □

Figures 3, 4, and 5 are three generic examples from [18] that demonstrate the performance of the deblurring model (22).

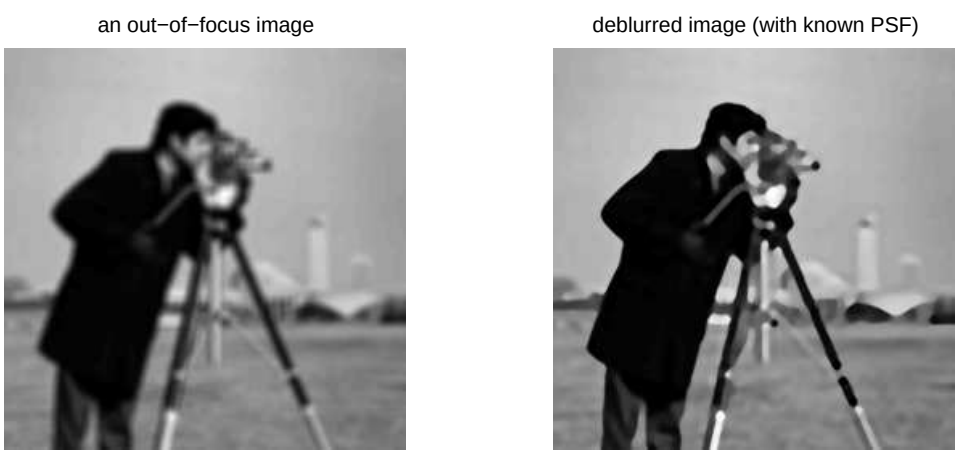

Fig. 3. Deblurring an out-of-focus image.

5.4. *Computation and Examples*

The variational deblurring model (22) has been computed more or less based on the formal Euler-Lagrange equation [1,9,60,61]:

$$\alpha \nabla \cdot \left[\frac{\nabla u}{|\nabla u|} \right] - \lambda K^*[K[u] - u_0] = 0, \tag{23}$$

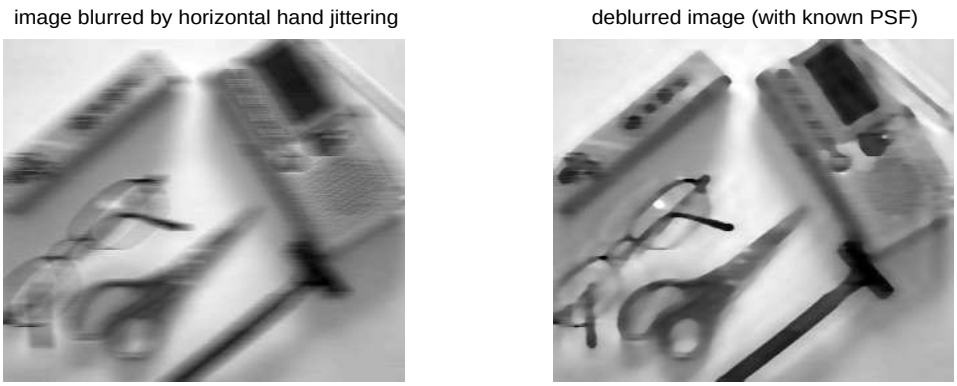

Fig. 4. Deblurring a motion-blurred image.

Fig. 5. Restoring another motion-blurred image.

with Neumann adiabatic condition $\partial u/\partial n = 0$ along the boundary $\partial\Omega$. Equation (23) holds in the distributional sense, i.e., for any compactly supported smooth test function ϕ, the solution satisfies

$$\alpha\langle\nabla\phi, \frac{\nabla u}{|\nabla u|}\rangle + \lambda\langle K[\phi], K[u] - u_0\rangle = 0.$$

The nonlinear degenerate elliptic equation (23) is often regularized to

$$\alpha\nabla\cdot\left[\frac{\nabla u}{|\nabla u|_a}\right] - \lambda K^*[K[u] - u_0] = 0, \tag{24}$$

where the notation $|x|_a$ denotes $\sqrt{x^2 + a^2}$ for some fixed positive parameter

a. It amounts to minimizing

$$E[u \mid u_0, K, a] = \alpha \int_\Omega \sqrt{|Du|^2 + a^2} + \frac{\lambda}{2} \int_\Omega (K[u] - u_0)^2 dx, \tag{25}$$

which is closely connected to the minimal surface problem [42].

Computationally, the most common algorithm has been based on the so-called *lagged-diffusivity* technique [9,18,72], which is an iterative procedure. Based on the current best estimation $u^{(n)}$, one solves for $u^{(n+1)}$ the following linearized equation:

$$\alpha \nabla \cdot \left[\frac{\nabla u^{(n+1)}}{|\nabla u^{(n)}|_a} \right] - \lambda K^*[K[u^{(n+1)}] - u_0] = 0, \tag{26}$$

with the Neumann condition. Notice that given $u^{(n)}$, the linear operator

$$L_n = -\alpha \nabla \cdot \frac{1}{|\nabla u^{(n)}|_a} \nabla + \lambda K^* K$$

is positive definite or strictly elliptic.

This algorithm guarantees convergence since it is equivalent to the alternating-minimization (AM) algorithm for the *augmented* energy

$$E_a[u, z \mid u_0, K] = \frac{\alpha}{2} \int_\Omega \left(z|\nabla u|^2 + z^{-1} \right) dx + \frac{\lambda}{2} \int_\Omega (K[u] - u_0)^2 dx,$$

where $z = z(x)$ is an auxiliary field, which corresponds to the edge signature function in image processing [17]. Then it can be easily shown that

$$\min_{u,z} E_a[u, z \mid u_0, K] = \min_u E_{\mathrm{TV}}[u \mid u_0, K].$$

Furthermore, the above lagged-diffusivity algorithm corresponds to exactly the AM algorithm for the augmented energy:

$$\cdots \to u^{(n)} \to z^{(n)} \to u^{(n+1)} \to \cdots .$$

6. Parametric Blind Deblurring

In all the above models, the blur K has been assumed known. We now develop variational deblurring models when the blur is unknown, a scenario often nicknamed "blind deblurring" [39,46,75]. Inspired by the theory of statistical estimation, we shall classify such models into ones that are *parametric* or *nonparametric*, or figuratively, *partially blind* or *completely blind*.

6.1. *Parametric Modeling*

Suppose that the unknown linear blur belongs to a known parametric family

$$\mathcal{K} = \{K_\theta \mid \theta \in I \subset \mathbb{R}^d\},$$

where $\theta = (\theta_1, \cdots, \theta_d)$ denotes a d-dimensional parametric vector and varies on a subset or domain I in $\mathbb{R}^d$. One is therefore not completely "blind" to the blur operator, and the uncertainty only arises from θ. A familiar example is the Gaussian family of shift-invariant blurs $K_\theta = g*$ given by

$$g(x \mid \theta) = \frac{1}{2\pi\theta} \exp\left(-\frac{x_1^2 + x_2^2}{2\theta}\right), \qquad \theta \in I = (0, \infty), \tag{27}$$

where in statistics θ precisely corresponds to the variance σ^2.

By the Bayesian rationale stated previously [18,53], parametric blind deblurring becomes the minimization of

$$E[u, \theta \mid u_0] = E[u_0 \mid u, \theta] + E[u] + E[\theta]. \tag{28}$$

The first two terms can be safely copied from the non-blind deblurring model discussed previously. Thus it suffices to incorporate some appropriate model for the parameter distribution $p(\theta)$ or $E[\theta]$.

Suppose $u \in \mathrm{BV}(\Omega)$, $\theta \in I \subset R^d$, and $E[\theta] = \phi(\theta)$ for some suitable function ϕ. Then the deblurring model is explicitly given by

$$E[u, \theta \mid u_0] = \alpha \int_\Omega |Du| + \frac{\lambda}{2} \int_\Omega (K_\theta[u] - u_0)^2 dx + \phi(\theta). \tag{29}$$

Assume that $\phi(\theta)$ is bounded below: $\phi(\theta) \geq M > -\infty$ for all $\theta \in I$. Otherwise it can attenuate the role of the first two terms in (29) and distort the real intention of the model. As an example, consider the Gaussian family in (27). Suppose the variance θ is subject to the exponential distribution with density function:

$$p(\theta) = a \exp(-a\theta), \qquad \theta \in I = (0, \infty), \quad \text{for some } a > 0. \tag{30}$$

Then $\phi(\theta) = E[\theta] = -\ln p(\theta) = a\theta - \ln a \geq -\ln a > -\infty$.

Following Theorem 3, K_θ is assumed to be injective and satisfy the DC-condition $K_\theta[1] = 1$. Then for any given θ, the conditional minimizer

$$\hat{u}_\theta = \operatorname{argmin} E[u \mid u_0, K_\theta] = \operatorname{argmin} \alpha \int_\Omega |Du| + \frac{\lambda}{2} \int_\Omega (K_\theta[u] - u_0)^2 dx \tag{31}$$

always exists and is unique by Theorem 3. The original model (29) is then reduced to an optimization problem on the parameter domain $I \subset \mathbb{R}^d$:

$$\min_{\theta \in I} e(\theta), \quad \text{with} \quad e(\theta) = E[\hat{u}_\theta, \theta \mid u_0].$$

$e(\theta)$ is, however, generally non-convex and consequently the global minimizer $(\hat{u}_{\theta_*}, \theta_*)$ could be non-unique.

6.2. *The AM Algorithm*

Such a multivariable optimization problem can usually be solved by the alternating-minimization (AM) algorithm [2,3,18,22,23,36,54]. One starts with some initial guess $\theta^{(0)}$, which could be drawn from $\operatorname{argmin}\phi(\theta)$ for instance. Then, one successively obtains the alternating conditional minimizers

$$\theta^{(0)} \to u^{(0)} \to \theta^{(1)} \to u^{(1)} \to \cdots \tag{32}$$

by optimizing the conditional energies:

$$\begin{aligned} u^{(n)} &= \operatorname{argmin} E[u \mid u_0, \theta^{(n)}], \quad \text{followed by} \\ \theta^{(n+1)} &= \operatorname{argmin} E[\theta \mid u_0, u^{(n)}], \quad \text{where} \\ E[\theta \mid u_0, u] &= \frac{\lambda}{2}\int_\Omega (K_\theta[u] - u_0)^2 dx + \phi(\theta). \end{aligned} \tag{33}$$

Notice that in the language of conditional probabilities, the Markov property holds for the zigzag sequence (32):

$$\begin{aligned} &\text{Prob}(\theta^{(n+1)} \mid u^{(n)}, \theta^{(n)}, u^{(n-1)}, \cdots) = \text{Prob}(\theta^{(n+1)} \mid u^{(n)}), \\ &\text{Prob}(u^{(n)} \mid \theta^{(n)}, u^{(n-1)}, \theta^{(n-1)}, \cdots) = \text{Prob}(u^{(n)} \mid \theta^{(n)}). \end{aligned}$$

By Theorem 3, the conditional update $\theta^{(n)} \to u^{(n)}$ must be unique, while the conditional parameter estimation $u^{(n)} \to \theta^{(n+1)}$ could be nonunique. Uniqueness can, however, still be enforced by some extra sorting scheme, e.g.,

$$\theta^{(n+1)} = \operatorname{argmin}\{\phi(\theta) \mid \theta \in \operatorname{argmin} E[\theta \mid u_0, u^{(n)}]\},$$

provided that $\phi(\theta)$ is strictly convex. The following is evident for AM.

Proposition 5: *(Alternating Minimization is Monotone) For each $n \geq 0$,*

$$E[u^{(n+1)}, \theta^{(n+1)} \mid u_0] \leq E[u^{(n)}, \theta^{(n)} \mid u_0].$$

Let $B(L^1, L^2)$ denote the Banach space of all bounded linear operators from $L^1(\Omega)$ to $L^2(\Omega)$. Then the following convergence result holds, whose proof can be found in our monograph [18].

Theorem 6: (Convergence of Alternating Minimization) Assume that

(a) the blur parametrization

$$K : I \subset \mathbb{R}^d \to B(L^1, L^2), \quad \theta \to K_\theta$$

is a continuous mapping; and

(b) $\phi(\theta)$ is lower semi-continuous in $\theta \in I$.

Then, if as $n \to \infty$, $u^{(n)} \to u_*$ in $L^1(\Omega)$ and $\theta^{(n)} \to \theta_* \in I$, the limit pair (u_*, θ_*) satisfies

$$u_* = \operatorname{argmin} E[u \mid u_0, \theta_*], \quad \theta_* = \operatorname{argmin} E[\theta \mid u_0, u_*]. \tag{34}$$

We must point out that the continuity on blur parametrization is strong but not baseless. Consider, for example, the shift-invariant Gaussian family (27) on $\Omega = \mathbb{R}^2$. By Young's inequality [48], one has

$$\|(K_\theta - K_{\theta'})[u]\|_2 = \|(g(x \mid \theta) - g(x \mid \theta')) * u\|_2 \le \|g(x \mid \theta) - g(x \mid \theta')\|_2 \|u\|_1.$$

Therefore, $\|K_\theta - K_{\theta'}\| \le \|g(x \mid \theta) - g(x \mid \theta')\|_2$, which indeed converges to zero for any $\theta' > 0$, and $\theta \to \theta'$.

In terms of the first formal variations,

$$\frac{\partial}{\partial u} E[u_* \mid u_0, \theta_*] = \frac{\partial}{\partial u} E[u_*, \theta_* \mid u_0]$$
$$\frac{\partial}{\partial \theta} E[\theta_* \mid u_0, u_*] = \frac{\partial}{\partial \theta} E[u_*, \theta_* \mid u_0].$$

Thus the limit (u_*, θ_*) does satisfy the equilibrium equations of the deblurring model $E[u, \theta \mid u_0]$, and consequently offers a good candidate for optimal deblurring. In particular, if $E[u, \theta \mid u_0]$ is strictly convex on $(u, \theta) \in \mathrm{BV}(\Omega) \times I$, (u_*, θ_*) must be the unique global minimizer.

7. Non-Parametric Blind Deblurring: Double-BV Model

7.1. *General Formulation of Blind Deblurring*

If the blur operator K is completely unknown, deblurring is conceivably much more challenging than the previous cases. Instead of estimating a few parameters, now one has to reconstruct the *entire* blur process K.

Herein we study only the shift-invariant case when the image observation u_0 is defined on $\Omega = \mathbb{R}^2$ with a PSF $k(x)$. The blur operator is thus reduced to a function, which is simpler than the general situation and can be managed with proper regularizations.

By the general Bayesian/Tikhonov framework, one attempts to minimize the posterior energy

$$E[u, k \mid u_0] = E[u] + E[u_0 \mid u, k] + E[k],$$

provided that the blur is independent of the image, as discussed earlier. In the case of BV images and Gaussian white noise, one has

$$E[u] = \alpha \int_{\mathbb{R}^2} |Du|, \qquad E[u_0 \mid u, k] = \frac{\lambda}{2} \int_{\mathbb{R}^2} (k * u - u_0)^2 dx.$$

Thus the key to successful deblurring lies in a proper proposal for the blur prior $E[k]$.

When the blur k is smooth, e.g., a Gaussian kernel, one may naturally enforce the Sobolev regularity [75]: $E[k] = \beta \int_{\mathbb{R}^2} |\nabla k|^2 dx$. Generally, such prior knowledge must be formulated based on the physical mechanism that drives the blur process, e.g., the atmospheric turbulence.

7.2. *Double-BV Blind Deblurring Model of Chan and Wong*

In motion blurs due to sudden jitters or out-of-focus blurs arising from ideal diffraction-free lenses (see, e.g., Chan and Shen [18]), the PSF's are typically compactly supported with sharp cutoff boundaries. In such scenarios, as for images with sharp edges, the total variation regularity seems more appealing for the blur k as well. This leads to the double-BV blind deblurring model of Chan and Wong [23]:

$$E[u, k \mid u_0 = \alpha \int_{\mathbb{R}^2} |Du| + \beta \int_{\mathbb{R}^2} |Dk| + \frac{\lambda}{2} \int_{\mathbb{R}^2} (k * u - u_0)^2 dx. \tag{35}$$

The detailed analysis for the double-BV model first appeared in our recent monograph [18]. Herein we only briefly introduce the most essential ingredients and refer the reader to [18] for more involved proofs and explanations.

For $\Omega = \mathbb{R}^2$, the BV norm is conventionally defined as [37,42]

$$\|u\|_{\mathrm{BV}} = \|u\|_{L^1(\mathbb{R}^2)} + |Du|(\mathbb{R}^2). \tag{36}$$

While most results on BV functions in the literature are for bounded domains, it is worthwhile to pay extra attention to the complexity arising from unboundedness. We refer the reader to the more detailed discussion in Chan and Shen [18].

We now first extend a Poincaré inequality from bounded domains to $\mathbb{R}^2$, whose proof can be found in [18].

Theorem 7: (Poincaré Inequality for BV($\mathbb{R}^2$)) Suppose u belongs to BV($\mathbb{R}^2$) with finite BV-norm defined as in (36). Then $u \in L^2(\mathbb{R}^2)$, and more specifically,

$$\|u\|_{L^2(\mathbb{R}^2)} \leq C|Du|(\mathbb{R}^2), \qquad \text{for some constant } C \text{ independent of } u.$$

For the double-BV blind deblurring model of Chan and Wong [23], we impose three conditions:

Condition A. The observation $u_0 \in L^2(\mathbb{R}^2) \cap L^\infty(\mathbb{R}^2)$.
Condition B. The ideal image u belongs to $\mathrm{BV}(\mathbb{R}^2)$.
Condition C. The blur PSF k belongs to $\mathrm{BV}(\mathbb{R}^2)$, and satisfies the DC-condition: $\int_{\mathbb{R}^2} k(x)dx = 1$.

The L^2 constraint in Condition A naturally comes from the data model in (35), while the L^∞ constraint is satisfied by most real imaging devices and is convenient for mathematical analysis. Furthermore, according to the Poincaré inequality in Theorem 7, Condition B implies that $u \in L^2(\mathbb{R}^2)$. Then by Young's inequality [48], one has

$$\|k * u\|_{L^2(\mathbb{R}^2)} \leq \|k\|_{L^1(\mathbb{R}^2)} \|u\|_{L^2(\mathbb{R}^2)},$$

which makes the data fitting term in (35) finite and well defined.

7.3. *On the Uniqueness: Hidden Symmetries*

In what follows we reveal some special symmetries hidden in the double-BV deblurring model (35). Such symmetries could lead to nonuniqueness of solutions.

Theorem 8: (Image-PSF Uncertainty) Suppose (u_*, k_*) minimizes the double-BV deblurring model (35) with (α, β, λ). Assume in addition that

$$m = \int_{\mathbb{R}^2} u_*(x)dx = \beta/\alpha.$$

Then $(u_+, k_+) = (mk_*, u_*/m)$ must be a minimizer as well.

The proof is a straightforward verification. Now for any given $a = (a_1, a_2) \in \mathbb{R}^2$, define the shifting operator

$$S_a : g(x) \to S_a[g] = g(x - a), \quad \text{for any measurable function } g.$$

Then it is well known that shifting commutes with convolution:

$$S_a[k * u] = S_a[k] * u = k * S_a[u].$$

Furthermore, for any $u, k \in \mathrm{BV}(\mathbb{R}^2)$, the invariance holds:

$$\int_{\mathbb{R}^2} |DS_a[u]| = \int_{\mathbb{R}^2} |Du| \ \text{ and } \ \int_{\mathbb{R}^2} S_a[k]dx = \int_{\mathbb{R}^2} kdx,$$

which induces the following symmetry.

Theorem 9: (Dual-Translation Uncertainty) Suppose (u_*, k_*) minimizes the double-BV deblurring model (35). Then for any $a \in \mathbb{R}^2$, $(S_a[u], S_a[k])$ is also a minimizer.

In order to better understand the double-BV deblurring model, consider now an easier but intimately related model - the double-Sobolev blind deblurring model $E_2[u, k \mid u_0]$ given by

$$\frac{\alpha}{2}\int_{\mathbb{R}^2}|\nabla u|^2 dx + \frac{\beta}{2}\int_{\mathbb{R}^2}|\nabla k|^2 dx + \frac{\lambda}{2}\int_{\mathbb{R}^2}(k * u - u_0)^2 dx, \tag{37}$$

for which both u and k belong to the Sobolev space $H^1(\mathbb{R}^2)$.

The unitary Fourier transform of a function $g(x)$ on $\mathbb{R}^2$ is defined by

$$G(\omega) = G(\omega_1, \omega_2) = \int_{\mathbb{R}^2} g(x) e^{-i2\pi\omega\cdot x} dx.$$

Then the unitary property of Fourier transform gives:

$$\int_{\mathbb{R}^2}|G(\omega)|^2 d\omega = \int_{\mathbb{R}^2}|g(x)|^2 dx, \text{ and } \int_{\mathbb{R}^2}|\nabla g(x)|^2 dx = 4\pi^2\int_{\mathbb{R}^2}\omega^2|G(\omega)|^2 d\omega,$$

with $\omega^2 = |\omega|^2 = \omega_1^2 + \omega_2^2$. Notice that the Fourier transform of $k * u$ is given by a direct product $K(\omega)U(\omega)$. Therefore, in the Fourier domain, the double-Sobolev blind deblurring energy $E_2[u, k \mid u_0]$ becomes $E_2[U, K \mid U_0]$, which is simply given by

$$2\pi^2\alpha\int_{\mathbb{R}^2}\omega^2|U(\omega)|^2 d\omega + 2\pi^2\beta\int_{\mathbb{R}^2}\omega^2|K(\omega)|^2 d\omega + \frac{\lambda}{2}\int_{\mathbb{R}^2}|K(\omega)U(\omega) - U_0(\omega)|^2 d\omega. \tag{38}$$

The DC-condition now requires $K(0) = 1$. Furthermore, since u, k, and u_0 are all real, one requires that both U and K satisfy the conjugate condition

$$\bar{U}(\omega) = U(-\omega) \text{ and } \bar{K}(\omega) = K(-\omega), \qquad \omega \in \mathbb{R}^2. \tag{39}$$

This leads to a nonuniqueness theorem more general than Theorem 9.

Theorem 10: (Dual-Phase Uncertainty) Let $(u_*, k_*) \in H^1(\mathbb{R}^2) \times H^1(\mathbb{R}^2)$ be a minimizer to the double-Sobolev blind deblurring model (37). And let

$$\phi(\omega) : \mathbb{R}^2 \to \mathbb{R}, \quad \omega \to \phi(\omega)$$

be any real smooth phase factor that is odd: $\phi(-\omega) = -\phi(\omega)$. Then

$$(u_+, k_+) = \text{Inverse Fourier Transforms of } (U_*(\omega)e^{i\phi(\omega)}, K_*(\omega)e^{-i\phi(\omega)})$$

must be also a minimizer.

Proof: It is straightforward to verify on the Fourier domain that

$$E[u_+, k_+ \mid u_0] = E[u_*, k_* \mid u_0],$$

and that both u_+ and k_+ are indeed real. Furthermore, k_+ does satisfy the DC-condition since

$$\int_{\mathbb{R}^2} k_+ dx = K_+(0) = K_*(0)e^{-i\phi(0)} = K_*(0) = 1. \qquad \square$$

In particular, by taking $\phi(\omega) = a \cdot \omega = a_1\omega_1 + a_2\omega_2$, one recovers the dual-translation uncertainty stated in Theorem 9. For uniqueness, it is therefore desirable to impose further conditions to break up the potential symmetries.

7.4. *The Existence Theorem*

The Poincaré's inequality in Theorem 7 can be further improved by dropping off the L^1 condition [18].

Theorem 11: (Poincaré's Inequality) For any $u \in L^2(\mathbb{R}^2)$, the Poincaré inequality holds:

$$\|u\|_{L^2(\mathbb{R}^2)} \leq C|Du|(\mathbb{R}^2), \quad \text{for some constant } C \text{ independent of } u.$$

The finiteness of the L^2-norm appears necessary due to counterexamples like $u \equiv 1$. The proof can be found in Chan and Shen [18].

Define the space BV_2 by

$$\mathrm{BV}_2(\mathbb{R}^2) = \{u \in L^2(\mathbb{R}^2) \mid |Du|(\mathbb{R}^2) < \infty\}.$$

Then by Theorem 7, $\mathrm{BV}(\mathbb{R}^2) \subset \mathrm{BV}_2(\mathbb{R}^2)$. The larger space BV_2 shall play a natural role for the blind deblurring model to be discussed below. We now study the existence of the double-BV blind deblurring model

$$E[u, k \mid u_0] = \alpha \int_{\mathbb{R}^2} |Du| + \beta \int_{\mathbb{R}^2} |Dk| + \frac{\lambda}{2} \int_{\mathbb{R}^2} (k * u - u_0)^2 dx. \tag{40}$$

The following conditions will be assumed for the study of existence.

Condition (a). Observation $u_0 \in L^2(\mathbb{R}^2) \cap L^\infty(\mathbb{R}^2)$.
Condition (b). Image $u \in \mathrm{BV}_2(\mathbb{R}^2)$, and $\|u\|_{L^\infty} \leq \|u_0\|_{L^\infty}$.
Condition (c). PSF $k \in \mathrm{BV}(\mathbb{R}^2)$, nonnegative, and satisfies the DC-condition: $\langle k, 1\rangle = 1$.

Notice that The constraints put differently on u and k help break their symmetric role in the model. However, even under Condition (b) and (c), the dual-translation uncertainty addressed by Theorem 9 is still not got rid of, since both conditions are still translation invariant.

For this purpose, Chan and Wong adopted the following centrosymmetry condition [23] to break the symmetry:

Condition (d'). The PSF is centrosymmetric: $k(-x) = k(x)$.

It amounts to requiring that the blur operator K is Hermitian. Numerical evidences in [23] seem to suggest that this condition can stably lead to unique deblurring solutions, though the theory has not been explored.

Herein to restrict the PSF to be highly concentrated near the origin, we impose the condition on the "tail" behavior of k.

Condition (d). There exists some nonnegative function $F(x) \in L^1(\mathbb{R}^2)$, and some positive radius $R > 0$, so that

$$0 \leq k(x) \leq F(x), \quad \forall x \in \mathbb{R}^2 : \ |x| \geq R. \tag{41}$$

For example, if $F(x) \equiv 0$ for all $|x| \geq R$, Condition (d) amounts to requiring k to be compactly supported on the disk $B_R = \{x \in \mathbb{R}^2 : |x| < R\}$.

Theorem 12: (Existence of Double-BV Blind Deblurring) Under Conditions (a), (b), (c), and (d), the minimizers to the double-BV blind deblurring model (40) exist.

The more technical proof could be found in our recent monograph [18]. Computationally, the double-BV blind deblurring model (40) can be also implemented via the AM algorithm, similar to that described in Eqn. (33) for parametric blind deblurring. More computational details can be found in the work of Chan and Wong [23].

8. Beblurring Besov Images via Iterated Shrinkage

In this section, we introduce the iterated-shrinkage algorithm of Daubechies et al. [30,31] for wavelet-based image deblurring.

Shrinkage has been one of the most efficient algorithms for image denoising and compression due to its low complexity and simple implementation, as studied in the stochastic framework by Donoho and Johnstone [34,35], and also in the variational framework by DeVore, et al. [8,32,33]. For deblurring, a direct shrinkage scheme becomes infeasible due to the global spatial correlation induced by the blur (integral) operator. Consequently,

Daubechies et al. in the aforementioned works developed the iterated-shrinkage algorithm to still take advantage of the efficiency of the shrinkage scheme.

8.1. *Wavelets and Besov Images*

We first briefly introduce wavelets and Besov images on $\mathbb{R}^2$. The reader is referred to, e.g., [18,29,51,52,67,74] for more details on these topics.

Let $\phi(x) = \phi(x_1, x_2)$ denote the scaling function, assumed to be compactly supported and sufficiently smooth for simplicity. For example, one can take the tensor product $\phi(x_1) \otimes \phi(x_2)$ of a 1D scaling function by Daubechies' design [28,29]. Assume that the three canonical wavelets associated to the multiresolution analysis of ϕ are given by

$$\psi^{(0,1)}(x), \quad \psi^{(1,0)}(x), \text{ and } \psi^{(1,1)}(x).$$

In the tensor-product framework, these can similarly be constructed from a 1D scaling function and its associated canonical wavelet. The wavelets are similarly assumed to be compactly supported and sufficiently smooth. Let

$$t = (t_1, t_2) \in T = \{(0,1), (1,0), (1,1)\}$$

denote one of the three wavelet types. Assume that each $\psi^t(x)$ has been normalized to have a unit L^2 norm, and the associated multiresoltuion analysis is orthogonal (*biorthogonalizty* imposes no extra challenge).

For any triple index

$$\lambda = (j, n, t) = (j, (n_1, n_2), t) \in \mathbb{Z} \times \mathbb{Z}^2 \times T,$$

define the L^2 normalized copy

$$\psi_\lambda(x) = \psi^t_{j,n}(x) = 2^j \psi^t(2^j x - n).$$

Similarly, define $\phi_n(x) = \phi(x - n)$ for any $n \in \mathbb{Z}^2$. Then $L^2(\mathbb{R}^2)$ has the *homogenous* orthonormal wavelet basis:

$$\left\{\psi_\lambda(x) : \lambda = (j, n, t) \in \mathbb{Z} \times \mathbb{Z}^2 \times T\right\}. \tag{42}$$

For $\lambda = (j, n, t)$, one defines $|\lambda| = j$ to be the associated resolution index. For any *locally integral* image u, its wavelet coefficients are defined by

$$u_\lambda = u_{(j,n,t)} = \langle u, \psi_\lambda \rangle = \int_{\mathbb{R}^2} u(x)\psi_\lambda(x)dx, \qquad \lambda \in \mathbb{Z} \times \mathbb{Z}^2 \times T.$$

In 2D, the space of Besov images, $B_q^\alpha(L^p)$ with $\alpha > 0, q, p \geq 1$, can be equivalently defined by:

$$\|u\|_{B_q^\alpha(L^p)} = \|\langle u, \phi_\bullet \rangle\|_{l^p} + \left[\sum_{j \geq 0} 2^{jq(\alpha+1-2/p)} \|u_{(j,\bullet,\bullet)}\|_{l^p}^q \right]^{1/q}.$$

The *homogeneous* Besov semi-norm $|u|_{\dot{B}_q^\alpha(L^p)}$, on the other hand, can be characterized by merely the wavelet coefficients:

$$|u|_{\dot{B}_q^\alpha(L^p)} = \left[\sum_{j \in \mathbb{Z}} 2^{jq(\alpha+1-2/p)} \|u_{(j,\bullet,\bullet)}\|_{l^p}^q \right]^{1/q}.$$

One Besov space of particular interest to image processing is when $\alpha = 1$ and $p = q = 1$, for which the semi-norm takes the simple form of:

$$|u|_{\dot{B}_1^1(L^1)} = \sum_{j \in \mathbb{Z}} \|u_{(j,\bullet,\bullet)}\|_{l^1} = \sum_\lambda |u_\lambda|. \tag{43}$$

The BV image prior, which has been extensively employed in the previous sections, is closely related to the Besov class $B_1^1(L^1)$. Roughly speaking, BV is somewhere between $B_1^1(L^1)$ and a weaker version of $B_1^1(L^1)$, which is the remarkable result established by Cohen et al. [25,26]. Thus in the wavelet literature, the BV image prior has often been approximated by the $B_1^1(L^1)$ [8,21], which shall also be adopted herein.

8.2. *Besov Image Deblurring via Iterated Shrinkage*

Consider the linear blur model with a known blur K and additive Gaussian noises:

$$u^0(x) = K[u](x) + n(x), \qquad x = (x_1, x_2) \in \mathbb{R}^2.$$

Assume that the blur $K : L^2 \to L^2$ is bounded and with operator norm $\|K\| \leq 1$. This is always true for any shift-invariant blur with a PSF $k(x)$ that is nonnegative everywhere. Then by Young's inequality, one has

$$\|k * u\|_{L^2} \leq \|k\|_{L^1} \, \|u\|_{L^2}, \tag{44}$$

and consequently the operator norm $\|K\| \leq \|k\|_{L^1} = 1$ since k satisfies the lowpass condition $\int_{\mathbb{R}^2} k(x)dx = 1$ and is nonnegative.

By the general Bayesian/Tikhonov framework discussed earlier, if the prior image model is taken to be the Besov space $B_1^1(L^1)$, the deblurring model is then given by

$$\hat{u} = \operatorname{argmin}_u \, E[u \mid u^0, K] = 2\alpha |u|_{\dot{B}_1^1(L^1)} + \beta \int_{\mathbb{R}^2} (K[u] - u^0)^2 dx, \tag{45}$$

where $\alpha, \beta > 0$ with β inversely proportional to the variance σ^2 of the noise, and α characterizing the sensitivity to signal roughness.

Under the wavelet basis, the deblurring energy takes the simple form of

$$E[u \mid u^0, K] = \sum_{\lambda} \left(2\alpha|u_\lambda| + \beta(K[u]_\lambda - u^0_\lambda)^2\right). \tag{46}$$

Unlike image denoising or compression for which $K = Id$ is the identity operator and the energy is completely decoupled, the present model is coupled across all the resolutions due to the blur operator K. This makes the classical wavelet shrinkage algorithms [34,35] not directly applicable.

Let $\delta u = \sum_\lambda \delta u_\lambda \psi_\lambda$ denote a perturbation. Then the first variation of the model energy $E[u \mid u^0, K]$ in (45) is given by

$$\begin{aligned} \delta E/2 &= \alpha \sum_{\lambda} \operatorname{sign}(u_\lambda)\delta u_\lambda + \beta \int_{\mathbb{R}^2} K^*[K[u] - u^0]\delta u dx \\ &= \sum_{\lambda} \big(\alpha \operatorname{sign}(u_\lambda) + \beta(M[u]_\lambda - g_\lambda)\big)\delta u_\lambda, \end{aligned} \tag{47}$$

where $M = K^*K$ and $g(x) = K^*[u^0]$. As a result, we have established the following theorem.

Theorem 13: The optimal deblurring must satisfy the system of equations:

$$0 = r \operatorname{sign}(u_\lambda) + (M[u]_\lambda - g_\lambda), \qquad \lambda \in \mathbb{Z} \times \mathbb{Z}^2 \times T, \tag{48}$$

where $r = \alpha/\beta$ could be considered as the noise-to-signal ratio as inspired by Wiener filtering and BV deblurring discussed earlier.

As in classical wavelet analysis [18,34], define the soft-shrinkage operator $S_r(t)$ by

$$S_r(t) = \operatorname{sign}(t)(|t| - r)^+ = \operatorname{sign}(t)\max(|t| - r, 0), \quad \text{for } t \in \mathbb{R}.$$

Then if there is no blur so that both K and M are the identity operator, the system of equilibrium equations (48) are then completely decoupled, and the optimal solution is directly given by

$$u_\lambda = S_r(g_\lambda), \quad \text{for } \lambda \in \mathbb{Z} \times \mathbb{Z}^2 \times T.$$

For deblurring, generally $M = K^*K$ is a mixing operator which could be sparse but is often not the identity matrix. Then Daubechies et al. [30,31] proposed the following iterated-shrinkage algorithm. Similar ideas also appeared in the variational-PDE literature for deblurring-related applications (see, e.g., Chan and Shen [17]).

To proceed, one first modifies that equilibrium system (48) to

$$0 = r\ \mathrm{sign}(u_\lambda) + (u_\lambda - g_\lambda) - (u_\lambda - M[u]_\lambda).$$

The the iterated-shrinkage algorithm of Daubechies et al. [30,31] is based on the following iteration scheme $u^k \to u^{k+1}$ at each time step k:

$$0 = r\ \mathrm{sign}(u_\lambda^{k+1}) + u_\lambda^{k+1} - \left(u_\lambda^k + g_\lambda - M[u^k]_\lambda\right). \tag{49}$$

Notice that due to the one-step time delay, the new system for u^{k+1} is *decoupled*. Furthermore, it takes the form of a blur-free denoising problem! Therefore, we have the following [30,31].

Theorem 14: At each time step k, the iteration is efficiently carried out by the shrinkage operator applied to each wavelet channel:

$$u_\lambda^{k+1} = S_r\left(u_\lambda^k + g_\lambda - M[u^k]_\lambda\right), \quad \text{for } \lambda \in \mathbb{Z} \times \mathbb{Z}^2 \times T.$$

8.3. *Understanding the Iterated-Shrinkage Algorithm*

We now present two ways to better understand the above iterated-shrinkage algorithm of Daubechies et al. [30,31], from both the differential-equation and variational points of view.

8.3.1. *As Semi-Implicit Time Marching*

Suppose more generally that the equilibrium system is augmented to

$$0 = r\ \mathrm{sign}(u_\lambda) + (Au_\lambda - g_\lambda) - (Au_\lambda - M[u]_\lambda), \tag{50}$$

for some constant $A \gg 1$. Then the iteration algorithm is given by

$$0 = r\ \mathrm{sign}(u_\lambda^{k+1}) + Au_\lambda^{k+1} - (Au_\lambda^k + g_\lambda - M[u^k]_\lambda), \tag{51}$$

which again allows an explicit shrinkage solution. Suppose $\Delta t = A^{-1} \ll 1$. Then the last equation can be rewritten to

$$\frac{u_\lambda^{k+1} - u_\lambda^k}{\Delta t} = -\left(r\ \mathrm{sign}(u_\lambda^{k+1}) + (M[u^k]_\lambda - g_\lambda)\right). \tag{52}$$

If one introduces the continuous time variable t so that

$$u_\lambda^k = u_\lambda(t = k\Delta t), \qquad k = 0, 1, \cdots.$$

Then the iteration (52) is precisely a semi-implicit scheme for the infinite system of coupled ordinary differential equations:

$$\frac{d}{dt}u_\lambda(t) = -\left(r\ \mathrm{sign}(u_\lambda(t)) + (M[u(t)]_\lambda - g_\lambda)\right), \qquad \lambda \in \mathbb{Z} \times \mathbb{Z}^2 \times T,$$

where the right hand side is precisely the negative gradient $-\frac{1}{2}\frac{\partial E}{\partial u_\lambda}$ by (47).

Therefore, in the limit of $A \gg 1$, or equivalently, $\Delta t = A^{-1} \ll 1$, the iterated shrinkage algorithm of Daubechies et al. [30,31] can be considered as a semi-implicit scheme for gradient-descent time marching for the deblurring energy $E[u \mid u^0, K]$.

8.3.2. *Via Augmentation and Auxiliary Variables*

The second way to understand the iterated-shrinkage algorithm is via the variational method on auxiliary variables and augmented functionals [30,31].

If one introduces an auxiliary variable z, which is considered as a delayed version of the target image u during the iteration, then the system (50) can be rewritten to

$$0 = r\ \mathrm{sign}(u_\lambda) + A(u_\lambda - z_\lambda) + M[z]_\lambda - g_\lambda. \tag{53}$$

Consequently, the iterated-shrinkage algorithm (51) can be considered as the solution to the above equation given $z = u^k$.

This motivates one to introduce an augmented energy $E[u, z \mid u^0, K]$ whose conditional energy $E[u \mid z, u^0, K]$ is in the integral form of (53):

$$E[u \mid z, u^0, K] = 2r|u|_{\dot{B}^1_1(L^1)} + A\|u - z\|^2 + 2\langle M[z] - g, u\rangle,$$

where both the norm and inner product are in the L^2 sense. Then given z, the system (53) yields the optimal u for $E[u, \mid z, u^0, K]$.

As a result, the full augmented energy must be given in the form of

$$E[u, z \mid u^0, K] = E[u \mid z, u^0, K] + \Phi[z \mid u^0, K],$$

where the functional Φ is independent of u.

We look for the specific form of Φ, such that (i) the iterated-shrinkage algorithm (51) corresponds to the AM (alternating-minimization) algorithm for the augmented energy $E[u, z \mid u^0, K]$; and (ii)

$$E[u, z \mid u^0, K] \geq \beta^{-1} E[u \mid u^0, K], \tag{54}$$

and the equality (or minimum) holds when $z = u$. The equality condition leads to

$$2\langle M[u] - g, u\rangle + \Phi[u \mid u^0, K] = \|K[u] - u^0\|^2.$$

Since $\langle M[u] - g, u\rangle = \langle K[u] - u^0, Ku\rangle$, this gives explicitly

$$\Phi[z \mid u^0, K] = \|K[z] - u^0\|^2 - 2\langle K[z] - u^0, Kz\rangle = -\langle K[z] - u^0, K[z] + u^0\rangle.$$

Notice that

$$\begin{aligned}
&2\langle M[z]-g,u\rangle-\langle K[z]-u^0,K[z]+u^0\rangle\\
=&2\langle K[z]-u^0,Ku\rangle-\langle K[z]-u^0,K[z]+u^0\rangle\\
=&-\langle K[z],K[z]\rangle+2\langle K[z],K[u]\rangle-2\langle K[u],u^0\rangle+\langle u^0,u^0\rangle\\
=&-\|K[z]-K[u]\|^2+\|K[u]-u^0\|^2\\
=&-\|K[u-z]\|^2+\|K[u]-u^0\|^2.
\end{aligned}$$

Therefore, the augmented energy is ultimately given by

$$E[uz\mid u^0,K]=2r|u|_{\dot{B}^1_1(L^1)}+A\|u-z\|^2-\|K[u-z]\|^2+\|K[u]-u^0\|^2. \quad (55)$$

Since $A\gg 1$ (and in particular $A\geq 1$) and the operator norm $\|K\|\leq 1$ as explained in (44), the augmented energy is indeed bounded below by $\beta^{-1}E[u\mid u^0,K]$ as required in (54). To conclude, we have the following theorem.

Theorem 15: The iterated-shrinkage algorithm for $E[u\mid u^0,K]$ of Daubechies et al. [30,31] is exactly the AM algorithm for the augmented energy $E[u,z\mid u^0,K]$. In particular, the algorithm must be stable and satisfy the monotone condition $E[u^{k+1}\mid u^0,K]\leq E[u^k\mid u^0,K]$.

9. Further Reading

For the several more involved proofs that have been left out, we refer the reader to our recent monograph [18]. For readers who are interested in this area, we also recommend to explore and read about other methodologies or related works, for example, the recursive inverse filtering (RIF) technique of Richardson [59] and Lucy [49] arising from astronomy imaging, as well as numerous works by other active researchers such as James Nagy et al. [55], Chan, Chan, Shen, and Shen [10] on wavelet deblurring via spatially varying filters, and Kindermann, Osher, and Jones [44] on nonlocal deblurring.

Acknowledgements

The authors would like to acknowledge the Institute of Mathematical Sciences, National University of Singapore, for her generous support to the visiting of the two authors, both financially and academically. The current work has also been partially supported by the National Science Foundation (NSF) of USA under grant numbers DMS-9973341 (Chan) and DMS-0202565 (Shen), the Office of Naval Research (ONR) of USA under grant number N00014-03-1-0888 (Chan), as well as the National Institute

of Health (NIH) of USA under grant number P20MH65166 (Chan). We are also very thankful to Yoon-Mo Jung for his careful proofreading of the manuscript and numerous suggestions and feedbacks.

References

1. R. Acar and C. R. Vogel. Analysis of total variation penalty methods for ill-posed problems. *Inverse Prob.*, 10:1217–1229, 1994.
2. L. Ambrosio and V. M. Tortorelli. Approximation of functionals depending on jumps by elliptic functionals via Γ-convergence. *Comm. Pure Appl. Math.*, 43:999–1036, 1990.
3. L. Ambrosio and V. M. Tortorelli. On the approximation of free discontinuity problems. *Boll. Un. Mat. Ital.*, 6-B:105–123, 1992.
4. M. Bertalmio, A. L. Bertozzi, and G. Sapiro. Navier-Stokes, fluid dynamics, and image and video inpainting. *IMA Preprint 1772 at: www.ima.umn.edu/preprints/jun01*, June, 2001.
5. M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. *Computer Graphics, SIGGRAPH 2000*, July, 2000.
6. P. Brémaud. *Markov Chains: Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer-Verlag New York, Inc., 1998.
7. A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imag. Vision*, 20:89–97, 2004.
8. A. Chambolle, R. A. DeVore, N.-Y. Lee, and B. J. Lucier. Nonlinear wavelet image processing: variational problems, compression and noise removal through wavelet shrinkage. *IEEE Trans. Image Processing*, 7(3):319–335, 1998.
9. A. Chambolle and P. L. Lions. Image recovery via Total Variational minimization and related problems. *Numer. Math.*, 76:167–188, 1997.
10. R. Chan, T. Chan, L. Shen, and Z. Shen. Wavelet deblurring algorithms for sparially varying blur from high-resolution image reconstruction. *Linear. Alg. Appl.*, 366:139–155, 2003.
11. T. F. Chan, S.-H. Kang, and J. Shen. Euler's elastica and curvature based inpainting. *SIAM J. Appl. Math.*, 63(2):564–592, 2002.
12. T. F. Chan, S. Osher, and J. Shen. The digital TV filter and nonlinear denoising. *IEEE Trans. Image Process.*, 10(2):231–241, 2001.
13. T. F. Chan and J. Shen. Nontexture inpainting by curvature driven diffusions (CDD). *J. Visual Comm. Image Rep.*, 12(4):436–449, 2001.
14. T. F. Chan and J. Shen. Bayesian inpainting based on geometric image models. *in Recent Progress in Comput. Applied PDEs, Kluwer Academic, New York*, pages 73–99, 2002.
15. T. F. Chan and J. Shen. Inpainting based on nonlinear transport and diffusion. *AMS Contemp. Math.*, 313:53–65, 2002.
16. T. F. Chan and J. Shen. Mathematical models for local nontexture inpaintings. *SIAM J. Appl. Math.*, 62(3):1019–1043, 2002.
17. T. F. Chan and J. Shen. On the role of the BV image model in image restoration. *AMS Contemp. Math.*, 330:25–41, 2003.

18. T. F. Chan and J. Shen. *Image Processing and Analysis: variational, PDE, wavelet, and stochastic methods*. SIAM Publisher, Philadelphia, 2005.
19. T. F. Chan and J. Shen. Variational image inpainting. *Comm. Pure Applied Math.*, 58:579–619, 2005.
20. T. F. Chan, J. Shen, and L. Vese. Variational PDE models in image processing. *Notices Amer. Math. Soc.*, 50:14–26, 2003.
21. T. F. Chan, J. Shen, and H.-M. Zhou. Total variation wavelet inpainting. *J. Math. Imag. Vision, to appear*, 2005.
22. T. F. Chan and L. A. Vese. A level set algorithm for minimizing the Mumford-Shah functional in image processing. *IEEE/Computer Society Proceedings of the 1st IEEE Workshop on "Variational and Level Set Methods in Computer Vision"*, pages 161–168, 2001.
23. T. F. Chan and C. K. Wong. Total variation blind deconvolution. *IEEE Trans. Image Process.*, 7(3):370–375, 1998.
24. D. Chandler. *Introduction to Modern Statistical Mechanics*. Oxford University Press, New York and Oxford, 1987.
25. A. Cohen, W. Dahmen, I. Daubechies, and R. DeVore. Harmonic analysis of the space BV. *Revista Matematica Iberoamericana*, 19:235–263, 2003.
26. A. Cohen, R. DeVore, P. Petrushev, and H. Xu. Nonlinear approximation and the space $\mathrm{BV}(R^2)$. *Amer. J. Math.*, 121:587–628, 1999.
27. T. M. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, Inc., New York, 1991.
28. I. Daubechies. Orthogonal bases of compactly supported wavelets. *Comm. Pure. Appl. Math.*, 41:909–996, 1988.
29. I. Daubechies. *Ten lectures on wavelets*. SIAM, Philadelphia, 1992.
30. I. Daubechies, M. Defrise, and C. DeMol. An iterative thresholding algorithm for lienar inverse problems with a sparsity constraint. *to appear in Comm. Pure Applied Math.*, 2005.
31. I. Daubechies and G. Teschke. Variational image restoration by means of wavelets: Simultaneous decomposition, deblurring, and denoising. *Appl. Comput. Harmon. Anal.*, 19(1):1–16, 2005.
32. R. A. DeVore, B. Jawerth, and B. J. Lucier. Image compression through wavelet transform coding. *IEEE Trans. Information Theory*, 38(2):719–746, 1992.
33. R. A. DeVore, B. Jawerth, and V. Popov. Compression of wavelet coefficients. *Amer. J. Math.*, 114:737–785, 1992.
34. D. L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Information Theory*, 41(3):613–627, 1995.
35. D. L. Donoho and I. M. Johnstone. Ideal spacial adaption by wavelet shrinkage. *Biometrika*, 81:425–455, 1994.
36. S. Esedoglu and J. Shen. Digital inpainting based on the Mumford-Shah-Euler image model. *European J. Appl. Math.*, 13:353–370, 2002.
37. L. C. Evans. *Partial Differential Equations*. Amer. Math. Soc., 1998.
38. L. C. Evans and R. F. Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, Inc., 1992.
39. D. Fish, A. Brinicombe, and E. Pike. Blind deconvolution by means of the

richardsonlucy algorithm. *J. Opt. Soc. Am. A*, 12:58–65, 1996.

40. S. Geman and D. Geman. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, 6:721–741, 1984.
41. W. Gibbs. *Elementary Principles of Statistical Mechanics*. Yale University Press, 1902.
42. E. Giusti. *Minimal Surfaces and Functions of Bounded Variation*. Birkhäuser, Boston, 1984.
43. A. D. Hillery and R. T. Chin. Iterative Wiener filters for image restoration. *IEEE Trans. Signal Processing*, 39:1892–1899, 1991.
44. S. Kindermann, S. Osher, and P. W. Jones. Deblurring and denoising of images by nonlocal functionals. *UCLA CAM Tech. Report*, 04-75, 2004.
45. D. C. Knill and W. Richards. *Perception as Bayesian Inference*. Cambridge Univ. Press, 1996.
46. R.L. Lagendijk, A. M. Tekalp, and J. Biemond. Maximum likelihood image and blur identification: A unifying approach. *Opt. Eng.*, 29:422–435, 1990.
47. J. S. Lee. Digital image enhancement and noise filtering by use of local statistics. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 2:165–168, 1980.
48. E. H. Lieb and M. Loss. *Analysis*. Amer. Math. Soc., second edition, 2001.
49. L. B. Lucy. An iterative technique for the rectification of observed distributions. *Astron. J.*, 79:745–754, 1974.
50. F. Malgouyres. Mathematical analysis of a model which combines total variation and wavelet for image restoration. *Journal of information processes*, 2(1):1–10, 2002.
51. S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, 1998.
52. Y. Meyer. *Wavelets and Operators*. Cambridge University Press, 1992.
53. D. Mumford. *Geometry Driven Diffusion in Computer Vision*, chapter "The Bayesian rationale for energy functionals", pages 141–153. Kluwer Academic, 1994.
54. D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Applied. Math.*, 42:577–685, 1989.
55. J. G. Nagy. A detailed publication list is available at the URL: http://www.mathcs.emory.edu/~nagy/research/pubs.html. Department of Mathematics and Computer Science, Emory University, Atlanta, GA 30322, USA.
56. A. V. Oppenheim and R. W. Schafer. *Discrete-Time Signal Processing*. Prentice Hall Inc., New Jersey, 1989.
57. S. Osher and N. Paragios. *Geometric Level Set Methods in Imaging, Vision and Graphics*. Springer Verlag, 2002.
58. S. Osher and J. Shen. Digitized PDE method for data restoration. In G. A. Anastassiou, editor, *Analytical-Computational Methods in Applied Mathematics*. Chapman & Hall/CRC, FL, 2000.
59. W. H. Richardson. Bayesian-based iterative method of image restoration. *J. Opt. Soc. Am.*, 62:55–59, 1972.

60. L. Rudin and S. Osher. Total variation based image restoration with free local constraints. *Proc. 1st IEEE ICIP*, 1:31–35, 1994.
61. L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.
62. J. Shen. On the foundations of vision modeling I. Weber's law and Weberized TV (total variation) restoration. *Physica D: Nonlinear Phenomena*, 175:241–251, 2003.
63. J. Shen. Bayesian video dejittering by BV image model. *SIAM J. Appl. Math.*, 64(5):1691–1708, 2004.
64. J. Shen. Piecewise $H^{-1} + H^0 + H^1$ images and the Mumford-Shah-Sobolev model for segmented image decomposition. *Appl. Math. Res. Exp.*, 4:143-167, 2005.
65. G. Strang. *Introduction to Applied Mathematics*. Wellesley-Cambridge Press, MA, 1993.
66. G. Strang. *Introduction to Linear Algebra*. Wellesley-Cambridge Press, 3rd edition, 1998.
67. G. Strang and T. Nguyen. *Wavelets and Filter Banks*. Wellesley-Cambridge Press, Wellesley, MA, 1996.
68. R. Strichartz. *A Guide to Distribution Theory and Fourier Transforms*. CRC Press, Ann Arbor, MI, USA, 1994.
69. A. N. Tikhonov. Regularization of incorrectly posed problems. *Soviet Math. Dokl.*, 4:1624–1627, 1963.
70. L. A. Vese. A study in the BV space of a denoising-deblurring variational problem. *Appl. Math. Optim.*, 44(2):131–161, 2001.
71. L. A. Vese and S. J. Osher. Modeling textures with Total Variation minimization and oscillating patterns in image processing. *J. Sci. Comput.*, 19:553–572, 2003.
72. C. R. Vogel. *Computational Methods for Inverse Problems*. SIAM, Philadelphia, 2002.
73. C. R. Vogel and M. E. Oman. Iterative methods for total variation denoising. *SIAM J. Sci. Comput.*, 17(1):227–238, 1996.
74. P. Wojtaszczyk. *A Mathematical Introduction to Wavelets*. London Mathematical Society Student Texts 37. Cambridge University Press, 1997.
75. Y. You and M. Kaveh. A regularization approach to joint blur identification and image restoration. *IEEE Transactions on Image Processing*, 5:416–428, 1996.