# SOLUTION DYNAMICS, CAUSALITY, AND CRITICAL BEHAVIOR OF THE REGULARIZATION PARAMETER IN TOTAL VARIATION DENOISING PROBLEMS

ANDY M. YIP* AND FREDERICK PARK*

**Abstract.** We analyze the role of the regularization parameter $\lambda$ in the total variation (TV) denoising model. There are several contributions in this paper. (1) We realize that, beyond controlling the smoothness of the solution, $\lambda$ exhibits another significant behavior — causality. This property allows us to solve the problem by incrementally increasing $\lambda$ so that each solution at a given regularization parameter can be used to solve the problem at a larger parameter efficiently. We call such a technique *parameter marching*. (2) While $\lambda$ is allowed to take a continuum of non-negative values, we show that only a unique finite number of them are critical and useful in the sense that they correspond to meaningful changes in signal features. Furthermore, we present the construction of these critical $\lambda$'s. (3) Since the analysis of the TV model is carried out by deriving exact solutions and investigating the dynamics of the solutions as $\lambda$ varies, many properties of the model which previously have not been studied to a rigorous extent are clearly revealed. A discussion of other possible further research within the parameter marching framework is also given. In particular, we present some insights into the improvement of the TV model provided by our analysis. We also discuss the possibility of applying parameter marching techniques to solving general Tikhonov-type regularization problems.

**Key words.** Total variation, denoising, image processing, critical parameters, causality, parameter marching

**AMS subject classifications.** 94A08, 94A12, 90C20

## 1. Introduction.

**1.1. The TV Denoising Model.** Denoising problems are of very great interest in various subjects including signal and image processing and data analysis. These problems are well-studied and many fast and reliable algorithms have been proposed. One of the denoising models that has drawn a lot of attention is the *Total Variation* (TV) model proposed by Rudin, Osher and Fatemi [14], also known as the Rudin-Osher-Fatemi model in literature. This model was shown to be very useful especially in image denoising problems for its preservation of edge information. Classical techniques such as linear low-pass filtering and minimization of mean square error with $L^2$-norm or $H^1$-norm regularization were shown to be unsuitable as edges, which contain very high frequency components, will be destroyed. Another common technique is (nonlinear) median filtering which is local in nature and thus is a good candidate for stream processing for its speed. However, due to the same reason, such a local technique ignores the global structure of the signal. Recently, several higher-order PDE models have been proposed which are evolved from the TV model, see [11, 12] and the references therein. The method in [11] attempts to overcome the undesirable staircasing effect of the TV model. The method in [12] aims at decomposing an image into texture and non-texture components such that noise removal is accomplished without destroying textures. Still, the TV model is an important building block for these higher-order PDE models. Therefore, it is worthy of investigation.

Given an observed signal $u^{(0)}(\mathbf{x})$ defined on $\Omega$ (usually a finite interval in $\mathbb{R}$ or a rectangle in $\mathbb{R}^2$) and a regularization parameter $\lambda \geq 0$, the TV model seeks for the solution to the minimization of a fitting term plus a fidelity term balanced by a

*Department of Mathematics, University of California, 405 Hilgard Avenue, Los Angeles, CA 90095-1555 ({mhyip,fpark}@math.ucla.edu).

regularization parameter $\lambda \geq 0$,

$$\min_{u \in BV(\Omega)} \frac{1}{2} \int_\Omega \left[u(\mathbf{x}) - u^{(0)}(\mathbf{x})\right]^2 d\mathbf{x} + \lambda \int_\Omega |\nabla u(\mathbf{x})| d\mathbf{x} \tag{1.1}$$

where $BV$ is the Banach space of all functions with bounded variation. Various algorithms have been proposed to solve the above minimization problem efficiently—an explicit artificial time marching scheme [14] in the original paper of Rudin et al, lagged diffusitivity fixed point iteration [17], multigrid techniques [18, 4], Newton continuation method [8], nonlinear filtering techniques [6], nonlinear primal-dual method [5], and TV minimization based on the dual formulation [3]. Two of the common characteristics of these algorithms are that (a) all of them are essentially variational methods based on solving the Euler-Lagrange equation for (1.1); (b) they are iterative methods which converge to the solution asymptotically.

**1.2. Objectives.** We show that in the TV denoising model, the regularization parameter exhibits a critical behavior not previously studied before. Moreover, solutions of the TV denoising model possess a causality property which suggests a scale space induced by $\lambda$. Utilizing exact solution formulas, we also study the dynamics of the solution as $\lambda$ varies. Thus, we identify $\lambda$ as a time variable. As far as the authors concerned, no previous work has considered this approach. In this paper, we consider TV denoising problems in the 1-D semi-discrete case (discrete in space and continuous in $\lambda$) where the minimization problem (1.1) becomes

$$J(\mathbf{u}) = \frac{1}{2} \sum_{i=1}^{n} \left(u_i - u_i^{(0)}\right)^2 + \lambda \sum_{i=1}^{n-1} |u_{i+1} - u_i| \to \min_{\mathbf{u} \in \mathbb{R}^n} J(\mathbf{u}). \tag{1.2}$$

Here, $\mathbf{u}^{(0)} = \left(u_1^{(0)}, \dots, u_n^{(0)}\right)^T \in \mathbb{R}^n$ is a given observed signal corrupted by additive noise, $\mathbf{u} = (u_1, \dots, u_n)^T \in \mathbb{R}^n$, and $\lambda \geq 0$ is a user-define regularization parameter.

**1.2.1. Critical $\lambda$'s.** Noticeably, in TV denoising problems, the solution depends continuously on the regularization parameter $\lambda$. Moreover, $\lambda$ is allowed to take a continuum of non-negative values. However, we show that only a unique finite number of them are critical and useful in the sense that they correspond to meaningful changes in signal features. Here, by a meaningful change, we refer to a change in shape of the signals. More precisely, in a discrete setting, suppose a signal $(v_1, \dots, v_n)$ is changed to $(u_1, \dots, u_n)$, the change is significant if there exists an $i$ such that $\operatorname{sgn}(u_{i+1} - u_i) = \operatorname{sgn}(v_{i+1} - v_i) \neq 0$ but $\operatorname{sgn}(v_{i+1} - v_i) = 0$; this corresponds to the merging of two signal features (removal of a discontinuity). In this sense, given a signal $\mathbf{u}^{(0)} = \left(u_1^{(0)}, \dots, u_n^{(0)}\right)^T$ there are at most $n-1$ number of $\hat{\lambda}^{(k)}$'s that correspond to changes in signal features. Here, each signal $\mathbf{u}(\hat{\lambda}^{(k)})$ differs from $\mathbf{u}(\hat{\lambda}^{(k+1)})$ by one oscillation. To the best of the authors' combined knowledge, none of the previous works realize this important property.

**1.2.2. Causality.** We prove that, beyond controlling the smoothness of the solution, $\lambda$ exhibits another significant behavior — causality — solution at each "time" $\tilde{\lambda}$ can be treated as an initial solution for computing solution at a "later time" $\lambda \geq \tilde{\lambda}$ [2]. Our method and analysis adapt this unique feature. In contrast to the above mentioned methods, iterative methods do not directly utilize this feature. Rather, iterative methods usually solve an Euler-Lagrange equation corresponding to a fixed $\lambda$ by starting at an arbitrary initial guess.

A similar idea having to do with the causality property is mentioned in [13] by Radmoser et al. Radmoser et al identify the Euler-Lagrange equation for (1.1) with a discrete approximation of a nonlinear diffusion equation when $\lambda$ is small. Thus, suggesting a link between the associated scale space of diffusion equations and the Euler-Lagrange equation. However, a rigorous foundation is absent in [13].

**1.2.3. Dynamics of Solutions.** Through our analysis and coupling with the above properties we derive a direct method which solves (1.2) by incrementally increasing the regularization parameter $\lambda$. We call such a technique *parameter marching*. Such a method identifies the dynamics of the solution which provides some alternative insights into the properties of the TV denoising model. These insights allow a better understanding of the model and some possible improvements. Before deriving an exact solution to (1.2) we make the preliminary observation that TV denoising problems can be reduced to quadratic programming problems. With this observation, our approach is to turn the discretized problem into a finite sequence of simple quadratic programming problems (QPP) which can be solved exactly in a fast manner. Each QPP corresponds to the TV minimization problem with respect to a critical regularization parameter $\hat{\lambda}^{(k)}$. Moreover, solutions to the sequence of QPPs are of practical interest since each solution retains the shape of the previous one except that the most obsolete oscillations are removed. Thus, the solution with respect to an arbitrary $\lambda$ which is in-between two consecutive critical parameters, i.e., $\hat{\lambda}^{(k)} < \lambda < \hat{\lambda}^{(k+1)}$, will have the same shape as the one at $\hat{\lambda}^{(k)}$ except that the most obsolete oscillations are dampened but not completely removed (these oscillations are completely removed at $\hat{\lambda}^{(k+1)}$).

Since we study the dynamics of the TV model by deriving exact solutions, we are able to reveal many properties of the model which previously have not been studied to a rigorous extent. Furthermore, we describe the denoising process geometrically in the sense that a kink in the trace of the solution corresponds to a change in signal features.

The organization of the rest of the paper is as follows. In §2, we present the critical behavior of the $\lambda$'s. Causality behavior and some further properties of the TV model are studied in §3. §4 deals with the geometry of minimizers. In §5, we present some numerical examples to illustrate our analysis. Finally, we give some ideas for future work in §6.

**2. Construction of Critical $\lambda$'s.** In this section, we show the existence and uniqueness of a set of critical $\lambda$'s at which each corresponding solution has structural changes. This is done by explicit construction of these $\lambda$'s and their corresponding solutions.

We first fix some notations that are used throughout the rest of the paper. Let $\mathbf{u}^{(0)} = \left(u_1^{(0)}, \ldots, u_n^{(0)}\right)^T$ be a given noisy signal of size $n$. A vector that is parallel (or anti-parallel) to $(1, 1, \ldots, 1)^T$ is said to be a *constant vector*. The $n$-dimensional Euclidean space with $u_i$ as its $i$-th dimension is named the *phase space* for $\mathbf{u} = (u_1, \ldots, u_n)^T$. The optimal solution to (1.2) with respect to $\lambda$ is denoted by $\mathbf{u}(\lambda) = (u_1(\lambda), \ldots, u_n(\lambda))^T$. A sequence of critical parameters is denoted by $\{\hat{\lambda}^{(k)}\}_{k=0}^K$ where $0 \le K \le n-1$ and $\hat{\lambda}^{(0)} \equiv 0$. Also, when $\lambda = \hat{\lambda}^{(k)}$, we denote $\mathbf{u}(\hat{\lambda}^{(k)})$ by $\mathbf{u}^{(k)}$. Finally, we define the indicator variable $\delta_{i+1/2}(\mathbf{u})$ by

$$\delta_{i+\frac{1}{2}}(\mathbf{u}) = \begin{cases} 1 & \text{if} \quad u_{i+1} \geq u_i \\ -1 & \text{if} \quad u_{i+1} < u_i \end{cases} \qquad \text{for } i = 1, \ldots, n-1.$$

To begin, we establish the stability of solutions to (1.2) with respect to $\lambda$ using standard compactness arguments. Similar results for the continuous case (1.1) can be found in [1].

LEMMA 2.1. *For a given* $\mathbf{u}^{(0)}$*, the solution* $\mathbf{u}(\lambda)$ *to (1.2) is a continuous function in* $\lambda \geq 0$.

*Proof.* Fix a $\lambda \geq 0$. Let $\lambda_k \geq 0$ such that $\lambda_k \to \lambda$. Let

$$J_k(\mathbf{u}) = \frac{1}{2}\|\mathbf{u} - \mathbf{u}^{(0)}\|_2^2 + \lambda_k \|\mathbf{u}\|_{\mathrm{TV}}.$$

Denote by $\mathbf{u}(\lambda_k)$ and $\tilde{\mathbf{u}}$ the minimizer of $J_k(\mathbf{u})$ and $J(\mathbf{u})$ respectively. Thus, we have $J_k(\mathbf{u}(\lambda_k)) \leq J_k(\tilde{\mathbf{u}})$ for all $k$. The sequence $\{J_k(\mathbf{u}(\lambda_k))\}$ is bounded since

$$\limsup_k J_k(\mathbf{u}(\lambda_k)) \leq \limsup_k J_k(\tilde{\mathbf{u}}) = \lim_k J_k(\tilde{\mathbf{u}}) = J(\tilde{\mathbf{u}}) < +\infty.$$

Using this, we obtain boundedness of $\{\mathbf{u}(\lambda_k) - \mathbf{u}^{(0)}\}$ (and hence boundedness of $\{\mathbf{u}(\lambda_k)\}$) in $\|\cdot\|_2$ because

$$\limsup_k \frac{1}{2}\|\mathbf{u}(\lambda_k) - \mathbf{u}^{(0)}\|_2^2 \leq \limsup_k J_k(\mathbf{u}(\lambda_k)) \leq J(\tilde{\mathbf{u}}) < +\infty.$$

This shows that the sequence $\{\mathbf{u}(\lambda_k)\}$ is contained in a compact set in $\mathbb{R}^n$. Suppose to the contrary that $\|\mathbf{u}(\lambda_k) - \tilde{\mathbf{u}}\|_2$ does not tend to 0. By compactness, there exists a subsequence $\{\mathbf{u}(\lambda_{k_l})\}$ converging to some $\bar{\mathbf{u}} \neq \tilde{\mathbf{u}}$. This leads to a contradiction to the uniqueness of solution to (1.2) since

$$J(\bar{\mathbf{u}}) = \lim_l J(\mathbf{u}(\lambda_{k_l})) = \lim_l [J(\mathbf{u}(\lambda_{k_l})) - J_{k_l}(\mathbf{u}(\lambda_{k_l}))] + \lim_l J_{k_l}(\mathbf{u}(\lambda_{k_l})) \leq 0 + J(\tilde{\mathbf{u}}). \quad \square$$

**2.1. Reducing TV Minimization Problems to Quadratic Programming Problems.** A key observation in the construction of analytic solutions of discrete TV denoising problems (which are non-smooth) is that they are equivalent to quadratic programming problems (which are smooth) when the prior information $\mathrm{sgn}[u_{i+1}(\lambda) - u_i(\lambda)]$ is known.

LEMMA 2.2. *Suppose the sign of the difference between the values of each pair of adjacent points of the solution* $\mathbf{u}(\lambda)$ *to (1.2) is known, i.e.,* $\mathrm{sgn}[u_{i+1}(\lambda) - u_i(\lambda)]$ *and hence* $\delta_{i+1/2}(\mathbf{u}(\lambda))$ *are known constants for all* $i$*. Then the minimization problem (1.2) is equivalent to*

$$\text{(2.1)} \qquad \min_{\mathbf{u} \in \mathbb{R}^n} \quad \frac{1}{2}\sum_{i=1}^{n}\left(u_i - u_i^{(0)}\right)^2 + \lambda \sum_{i=1}^{n-1} \delta_{i+\frac{1}{2}}(\mathbf{u}(\lambda))(u_{i+1} - u_i)$$

$$\textit{subject to} \quad \delta_{i+\frac{1}{2}}(\mathbf{u}(\lambda))(u_{i+1} - u_i) \geq 0 \quad \textit{if} \quad \mathrm{sgn}[u_{i+1}(\lambda) - u_i(\lambda)] \neq 0$$
$$\delta_{i+\frac{1}{2}}(\mathbf{u}(\lambda))(u_{i+1} - u_i) = 0 \quad \textit{if} \quad \mathrm{sgn}[u_{i+1}(\lambda) - u_i(\lambda)] = 0.$$

*Proof.* We fix a $\lambda \geq 0$. Let $\tilde{J}(\mathbf{u})$ be the objective function of (2.1) and $\tilde{\mathbf{u}}(\lambda)$ be the corresponding solution.

It is clear that $\mathbf{u} = \mathbf{u}(\lambda)$ satisfies all the constraints in (2.1) because

$$\delta_{i+\frac{1}{2}}(\mathbf{u}(\lambda))[u_{i+1}(\lambda) - u_i(\lambda)] = |u_{i+1}(\lambda) - u_i(\lambda)| > 0 \quad \text{if} \quad \operatorname{sgn}[u_{i+1}(\lambda) - u_i(\lambda)] \neq 0$$
$$\delta_{i+\frac{1}{2}}(\mathbf{u}(\lambda))(u_{i+1}(\lambda) - u_i(\lambda)) = 0 \quad \text{if} \quad \operatorname{sgn}[u_{i+1}(\lambda) - u_i(\lambda)] = 0.$$

This shows that $\mathbf{u}(\lambda)$ is a feasible solution to (2.1). Since, $\tilde{\mathbf{u}}(\lambda)$ is the minimizer of (2.1), we have $\tilde{J}(\tilde{\mathbf{u}}(\lambda)) \leq \tilde{J}(\mathbf{u}(\lambda))$. On the other hand, since $\mathbf{u}(\lambda)$ is the minimizer of (1.2), we have $J(\mathbf{u}(\lambda)) \leq J(\tilde{\mathbf{u}}(\lambda))$. Due to the constraints in (2.1), we see that $\tilde{J}(\mathbf{u}(\lambda)) = J(\mathbf{u}(\lambda))$ and $\tilde{J}(\tilde{\mathbf{u}}(\lambda)) = J(\tilde{\mathbf{u}}(\lambda))$. All together, we have

$$J(\tilde{\mathbf{u}}(\lambda)) = \tilde{J}(\tilde{\mathbf{u}}(\lambda)) \leq \tilde{J}(\mathbf{u}(\lambda)) = J(\mathbf{u}(\lambda)) \leq J(\tilde{\mathbf{u}}(\lambda)).$$

Hence, $J(\mathbf{u}(\lambda)) = J(\tilde{\mathbf{u}}(\lambda))$. By uniqueness of solution to (1.2), we have $\mathbf{u}(\lambda) = \tilde{\mathbf{u}}(\lambda)$. ∎

Problem (2.1) can be solved in $O(n)$ floating point operations since it is a very simple QPP whose variables are decoupled in the quadratic terms.

Of course, for a general $\lambda$, the sign of the difference $\operatorname{sgn}[u_{i+1}(\lambda) - u_i(\lambda)]$ between the values of each pair of adjacent points of $\mathbf{u}(\lambda)$ is unknown in advance. A key to obtain these sign values is the important observation that if $\operatorname{sgn}[u_{i+1}(\hat{\lambda}^{(k)}) - u_i(\hat{\lambda}^{(k)})]$ is known for some $\hat{\lambda}^{(k)} \geq 0$, then for all $\lambda > \hat{\lambda}^{(k)}$ sufficiently close to $\hat{\lambda}^{(k)}$, we do know the values of $\operatorname{sgn}[u_{i+1}(\lambda) - u_i(\lambda)]$ for all $i$ because the "shape" of the reconstructed signal $\mathbf{u}(\lambda)$ is exactly the same as that of $\mathbf{u}^{(k)}$. In the other words, monotonicity of neighboring values is preserved for $\lambda \in [\hat{\lambda}^{(k)}, \lambda^{(k+1)})$. This fact is proved in the following lemma.

Lemma 2.3. *For a given solution $\mathbf{u}^{(k)}$ at $\hat{\lambda}^{(k)} < \infty$, there exists a $\lambda^{(k+1)} > \hat{\lambda}^{(k)}$ such that the sign of the difference between the values of each pair of adjacent points of the minimizer $\mathbf{u}(\lambda)$ of (1.2) remains the same as that of $\mathbf{u}^{(k)}$ for all $\lambda \in [\hat{\lambda}^{(k)}, \lambda^{(k+1)})$, i.e.,*

$$\operatorname{sgn}[u_{i+1}(\lambda) - u_i(\lambda)] = \operatorname{sgn}\left(u_{i+1}^{(k)} - u_i^{(k)}\right)$$

*for all $\lambda \in [\hat{\lambda}^{(k)}, \lambda^{(k+1)})$ and for all $1 \leq i \leq n-1$.*

*Proof.* For the case $u_{i+1}^{(k)} \neq u_i^{(k)}$, the result simply follows from Lemma 2.1 (for $\lambda^{(k+1)}$ close enough to $\hat{\lambda}^{(k)}$).

For the case $u_{i+1}^{(k)} = u_i^{(k)}$, we first note that if $\mathbf{u}^{(k)}$ is a constant vector, then the result immediately follows from the fact that $\|\mathbf{u}(\lambda)\|_{\mathrm{TV}}$ is a non-increasing function in $\lambda$. Let us suppose that $\mathbf{u}^{(k)}$ is a non-constant vector.

Suppose to the contrary that there exists an $i_0$ such that $u_{i_0+1}^{(k)} = u_{i_0}^{(k)}$ but $u_{i_0+1}(\lambda) \neq u_{i_0}(\lambda)$ for all $\lambda > \hat{\lambda}^{(k)}$. Let $p \geq 0$ be the largest integer such that $u_{i+1}^{(k)} = u_i^{(k)}$ for all $i_0 \leq i \leq i_0 + p$. Let $q \geq 0$ be the largest integer such that $u_{i+1}^{(k)} = u_i^{(k)}$ for all $i_0 - q \leq i \leq i_0$. Since $\mathbf{u}^{(k)}$ is a non-constant vector, we must have $|u_{i_0-q}^{(k)} - u_{i_0-q-1}^{(k)}| > 0$ with $i_0 - q > 1$ or $|u_{i_0+p+1}^{(k)} - u_{i_0+p}^{(k)}| > 0$ with $i_0 + p < n-1$ or both. Note that in case $i_0+p+1 = n+1$ (respectively $i_0-q-1 = 0$), we define $u_0^{(k)} := u_1^{(k)}$ (respectively $u_{n+1}^{(k)} := u_n^{(k)}$). Denote $\sum_{i=-q-1}^{p} |v_{i+1} - v_i|$ by $\mathrm{TV}(v_{-q-1}, v_{-q}, \ldots, v_p, v_{p+1})$. Thus, we must have

$$\gamma := \mathrm{TV}(u_{i_0-q-1}^{(k)}, \ldots, u_{i_0+p+1}^{(k)}) = \sum_{i=i_0-q-1}^{i_0+p} |u_{i+1}^{(k)} - u_i^{(k)}| \tag{2.2}$$
$$= |u_{i_0-q}^{(k)} - u_{i_0-q-1}^{(k)}| + |u_{i_0+p+1}^{(k)} - u_{i_0+p}^{(k)}| > 0.$$

Next, by continuity stated in Lemma 2.1, for each $\epsilon > 0$, there exists a $\lambda_\epsilon > \hat{\lambda}^{(k)}$ such that $|u_i(\lambda_\epsilon) - u_i^{(k)}| < \epsilon$ for all $i$. In particular, if we choose $\epsilon$ such that

$$0 < \epsilon \leq \begin{cases} \frac{1}{2}|u_{i_0-q}^{(k)} - u_{i_0-q-1}^{(k)}| & \text{if } |u_{i_0+p+1}^{(k)} - u_{i_0+p}^{(k)}| = 0 \\ \frac{1}{2}|u_{i_0+p+1}^{(k)} - u_{i_0+p}^{(k)}| & \text{if } |u_{i_0-q}^{(k)} - u_{i_0-q-1}^{(k)}| = 0 \\ \frac{1}{2}\min\{|u_{i_0}^{(k)} - u_{i_0-1}^{(k)}|, |u_{i_0+2}^{(k)} - u_{i_0+1}^{(k)}|\} & \text{otherwise,} \end{cases}$$

then we have

$$|u_{i_0-q}^{(k)} - u_{i_0-q-1}^{(k)}| + |u_{i_0-q}(\lambda_\epsilon) - u_{i_0-q-1}(\lambda_\epsilon)| \equiv |u_{i_0-q}^{(k)} - u_{i_0-q-1}(\lambda_\epsilon)| + |u_{i_0-q}(\lambda_\epsilon) - u_{i_0-q-1}^{(k)}| \tag{2.3}$$

and

$$|u_{i_0+p+1}^{(k)} - u_{i_0+p}^{(k)}| + |u_{i_0+p+1}(\lambda_\epsilon) - u_{i_0+p}(\lambda_\epsilon)| \equiv |u_{i_0+p+1}(\lambda_\epsilon) - u_{i_0+p}^{(k)}| + |u_{i_0+p+1}^{(k)} - u_{i_0+p}(\lambda_\epsilon)|. \tag{2.4}$$

Define the functions $G_{\hat{\lambda}^{(k)}}(v_{i_0-q}, \ldots, v_{i_0+p})$ and $G_{\lambda_\epsilon}(v_{i_0-q}, \ldots, v_{i_0+p})$ by

$$G_{\hat{\lambda}^{(k)}}(v_{i_0-q}, \ldots, v_{i_0+p}) := \frac{1}{2}\sum_{i=i_0-q}^{i_0+p}\left(v_i - u_i^{(0)}\right)^2 + \hat{\lambda}^{(k)}\mathrm{TV}(u_{i_0-q-1}^{(k)}, v_{i_0-q}, \ldots, v_{i_0+p}, u_{i_0+p+1}^{(k)})$$

$$G_{\lambda_\epsilon}(v_{i_0-q}, \ldots, v_{i_0+p}) := \frac{1}{2}\sum_{i=i_0-q}^{i_0+p}\left(v_i - u_i^{(0)}\right)^2 + \lambda_\epsilon\mathrm{TV}(u_{i_0-q-1}(\lambda_\epsilon), v_{i_0-q}, \ldots, v_{i_0+p}, u_{i_0+p+1}(\lambda_\epsilon)).$$

These two functions just collect the terms in the objective function $J(\mathbf{u})$ that depend on $u_{i_0-q}, \ldots, u_{i_0+p}$ and then plug-in $(u_{i_0-q-1}, u_{i_0+p+1})$ with values $(u_{i_0-q-1}^{(k)}, u_{i_0+p+1}^{(k)})$ and $(u_{i_0-q-1}(\lambda_\epsilon), u_{i_0+p+1}(\lambda_\epsilon))$ respectively. Hence, we have

$$G_{\hat{\lambda}^{(k)}}(u_{i_0-q}^{(k)}, \ldots, u_{i_0+p}^{(k)}) < G_{\hat{\lambda}^{(k)}}(u_{i_0-q}(\lambda_\epsilon), \ldots, u_{i_0+p}(\lambda_\epsilon)) \tag{2.5}$$

$$G_{\lambda_\epsilon}(u_{i_0-q}(\lambda_\epsilon), \ldots, u_{i_0+p}(\lambda_\epsilon)) < G_{\lambda_\epsilon}(u_{i_0-q}^{(k)}, \ldots, u_{i_0+p}^{(k)}) \tag{2.6}$$

which essentially say that $\mathbf{u}^{(k)}$ and $\mathbf{u}(\lambda_\epsilon)$ are the minimizers with respect to the regularization parameter $\hat{\lambda}^{(k)}$ and $\lambda_\epsilon$ respectively.

We now aim at showing that equations (2.2)–(2.5) imply

$$G_{\lambda_\epsilon}(u_{i_0-q}^{(k)}, \ldots, u_{i_0+p}^{(k)}) \leq G_{\lambda_\epsilon}(u_{i_0-q}(\lambda_\epsilon), \ldots, u_{i_0+p}(\lambda_\epsilon)) \tag{2.7}$$

which is a contradiction to (2.6) and thus the result follows. Notice that using (2.5), we obtain

$$G_{\lambda_\epsilon}(u_{i_0-q}^{(k)}, \ldots, u_{i_0+p}^{(k)}) = \frac{1}{2}\sum_{i=i_0-q}^{i_0+p}\left(u_i^{(k)} - u_i^{(0)}\right)^2 + \lambda_\epsilon\mathrm{TV}(u_{i_0-q-1}(\lambda_\epsilon), u_{i_0-q}^{(k)}, \ldots, u_{i_0+p}^{(k)}, u_{i_0+p+1}(\lambda_\epsilon))$$

$$
\begin{aligned}
&= \frac{1}{2}\sum_{i=i_0-q}^{i_0+p}\left(u_i^{(k)}-u_i^{(0)}\right)^2+\hat{\lambda}^{(k)}\mathrm{TV}(u_{i_0-q-1}^{(k)},u_{i_0-q}^{(k)},\ldots,u_{i_0+p}^{(k)},u_{i_0+p+1}^{(k)})\\
&\quad -\hat{\lambda}^{(k)}\mathrm{TV}(u_{i_0-q-1}^{(k)},u_{i_0-q}^{(k)},\ldots,u_{i_0+p}^{(k)},u_{i_0+p+1}^{(k)})\\
&\quad +\lambda_\epsilon\mathrm{TV}(u_{i_0-q-1}(\lambda_\epsilon),u_{i_0-q}^{(k)},\ldots,u_{i_0+p}^{(k)},u_{i_0+p+1}(\lambda_\epsilon))\\
&= G_{\hat{\lambda}^{(k)}}(u_{i_0-q}^{(k)},\ldots,u_{i_0+p}^{(k)})-\hat{\lambda}^{(k)}\mathrm{TV}(u_{i_0-q-1}^{(k)},u_{i_0-q}^{(k)},\ldots,u_{i_0+p}^{(k)},u_{i_0+p+1}^{(k)})\\
&\quad +\lambda_\epsilon\mathrm{TV}(u_{i_0-q-1}(\lambda_\epsilon),u_{i_0-q}^{(k)},\ldots,u_{i_0+p}^{(k)},u_{i_0+p+1}(\lambda_\epsilon))\\
&= G_{\hat{\lambda}^{(k)}}(u_{i_0-q}^{(k)},\ldots,u_{i_0+p}^{(k)})+\hat{\lambda}^{(k)}\{\mathrm{TV}(u_{i_0-q-1}(\lambda_\epsilon),u_{i_0-q}^{(k)},\ldots,u_{i_0+p}^{(k)},u_{i_0+p+1}(\lambda_\epsilon))\\
&\quad -\mathrm{TV}(u_{i_0-q-1}^{(k)},u_{i_0-q}^{(k)},\ldots,u_{i_0+p}^{(k)},u_{i_0+p+1}^{(k)})\}\\
&\quad +(\lambda_\epsilon-\hat{\lambda}^{(k)})\{\mathrm{TV}(u_{i_0-q-1}(\lambda_\epsilon),u_{i_0-q}^{(k)},\ldots,u_{i_0+p}^{(k)},u_{i_0+p+1}(\lambda_\epsilon))\}\\
&< G_{\hat{\lambda}^{(k)}}(u_{i_0-q}(\lambda_\epsilon),\ldots,u_{i_0+p}(\lambda_\epsilon))+\hat{\lambda}^{(k)}\{|u_{i_0-q}^{(k)}-u_{i_0-q-1}(\lambda_\epsilon)|\\
&\quad +|u_{i_0+p+1}(\lambda_\epsilon)-u_{i_0+p}^{(k)}|-|u_{i_0-q}^{(k)}-u_{i_0-q-1}^{(k)}|-|u_{i_0+p+1}^{(k)}-u_{i_0+p}^{(k)}|\}\\
&\quad +(\lambda_\epsilon-\hat{\lambda}^{(k)})\{\mathrm{TV}(u_{i_0-q-1}(\lambda_\epsilon),u_{i_0-q}^{(k)},\ldots,u_{i_0+p}^{(k)},u_{i_0+p+1}(\lambda_\epsilon))\}\\
&= G_{\lambda_\epsilon}(u_{i_0-q}(\lambda_\epsilon),\ldots,u_{i_0+p}(\lambda_\epsilon))+\hat{\lambda}^{(k)}\{|u_{i_0-q}^{(k)}-u_{i_0-q-1}(\lambda_\epsilon)|\\
&\quad +|u_{i_0+p+1}(\lambda_\epsilon)-u_{i_0+p}^{(k)}|-|u_{i_0-q}^{(k)}-u_{i_0-q-1}^{(k)}|-|u_{i_0+p+1}^{(k)}-u_{i_0+p}^{(k)}|\\
&\quad +|u_{i_0-q}(\lambda_\epsilon)-u_{i_0-q-1}^{(k)}|+|u_{i_0+p+1}^{(k)}-u_{i_0+p}(\lambda_\epsilon)|\\
&\quad -|u_{i_0-q}(\lambda_\epsilon)-u_{i_0-q-1}(\lambda_\epsilon)|-|u_{i_0+p+1}(\lambda_\epsilon)-u_{i_0+p}(\lambda_\epsilon)|\}\\
&\quad +(\lambda_\epsilon-\hat{\lambda}^{(k)})\{\mathrm{TV}(u_{i_0-q-1}(\lambda_\epsilon),u_{i_0-q}^{(k)},\ldots,u_{i_0+p}^{(k)},u_{i_0+p+1}(\lambda_\epsilon))\\
&\quad -\mathrm{TV}(u_{i_0-q-1}(\lambda_\epsilon),u_{i_0-q}(\lambda_\epsilon),\ldots,u_{i_0+p}(\lambda_\epsilon),u_{i_0+p+1}(\lambda_\epsilon))\}\\
&:= G_{\lambda_\epsilon}(u_{i_0-q}(\lambda_\epsilon),\ldots,u_{i_0+p}(\lambda_\epsilon))+\hat{\lambda}^{(k)}\{\mathrm{I}\}+(\lambda_\epsilon-\hat{\lambda}^{(k)})\{\mathrm{II}\}.
\end{aligned}
$$

Now, by (2.3) and (2.4), we have $\mathrm{I}\equiv 0$. Since $u_i(\lambda_\epsilon)\to u_i^{(k)}$ for all $i$ uniformly and $u_i^{(k)}=u_{i+1}^{(k)}$ for $i_0-q\le i\le i_0+p-1$, we have $|u_{i+1}(\lambda_\epsilon)-u_i(\lambda_\epsilon)|\to 0$ for $i_0-q\le i\le i_0+p-1$ uniformly. Together with (2.2), we obtain $\mathrm{II}\to -\gamma<0$ as $\epsilon\to 0$. Hence, for $\epsilon$ small enough, we have (2.7) which contradicts to (2.6). ∎
In Figure 2.1, we illustrate the idea that the shape of $\mathbf{u}(\lambda)$ retains that of $\mathbf{u}^{(k)}$ when $\lambda>\hat{\lambda}^{(k)}$ is small enough.

Lemma 2.3 implies that we can advance from $\hat{\lambda}^{(k)}$ to a larger value $\lambda^{(k+1)}$ until the shape of the reconstructed signal changes. Once we encounter such structural changes, the values $\mathrm{sgn}[u_{i+1}(\lambda^{(k+1)})-u_i(\lambda^{(k+1)})]$ are updated. By repeating this process, we can obtain the values $\mathrm{sgn}[u_{i+1}(\lambda)-u_i(\lambda)]$ at any desired $\lambda$. We will show in §2.4 that this process needs to be repeated at most $n-1$ times.

**2.2. Explicit Solution to the Quadratic Programming Problem.** Based on Lemmas 2.2 and 2.3, when $\lambda\in[\hat{\lambda}^{(k)},\lambda^{(k+1)})$, the minimization problem (1.2) is equivalent to

$$
\begin{aligned}
&\min_{\mathbf{u}\in\mathbb{R}^n}\quad \frac{1}{2}\sum_{i=1}^{n}\left(u_i-u_i^{(0)}\right)^2+\lambda\sum_{i=1}^{n-1}\delta_{i+\frac{1}{2}}(\mathbf{u}^{(k)})(u_{i+1}-u_i)\\
&\text{subject to}\quad \delta_{i+\frac{1}{2}}(\mathbf{u}^{(k)})(u_{i+1}-u_i)\ge 0\quad \text{if}\quad u_{i+1}^{(k)}\ne u_i^{(k)}\\
&\qquad\qquad\quad \delta_{i+\frac{1}{2}}(\mathbf{u}^{(k)})(u_{i+1}-u_i)=0\quad \text{if}\quad u_{i+1}^{(k)}= u_i^{(k)}.
\end{aligned}
\tag{2.8}
$$

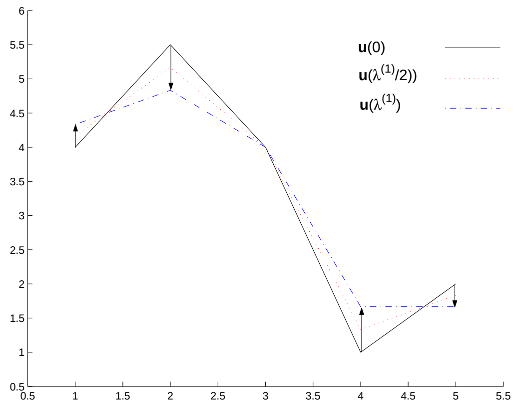


FIG. 2.1. *Solutions at $\lambda = 0$, $\hat{\lambda}^{(1)}/2$, and $\hat{\lambda}^{(1)}$. The shape of $\mathbf{u}(\hat{\lambda}^{(1)}/2)$ retains that of $\mathbf{u}^{(0)}$. The values $u_4(\lambda)$ and $u_5(\lambda)$ are the pair of adjacent values which approach each other faster than other pairs of adjacent values.*

The above problem is a quadratic programming problem with linear constraints which has a simple explicit solution given in the the following lemma.

Before stating the results, we fix some notations. For each $i$, we denote by $R_i^{(k)} := (i - q_i^{(k)}, \dots, i, \dots, i + p_i^{(k)})$ the maximal connected neighborhood of $i$ at which $\{u_i^{(k)}\}$ takes the same value, i.e., $p_i^{(k)}$ and $q_i^{(k)}$ are nonnegative integers such that

$$u_{i-q_i^{(k)}-1}^{(k)} \neq u_{i-q_i^{(k)}}^{(k)} = \dots = u_i^{(k)} = \dots = u_{i+p_i^{(k)}}^{(k)} \neq u_{i+p_i^{(k)}+1}^{(k)}.$$

Moreover, the average value of the entries of $\mathbf{u}^{(0)}$ on this neighborhood is denoted by

$$a_i^{(k)} := \frac{\sum_{l=-q_i^{(k)}}^{p_i^{(k)}} u_{i+l}^{(0)}}{p_i^{(k)} + q_i^{(k)} + 1}.$$

We also define

$$\alpha_i^{(k)} := \frac{\delta_{i+p_i^{(k)}+\frac{1}{2}}(\mathbf{u}^{(k)}) - \delta_{i-q_i^{(k)}-\frac{1}{2}}(\mathbf{u}^{(k)})}{p_i^{(k)} + q_i^{(k)} + 1}. \tag{2.9}$$

Here, $\delta_{1/2}(\mathbf{u}^{(k)})$ and $\delta_{n+1/2}(\mathbf{u}^{(k)})$ are defined to be 0. We remark that the identities $\delta_{1/2}(\mathbf{u}^{(k)})(u_1 - u_0) \equiv \delta_{n+1/2}(\mathbf{u}^{(k)})(u_{n+1} - u_n) \equiv 0$ can be interpreted as homogeneous Neumann boundary conditions.

LEMMA 2.4. *For each $\lambda \in [\hat{\lambda}^{(k)}, \lambda^{(k+1)})$, where $\hat{\lambda}^{(k)} < \infty$ and $\lambda^{(k+1)}$ is guaranteed by Lemma 2.3, the unique solution to (2.8) is given by*

$$u_i(\lambda) = a_i^{(k)} + \lambda \alpha_i^{(k)}, \tag{2.10}$$

*or, equivalently,*

$$u_i(\lambda) = u_i^{(k)} + (\lambda - \hat{\lambda}^{(k)}) \alpha_i^{(k)}. \tag{2.11}$$

We notice that (2.11) is also valid if $\mathbf{u}^{(k)}$ is a constant vector since $\alpha_i^{(k)} = 0$ for all $i$.

*Proof.* Let $J^{(k)}(\mathbf{u})$ be the objective function in (2.8). Let $L^{(k)}(\mathbf{u})$ be the corresponding Lagrangian function

$$L^{(k)}(\mathbf{u}) = J^{(k)}(\mathbf{u}) - \sum_{i=1}^{n-1} \beta_{i+\frac{1}{2}} \delta_{i+\frac{1}{2}}(\mathbf{u}^{(k)})(u_{i+1} - u_i)$$

where $\beta_{i+1/2}$'s are Lagrange multipliers. The solution $\mathbf{u}(\lambda)$ to (2.8) must satisfy the Karush-Kuhn-Tucker Conditions (KKT Conditions) [10, pp. 69–71],

$$\nabla L^{(k)}(\mathbf{u}(\lambda)) = \mathbf{0} \tag{2.12}$$

$$\beta_{i+\frac{1}{2}} \geq 0 \quad \text{if} \quad u_{i+1}^{(k)} \neq u_i^{(k)} \tag{2.13}$$

$$\beta_{i+\frac{1}{2}} = 0 \quad \text{if} \quad \delta_{i+\frac{1}{2}}(\mathbf{u}^{(k)})[u_{i+1}(\lambda) - u_i(\lambda)] > 0. \tag{2.14}$$

The components of the vector equation $\nabla L^{(k)}(\mathbf{u}(\lambda)) = \mathbf{0}$ read

$$u_i(\lambda) = u_i^{(0)} + \lambda \left[\delta_{i+\frac{1}{2}}(\mathbf{u}^{(k)}) - \delta_{i-\frac{1}{2}}(\mathbf{u}^{(k)})\right] + \beta_{i-\frac{1}{2}}\delta_{i-\frac{1}{2}}(\mathbf{u}^{(k)}) - \beta_{i+\frac{1}{2}}\delta_{i+\frac{1}{2}}(\mathbf{u}^{(k)}) \tag{2.15}$$

for $i = 1, \ldots, n$. By Lemma 2.3 and the choice of $p_i^{(k)}$ and $q_i^{(k)}$, we have

$$u_{i-q_i^{(k)}-1}(\lambda) \neq u_{i-q_i^{(k)}}(\lambda) \tag{2.16}$$

$$u_{i+l}(\lambda) = u_i(\lambda) \quad \text{for} \quad l = -q_i^{(k)}, \ldots, p_i^{(k)} \tag{2.17}$$

$$u_{i+p_i^{(k)}}(\lambda) \neq u_{i+p_i^{(k)}+1}(\lambda). \tag{2.18}$$

As a result, the KKT Condition (2.14), and equations (2.16) and (2.18) imply that $\beta_{i-q_i^{(k)}-1/2} = \beta_{i+p_i^{(k)}+1/2} = 0$. By summing up the equations (2.15) for $i-q_i^{(k)}, \ldots, i+p_i^{(k)}$, we obtain

$$\begin{aligned}
\sum_{l=-q_i^{(k)}}^{p_i^{(k)}} u_{i+l}(\lambda) &= \sum_{l=-q_i^{(k)}}^{p_i^{(k)}} u_{i+l}^{(0)} + \lambda \sum_{l=-q_i^{(k)}}^{p_i^{(k)}} \left[\delta_{i+l+\frac{1}{2}}(\mathbf{u}^{(k)}) - \delta_{i+l-\frac{1}{2}}(\mathbf{u}^{(k)})\right] \\
&\quad + \sum_{l=-q_i^{(k)}}^{p_i^{(k)}} \left[\beta_{i+l-\frac{1}{2}}\delta_{i+l-\frac{1}{2}}(\mathbf{u}^{(k)}) - \beta_{i+l+\frac{1}{2}}\delta_{i+l+\frac{1}{2}}(\mathbf{u}^{(k)})\right] \\
&= \sum_{l=-q_i^{(k)}}^{p_i^{(k)}} u_{i+l}^{(0)} + \lambda \left[\delta_{i+p_i^{(k)}+\frac{1}{2}}(\mathbf{u}^{(k)}) - \delta_{i-q_i^{(k)}-\frac{1}{2}}(\mathbf{u}^{(k)})\right] \\
&\quad + \beta_{i-q_i^{(k)}-\frac{1}{2}}\delta_{i-q_i^{(k)}-\frac{1}{2}}(\mathbf{u}^{(k)}) - \beta_{i+p_i^{(k)}+\frac{1}{2}}\delta_{i+p_i^{(k)}+\frac{1}{2}}(\mathbf{u}^{(k)}) \\
&= \sum_{l=-q_i^{(k)}}^{p_i^{(k)}} u_{i+l}^{(0)} + \lambda \left[\delta_{i+p_i^{(k)}+\frac{1}{2}}(\mathbf{u}^{(k)}) - \delta_{i-q_i^{(k)}-\frac{1}{2}}(\mathbf{u}^{(k)})\right].
\end{aligned}$$

Since $u_{i+l}(\lambda) = u_i(\lambda)$ for $l = -q_i^{(k)}, \ldots, p_i^{(k)}$, c.f. (2.17), we have

$$\begin{aligned}
u_i(\lambda) &= \frac{\sum_{l=-q_i^{(k)}}^{p_i^{(k)}} u_{i+l}^{(0)}}{p_i^{(k)} + q_i^{(k)} + 1} + \lambda \frac{\delta_{i+p_i^{(k)}+\frac{1}{2}}(\mathbf{u}^{(k)}) - \delta_{i-q_i^{(k)}-\frac{1}{2}}(\mathbf{u}^{(k)})}{p_i^{(k)} + q_i^{(k)} + 1} \\
&= a_i^{(k)} + \lambda \alpha_i^{(k)}
\end{aligned} \tag{2.19}$$

for $i = 1, \ldots, n$ as desired. To prove (2.11), we simply note that by putting $\lambda = \hat{\lambda}^{(k)}$ in (2.10), we have $a_i^{(k)} = u_i^{(k)} - \hat{\lambda}^{(k)} \alpha_i^{(k)}$. ◻

**2.3. Critical Values of $\lambda$.** Next, we compute the maximum value $\hat{\lambda}^{(k+1)}$ of $\lambda^{(k+1)}$ for which Lemmas 2.3 and 2.4 hold. Thus, $\hat{\lambda}^{(k+1)}$ is the critical value at which the solution has structural changes compared to $\mathbf{u}^{(k)}$.

The idea is to compute the smallest value of $\lambda$ ($\lambda > \lambda^{(k)}$) such that the difference between a pair of adjacent values changes sign. This idea is illustrated in Figure 2.1 where we observe that a pair of adjacent values merge with one another.

LEMMA 2.5. *Let*

$$I(\mathbf{u}^{(k)}) := \left\{ \frac{u_{i+1}^{(k)} - u_i^{(k)}}{\alpha_i^{(k)} - \alpha_{i+1}^{(k)}} + \hat{\lambda}^{(k)} : \alpha_i^{(k)} \neq \alpha_{i+1}^{(k)},\ u_{i+1}^{(k)} \neq u_i^{(k)} \right\}. \tag{2.20}$$

*Then, the maximum value $\hat{\lambda}^{(k+1)}$ of $\lambda^{(k+1)}$ for which Lemmas 2.3 and 2.4 hold is given by*

$$\hat{\lambda}^{(k+1)} = \min I(\mathbf{u}^{(k)}). \tag{2.21}$$

*If $I(\mathbf{u}^{(k)}) = \phi$, then $\min I(\mathbf{u}^{(k)}) := +\infty$.*

*Proof.* To prove this lemma, we divide the proof into several steps.

<u>Step 1.</u> *If $\mathbf{u}^{(k)}$ is a constant vector, then $I(\mathbf{u}^{(k)}) = \phi$ (empty set) and $\hat{\lambda}^{(k+1)} = +\infty$.*

If $\mathbf{u}^{(k)}$ is a constant vector, then $\mathbf{u}(\lambda) = \mathbf{u}^{(k)}$ for all $\lambda \geq \hat{\lambda}^{(k)}$. Thus, $\hat{\lambda}^{(k+1)} = +\infty$. It is clear that $I(\mathbf{u}^{(k)}) = \phi$ since $u_{i+1}^{(k)} = u_i^{(k)}$ for all $i$.

<u>Step 2.</u> *If $\mathbf{u}^{(k)}$ is a nonconstant vector, then $I(\mathbf{u}^{(k)}) \neq \phi$.*

Since $\mathbf{u}^{(k)}$ is a nonconstant vector, there exists a smallest positive integer $i^*$ such that $u_{i^*+1}^{(k)} \neq u_{i^*}^{(k)}$. In fact, $i^* \equiv 1 + p_1^{(k)}$. We first note that $p_{i^*}^{(k)} \equiv q_{i^*+1}^{(k)} \equiv 0$ and $i^* - q_{i^*}^{(k)} - 1/2 \equiv 1/2$. By (2.9) and $\delta_{1/2}^{(k)} = 0$, we have

$$\alpha_{i^*}(\mathbf{u}^{(k)}) = \frac{\delta_{i^* + p_{i^*}^{(k)} + \frac{1}{2}}(\mathbf{u}^{(k)}) - \delta_{i^* - q_{i^*}^{(k)} - \frac{1}{2}}(\mathbf{u}^{(k)})}{p_{i^*}^{(k)} + q_{i^*}^{(k)} + 1} = \frac{\delta_{i^* + \frac{1}{2}}(\mathbf{u}^{(k)})}{q_{i^*}^{(k)} + 1}$$

and

$$\begin{aligned} \alpha_{i^*+1}(\mathbf{u}^{(k)}) &= \frac{\delta_{i^* + p_{i^*+1}^{(k)} + \frac{3}{2}}(\mathbf{u}^{(k)}) - \delta_{i^* - q_{i^*+1}^{(k)} + \frac{1}{2}}(\mathbf{u}^{(k)})}{p_{i^*+1}^{(k)} + q_{i^*+1}^{(k)} + 1} \\ &= \frac{\delta_{i^* + p_{i^*+1}^{(k)} + \frac{3}{2}}(\mathbf{u}^{(k)}) - \delta_{i^* + \frac{1}{2}}(\mathbf{u}^{(k)})}{p_{i^*+1}^{(k)} + 1}. \end{aligned}$$

As a result, we have $\alpha_{i^*}(\mathbf{u}^{(k)}) \neq \alpha_{i^*+1}(\mathbf{u}^{(k)})$ for any $\delta_{i^*+1/2}(\mathbf{u}^{(k)}) \in \{\pm 1\}$ and for any $\delta_{i^* + p_{i^*+1}^{(k)} + 3/2}(\mathbf{u}^{(k)}) \in \{\pm 1\}$. Together with $u_{i^*+1}^{(k)} \neq u_{i^*}^{(k)}$, we have $I(\mathbf{u}^{(k)}) \neq \phi$.

<u>Step 3.</u> *If $\mathbf{u}^{(k)}$ is a nonconstant vector, then $\hat{\lambda}^{(k+1)} = \min I(\mathbf{u}^{(k)})$ and $\hat{\lambda}^{(k+1)} < +\infty$.*

In view of Lemma 2.3, this lemma is equivalent to finding the largest possible value $\hat{\lambda}^{(k+1)}$ of $\lambda^{(k+1)}$ such that $\mathrm{sgn}[u_{i+1}(\lambda) - u_i(\lambda)] = \mathrm{sgn}[u_{i+1}^{(k)} - u_i^{(k)}]$ for any $\lambda \in [\hat{\lambda}^{(k)}, \hat{\lambda}^{(k+1)})$. Let $\hat{\lambda}^{(k+1)} \leq +\infty$ be the maximum value of $\lambda^{(k+1)}$ such that Lemma 2.3 holds. Note that Lemma 2.3 guarantees that $\hat{\lambda}^{(k+1)} > \hat{\lambda}^{(k)}$. For each $i$, we consider three different cases.

*Case (i).* $u_{i+1}^{(k)} = u_i^{(k)}$.

By Lemma 2.3, we obtain $\operatorname{sgn}[u_{i+1}(\lambda) - u_i(\lambda)] = \operatorname{sgn}(u_{i+1}^{(k)} - u_i^{(k)}) = 0$ for any $\lambda \in [\hat{\lambda}^{(k)}, \hat{\lambda}^{(k+1)})$. As a result, such $i$ puts no constraint on $\hat{\lambda}^{(k+1)}$.

*Case (ii).* $\alpha_{i+1}(\mathbf{u}^{(0)}) = \alpha_i(\mathbf{u}^{(0)})$.

In this case, we must have $\delta_{i-q_i^{(k)}-1/2} = \delta_{i+1/2} = \delta_{i+p_{i+1}^{(k)}+3/2}$. Then, according to (2.11), we have $u_i(\lambda) = u_i^{(k)}$ and $u_{i+1}(\lambda) = u_{i+1}^{(k)}$ for any $\lambda \in [\hat{\lambda}^{(k)}, \hat{\lambda}^{(k+1)})$. Thus, $\operatorname{sgn}[u_{i+1}(\lambda) - u_i(\lambda)] = \operatorname{sgn}[u_{i+1}^{(k)} - u_i^{(k)}] = 0$ for any $\lambda \in [\hat{\lambda}^{(k)}, \hat{\lambda}^{(k+1)})$ and for any $\hat{\lambda}^{(k+1)} > \hat{\lambda}^{(k)}$. As a result, such $i$ puts no constraint on $\hat{\lambda}^{(k+1)}$.

*Case (iii).* $u_{i+1}^{(k)} \neq u_i^{(k)}$ *and* $\alpha_{i+1}(\mathbf{u}^{(k)}) \neq \alpha_i(\mathbf{u}^{(k)})$.

By (2.11), $u_{i+1}(\lambda) = u_i(\lambda)$ if and only if $\lambda = \frac{u_{i+1}^{(k)} - u_i^{(k)}}{\alpha_i(\mathbf{u}^{(k)}) - \alpha_{i+1}(\mathbf{u}^{(k)})} + \hat{\lambda}^{(k)}$. So, $\operatorname{sgn}[u_{i+1}(\lambda) - u_i(\lambda)] = \operatorname{sgn}[u_{i+1}^{(k)} - u_i^{(k)}]$ for all $\lambda \in [\hat{\lambda}^{(k)}, \hat{\lambda}^{(k+1)})$ if and only if $\hat{\lambda}^{(k+1)} \leq \frac{u_{i+1}^{(k)} - u_i^{(k)}}{\alpha_i(\mathbf{u}^{(k)}) - \alpha_{i+1}(\mathbf{u}^{(k)})} + \hat{\lambda}^{(k)}$. Such condition must be satisfied for $i$ satisfying $u_{i+1}^{(k)} \neq u_i^{(k)}$ and $\alpha_{i+1}(\mathbf{u}^{(k)}) \neq \alpha_i(\mathbf{u}^{(k)})$. Hence, $\hat{\lambda}^{(k+1)} \leq \min I(\mathbf{u}^{(k)})$. Note that by the result in Step 2, $I(\mathbf{u}^{(k)}) \neq \phi$, and hence $\hat{\lambda}^{(k+1)} \leq \min I(\mathbf{u}^{(k)}) < +\infty$.

Combining the results in Cases (i)–(iii), the only restriction on $\hat{\lambda}^{(k+1)}$ is $\hat{\lambda}^{(k+1)} \leq \min I(\mathbf{u}^{(k)})$. Thus, the largest possible value of $\hat{\lambda}^{(k+1)}$ is $\min I(\mathbf{u}^{(k)})$. ∎

We remark that Lemma 2.4 cannot be directly applied to $\lambda = \hat{\lambda}^{(k+1)}$ because the KKT condition (2.14) does not hold in this case. However, by using continuity of $\mathbf{u}(\lambda)$ with respect to $\lambda$, the following lemma extends Lemma 2.4 to the case when $\lambda = \hat{\lambda}^{(k+1)}$.

Lemma 2.6. *Suppose $\hat{\lambda}^{(k+1)}$ obtained by (2.21) is finite. Then, the solution to (1.2) with $\lambda = \hat{\lambda}^{(k+1)}$ is given by (2.10).*

*Proof.* This follows from the fact that $\mathbf{u}(\lambda)$ is continuous in $\lambda$, i.e.,

$$u_i(\hat{\lambda}^{(k+1)}) = \lim_{\lambda \to \hat{\lambda}^{(k+1)}} u_i(\lambda) = \lim_{\lambda \to \hat{\lambda}^{(k+1)}} a_i^{(k)} + \lambda \alpha_i^{(k)} = a_i^{(k)} + \hat{\lambda}^{(k+1)} \alpha_i^{(k)}. \qquad ∎$$

In the above lemma, $\operatorname{sgn}[u_{i+1}(\hat{\lambda}^{(k+1)}) - u_i(\hat{\lambda}^{(k+1)})] \neq \operatorname{sgn}(u_{i+1}^{(k)} - u_i^{(k)})$ for those $i$'s which minimize (2.21). In fact, for these $i$'s, the value $\operatorname{sgn}[u_{i+1}(\hat{\lambda}^{(k+1)}) - u_i(\hat{\lambda}^{(k+1)})]$ changes from $\pm 1$ to 0. Thus, the values $\operatorname{sgn}[u_{i+1}(\hat{\lambda}^{(k+1)}) - u_i(\hat{\lambda}^{(k+1)})]$'s can be obtained by updating $\operatorname{sgn}[u_{i+1}(\hat{\lambda}^{(k)}) - u_i(\hat{\lambda}^{(k)})]$'s accordingly.

**2.4. Maximum Number of Critical $\lambda$'s.** In view of Lemmas 2.4–2.6, given an observed signal $\mathbf{u}^{(0)}$, we may generate a sequence of solutions $\{\mathbf{u}^{(k)}\}_{k=0}^K$ and the corresponding critical parameters $\{\hat{\lambda}^{(k)}\}_{k=0}^K$ such that $0 = \hat{\lambda}^{(0)} < \ldots < \hat{\lambda}^{(k)} < \ldots < \hat{\lambda}^{(K)}$ for some $K \geq 0$. It remains to establish the fact that given a $\lambda > 0$, there exists a unique $1 \leq K \leq n-1$ such that $\hat{\lambda}^{(K-1)} < \lambda \leq \hat{\lambda}^{(K)}$. Thus, by redefining $\hat{\lambda}^{(K)} := \lambda$, the sequence of solutions converges to $\mathbf{u}(\lambda)$ and contains at most $n-1$ elements (solutions) excluding $\mathbf{u}^{(0)}$.

Let $E(\mathbf{u})$ be the number of adjacent pairs of entries of $\mathbf{u}$ having equal value. It is obvious that $0 \leq E(\mathbf{u}) \leq n-1$ and that $\mathbf{u}$ is a constant vector if and only if $E(\mathbf{u}) = n-1$ for any $n$-vector $\mathbf{u}$. Our next result shows that $E(\mathbf{u}^{(k)})$ is strictly increasing in $k$ as long as $\mathbf{u}^{(k)}$ remains a nonconstant vector.

Lemma 2.7. *Let $E(\mathbf{u}) = |\{i : u_{i+1} = u_i,\ 1 \leq i \leq n-1\}|$. Suppose $E(\mathbf{u}^{(k)}) < n-1$, i.e., $\mathbf{u}^{(k)}$ is a nonconstant vector. Then,*

$$E(\mathbf{u}^{(k+1)}) \geq E(\mathbf{u}^{(k)}) + 1.$$

*Proof.* If $E(\mathbf{u}^{(k)}) < n-1$, then $I(\mathbf{u}^{(k)}) \neq \phi$ by Step 2 of the proof of Lemma 2.5 and $\hat{\lambda}^{(k+1)}$ defined by (2.21) is finite. By the construction of $\hat{\lambda}^{(k+1)}$, we have $u_{i+1}^{(k+1)} = u_i^{(k+1)}$ for those $i$'s which minimize (2.21) while $u_{i+1}^{(k)} \neq u_i^{(k)}$ (at least 1 such $i$ must exist by assumption). Thus, $E(\mathbf{u}^{(k+1)}) \geq E(\mathbf{u}^{(k)}) + 1$. ∎

COROLLARY 2.8. *The solution $\mathbf{u}^{(K)}$ is a constant vector for some $0 \leq K \leq n-1$.*

*Proof.* If $\mathbf{u}^{(k)}$ is not a constant vector for all $0 \leq k \leq n-2$, then by Lemma 2.7, $E(\mathbf{u}^{(n-1)}) \geq E(\mathbf{u}^{(0)}) + n - 1 \geq n-1$. Since $E(\mathbf{u}) \leq n-1$ for any $n$-vector $\mathbf{u}$, we must have $E(\mathbf{u}^{(n-1)}) = n-1$ implying that $\mathbf{u}^{(n-1)}$ is a constant vector. Thus, $\mathbf{u}^{(K)}$ is a constant vector for some $0 \leq K \leq n-1$. ∎

In the sequel, we define $\hat{\lambda}^{(K+1)} = \infty$ if $0 \leq K \leq n-1$ is the smallest integer such that $\mathbf{u}^{(K)}$ is a constant vector.

COROLLARY 2.9. *For any $\lambda > 0$, there exists a unique $1 \leq K \leq n-1$ such that $\hat{\lambda}^{(K)} < \lambda \leq \hat{\lambda}^{(K+1)}$. Moreover, the solution to (1.2) at $\lambda$ is given by*

$$u_i(\lambda) = u_i^{(K)} + (\lambda - \hat{\lambda}^{(K)})\alpha_i^{(K)}.$$

*Proof.* We simply note that the sequence $\{\hat{\lambda}^{(k)}\}$ is strictly increasing and that $\hat{\lambda}^{(K+1)} = \infty$ for some $1 \leq K \leq n-1$ by Corollary 2.8. Thus, there exists a unique $1 \leq K \leq n-1$ such that $\hat{\lambda}^{(K)} < \lambda \leq \hat{\lambda}^{(K+1)}$. By Lemma 2.4, the solution at $\lambda$ is given by (2.11). ∎

In view of the above corollary, for a given $\lambda \geq 0$, we can construct a sequence of solutions $\{\mathbf{u}^{(k)}\}_{k=0}^{K}$ with $\hat{\lambda}^{(K)} = \lambda$ for some $0 \leq K \leq n-1$. We call such a technique (i.e. compute the final solution by iteratively increasing $\lambda$) *parameter marching.*

**2.5. Dealing with Non-differentiability of TV-norm.** TV minimization problems (1.1) possess singularities because the objective functions are not differentiable with respect to $u$ at flat regions (where $u_x(x) = 0$). Consider the Euler-Lagrange equation for (1.1) under homogeneous Neumann boundary conditions,

$$u(x) = u^{(0)}(x) + \lambda \left( \frac{u_x(x)}{|u_x(x)|} \right)_x \qquad \text{in } \Omega \tag{2.22}$$

$$\frac{\partial u}{\partial n} = 0 \qquad \text{on } \partial\Omega.$$

If $u_x(x) \neq 0$, then the curvature term $(u_x(x)/|u_x(x)|)_x = 0$. If $u_x(x) = 0$ (flat region), then the curvature term possesses a singularity at $x$. These intrinsic singularities also appear when we solve the minimization problem by the direct method presented above.

Our parameter marching technique deals with this problem gracefully by replacing a QPP by another. More precisely, each critical parameter $\hat{\lambda}^{(k)}$ is constructed such that flat regions are unchanged when $\lambda \in [\hat{\lambda}^{(k)}, \hat{\lambda}^{(k+1)})$. Thus, the $(k+1)$-th QPP is a differentiable problem with respect to $\lambda$ in its corresponding interval $[\hat{\lambda}^{(k)}, \hat{\lambda}^{(k+1)})$. But the $(k+1)$-th QPP is no longer differentiable at $\hat{\lambda}^{(k+1)}$. However, once we encounter such a singularity when solving the $(k+1)$-th QPP, we switch to the $(k+2)$-th QPP whose objective function and constraints are changed accordingly. As a result, the singularity problems are handled explicitly by a "jump in problems".

In Figure 2.2, we illustrate schematically the trace of the solutions $\mathbf{u}(\lambda)$ in the phase space for $\lambda \in [0, \infty)$. The dotted line is defined by $\{\mathbf{u} \in \mathbb{R}^n : u_1 = \ldots = u_n\}$

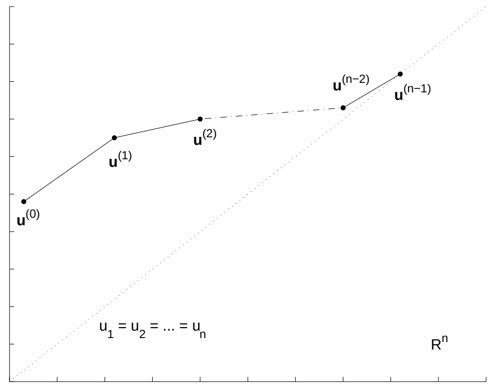


FIG. 2.2. *The trace of the solution* $\mathbf{u}(\lambda)$ *in the phase space for* $\lambda \in [0, \infty)$*. The dotted line represents the line* $u_1 = u_2 = \ldots = u_n$*. Each critical* $\lambda$ *induces a turn in the trace. The trace is a straight line in-between two consecutive critical* $\lambda$*'s. Note that* $\mathbf{u}^{(n-1)}$ *is a constant vector and that* $\mathbf{u}(\lambda) = \mathbf{u}^{(n-1)}$ *for all* $\lambda \geq \hat{\lambda}^{(n-1)}$*.*

on which $\mathbf{u}(\lambda)$ is a constant vector. We note that the trace is a continuous piecewise linear line in $\mathbb{R}^n$ with at most $n-1$ corners. Thus, $\mathbf{u}(\lambda)$ changes smoothly almost for all $\lambda$ except at the corners. Our method explicitly computes the $\lambda$'s ($\{\hat{\lambda}^{(k)}\}$) at which $\mathbf{u}(\lambda)$ is not differentiable, and hence computing $\mathbf{u}^{(k+1)}$ from $\mathbf{u}^{(k)}$ is easily done.

**3. Causality and Dynamics of the Exact Solutions $\mathbf{u}^{(k)}$.** In this section, we investigate the associated causality of TV denoising problems. We also present some further properties of the sequence of solutions $\mathbf{u}^{(k)} \equiv \mathbf{u}(\lambda^{(k)})$ generated by the formulae in Lemma 2.4. These properties are relatively easy to derive as exact solutions are available.

The next proposition proves that $\lambda$ resembles the role of the time variable in memoryless time-dependent problems (i.e. causality) in the sense that given the solution $\mathbf{u}(\lambda_0)$ at some $\lambda_0$, the solution at $\lambda > \lambda_0$ depends only on the solution at $\lambda_0$ but not on solutions at parameters less than $\lambda_0$. Thus, to obtain the solution at $\lambda > \lambda_0$, we may solve (1.2) with the observed signal $\mathbf{u}^{(0)}$ replaced by $\mathbf{u}(\lambda_0)$ and $\lambda$ replaced by $\lambda - \lambda_0$. In other words, if we let $T_\lambda$ be an operator which maps a signal $\mathbf{u}^{(0)}$ to its denoised version $\mathbf{u}(\lambda)$ with respect to the parameter $\lambda$, then we have $T_{\lambda_1+\lambda_2} = T_{\lambda_1} \circ T_{\lambda_2}$, c.f. [2]. Such a property further justifies our idea of marching the regularization parameter.

PROPOSITION 3.1. *Given the solution* $\mathbf{u}^{(k)}$ *of (1.2) at* $\hat{\lambda}^{(k)}$ *and given* $\hat{\lambda}^{(k+1)}$ *calculated by (2.21). The solution* $\mathbf{u}(\lambda)$ *of (1.2) for any* $\lambda \in [\hat{\lambda}^{(k)}, \hat{\lambda}^{(k+1)}]$ *can be obtained by solving*

$$\begin{aligned} (3.1) \qquad & \min_{\mathbf{u}\in\mathbb{R}^n} \quad \frac{1}{2}\sum_{i=1}^{n}\left(u_i - u_i^{(k)}\right)^2 + \mu\sum_{i=1}^{n-1}\delta_{i+\frac{1}{2}}(\mathbf{u}^{(k)})(u_{i+1}-u_i) \\ & \textit{subject to} \quad \delta_{i+\frac{1}{2}}(\mathbf{u}^{(k)})(u_{i+1}-u_i) \geq 0 \quad \textit{if} \quad \operatorname{sgn}[u_{i+1}^{(k)} - u_i^{(k)}] \neq 0 \\ & \qquad\qquad\quad \delta_{i+\frac{1}{2}}(\mathbf{u}^{(k)})(u_{i+1}-u_i) = 0 \quad \textit{if} \quad \operatorname{sgn}[u_{i+1}^{(k)} - u_i^{(k)}] = 0 \end{aligned}$$

*where* $\mu := \lambda - \hat{\lambda}^{(k)}$*.*

*Proof.* If we treat $\mathbf{u}^{(k)}$ as the observed signal, then Corollary 2.9 says that there exists a $\hat{\mu}^{(k+1)} > 0$ such that the solution to (3.1) is given by

$$u_i(\mu) = u_i^{(k)} + \mu\alpha_i^{(k)} \tag{3.2}$$

for any $\mu \in [0, \hat{\mu}^{(k+1)}]$ where

$$\hat{\mu}^{(k+1)} = \min\left\{ \frac{u_{i+1}^{(k)} - u_i^{(k)}}{\alpha_i^{(k)} - \alpha_{i+1}^{(k)}} \mid \alpha_i^{(k)} \neq \alpha_{i+1}^{(k)},\ u_{i+1}^{(k)} \neq u_i^{(k)} \right\} \tag{3.3}$$

and $\alpha_i^{(k)}$ is determined by (2.9). Comparing (3.2) with (2.11), we have $\mu = \lambda - \hat{\lambda}^{(k)}$ for any $\lambda \in [\hat{\lambda}^{(k)}, \hat{\lambda}^{(k+1)}]$. □

One of the properties of continuous TV minimization problems (1.1) is that the mean value of the solution for each $\lambda \geq 0$ is preserved, i.e., $\int_\Omega u(x,\lambda)dx = \int_\Omega u^{(0)}(x)dx$ for all $\lambda \geq 0$, see [14]. The following proposition states that such a mean-preserving property is also true for the discrete case.

Proposition 3.2. *For any $\lambda \geq 0$,*

$$\frac{1}{n}\sum_{i=1}^{n} u_i(\lambda) \equiv \frac{1}{n}\sum_{i=1}^{n} u_i^{(0)}.$$

*Proof.* Suppose $\lambda \in [\hat{\lambda}^{(k)}, \hat{\lambda}^{(k+1)})$ for some $k \geq 0$. Summing up the formula in (2.15) for $i = 1, \ldots, n$ implies that

$$\sum_{i=1}^{n} u_i(\lambda) \equiv \sum_{i=1}^{n} a_i^{(k)} \equiv \sum_{i=1}^{n} u_i^{(k)}$$

for any $\lambda \in [\lambda^{(k)}, \lambda^{(k+1)}]$. Inductively, we have

$$\sum_{i=1}^{n} u_i(\lambda) \equiv \sum_{i=1}^{n} u_i^{(0)}.$$ □

It is obvious that the solution $\mathbf{u}(\lambda)$ becomes smoother and smoother as $\lambda$ increases; because the total variation decreases as $\lambda$ increases. A less obvious question is how $\mathbf{u}(\lambda)$ changes with $\lambda$ exactly. Our next proposition provides some insight to answer this question. We remark that Chan and Strong [15] gave analogous results to continuous TV denoising problems.

Proposition 3.3. *Suppose $\lambda \in [\hat{\lambda}^{(k)}, \hat{\lambda}^{(k+1)}]$ and $u_i^{(k)}$ lies in a flat region $R_i^{(k)}$ with grid points $(i - q_i^{(k)}, \ldots, i, \ldots, i + p_i^{(k)})$. Suppose also that $\mathbf{u}^{(k)}$ is a non-constant vector. Denote the length of $R_i^{(k)}$ by $|R_i^{(k)}| \equiv p_i^{(k)} + q_i^{(k)} + 1$. If $i$ is such that $i - q_i^{(k)} > 1$ and $i + p_i^{(k)} < n$, then*

$$u_i(\lambda) - u_i^{(k)} = (\lambda - \hat{\lambda}^{(k)}) \begin{cases} 0 & \text{if} \quad u_{i-q_i^{(k)}-1}^{(k)} < u_i^{(k)} < u_{i+p_i^{(k)}+1}^{(k)} \\ \frac{2}{|R_i^{(k)}|} & \text{if} \quad u_i^{(k)} < \min\left\{ u_{i-q_i^{(k)}-1}^{(k)}, u_{i+p_i^{(k)}+1}^{(k)} \right\} \\ \frac{-2}{|R_i^{(k)}|} & \text{if} \quad u_i^{(k)} > \max\left\{ u_{i-q_i^{(k)}-1}^{(k)}, u_{i+p_i^{(k)}+1}^{(k)} \right\} \\ 0 & \text{if} \quad u_{i-q_i^{(k)}-1}^{(k)} > u_i^{(k)} > u_{i+p_i^{(k)}+1}^{(k)} \end{cases}.$$

*If $i$ is such that $i - q_i^{(k)} = 1$, then*

$$u_i(\lambda) - u_i^{(k)} = (\lambda - \hat{\lambda}^{(k)}) \begin{cases} \frac{1}{|R_i^{(k)}|} & \text{if} \quad u_i^{(k)} < u_{i+p_i^{(k)}+1}^{(k)} \\ \frac{-1}{|R_i^{(k)}|} & \text{if} \quad u_i^{(k)} > u_{i+p_i^{(k)}+1}^{(k)} \end{cases} .$$

*If $i$ is such that $i + p_i^{(k)} = n$, then*

$$u_i(\lambda) - u_i^{(k)} = (\lambda - \hat{\lambda}^{(k)}) \begin{cases} \frac{1}{|R_i^{(k)}|} & \text{if} \quad u_i^{(k)} < u_{i-q_i^{(k)}-1}^{(k)} \\ \frac{-1}{|R_i^{(k)}|} & \text{if} \quad u_i^{(k)} > u_{i-q_i^{(k)}-1}^{(k)} \end{cases} .$$

*Proof.* The results are direct consequences of Lemma 2.4. ∎

Thus, if $\mathbf{u}^{(k)}$ is a nonconstant vector, then for $u_i^{(k)}$ in a region $R_i^{(k)}$ which does not contain any boundary point, the value $u_i(\lambda)$ remains unchanged if $\mathbf{u}^{(k)}$ is monotone in a neighborhood of $R_i^{(k)}$ and is increased (decreased) by $2(\lambda - \hat{\lambda}^{(k)})/|R_i^{(k)}|$ if $R_i^{(k)}$ is a local minima (maxima). For $u_i^{(k)}$ in a region $R_i^{(k)}$ which contains a boundary point, the value $u_i(\lambda)$ must be increased (decreased) by $(\lambda - \hat{\lambda}^{(k)})/|R_i^{(k)}|$ if $R_i^{(k)}$ is a local minima (maxima).

**4. Geometry of Minimizers.** In the previous sections, we derive algebraically exact solutions of TV denoising problems and their properties. In this section, we give their geometric interpretation in order to further understand the TV denoising model. In the sequel, the constant vector $(1, 1, \dots, 1)^T$ is denoted by $\mathbf{1}$.

To begin, we make an observation:

$$\|\mathbf{u}\|_{\mathrm{TV}} = \|\mathbf{u} + c\mathbf{1}\|_{\mathrm{TV}} \text{ for any } c \in \mathbb{R}.$$

Here, the TV norm is invariant in the $\mathbf{1}$ direction. From this observation, we notice that (a) the level set $\|\mathbf{u}\|_{\mathrm{TV}} = 0$ in the phase space is a straight line which passes through the origin with tangent vector $\mathbf{1}$; (b) the level set $\|\mathbf{u}\|_{\mathrm{TV}} = r$ for $r > 0$ in the phase space is the shell of a hyperrectangular tube which extends infinitely in the directions $\pm\mathbf{1}$. In Figure 4.1, we illustrate the level set $\|\mathbf{u}\|_{\mathrm{TV}} = r$ for $r > 0$ in the 3-dimensional case, i.e., $\mathbf{u} = (u_1, u_2, u_3)^T$.

A close inspection to Figure 4.1 reveals that the main reason leading to the mean-preserving property stated in Proposition 3.2 is that the TV-norm is invariant in the $\mathbf{1}$ direction. More precisely, on any straight line in $\mathbb{R}^n$ having a tangent vector $\mathbf{1}$, the point that is closest to $\mathbf{u}^{(0)}$ in Euclidean norm is the orthogonal projection of $\mathbf{u}^{(0)}$ onto the line. It means that any solution to the TV denoising problem must lie on the plane $\langle \mathbf{u} - \mathbf{u}^{(0)}, \mathbf{1} \rangle = 0$. Generally, any other regularization functional which is invariant in the $\mathbf{1}$ direction would posses the mean-preserving property. In the following proposition, we prove these facts based on geometric arguments.

Proposition 4.1. *Given an initial data $\mathbf{u}^{(0)}$, any solution $\mathbf{u}(\lambda)$ to the minimization problem (1.2) lies on the plane $\langle \mathbf{u}(\lambda) - \mathbf{u}^{(0)}, \mathbf{1} \rangle = 0$. Equivalently, the mean of $\mathbf{u}(\lambda)$ is preserved for all $\lambda \geq 0$.*

*Proof.* We first note that the TV-norm of the solution $\mathbf{u}(\lambda)$ is invariant in the $\mathbf{1}$ direction, i.e., $\|\mathbf{u}(\lambda) + c\mathbf{1}\|_{\mathrm{TV}} = \|\mathbf{u}(\lambda)\|_{\mathrm{TV}}$ for all $c \in \mathbb{R}$. Let $F(c) = \frac{1}{2}\|\mathbf{u}(\lambda) + c\mathbf{1} - \mathbf{u}^{(0)}\|_2^2$ be the value of the fitting term with data $\mathbf{u}(\lambda) + c\mathbf{1}$. Thus, a necessary condition for $\mathbf{u}(\lambda)$ to be a minimizer is that it gives the smallest value of the fitting

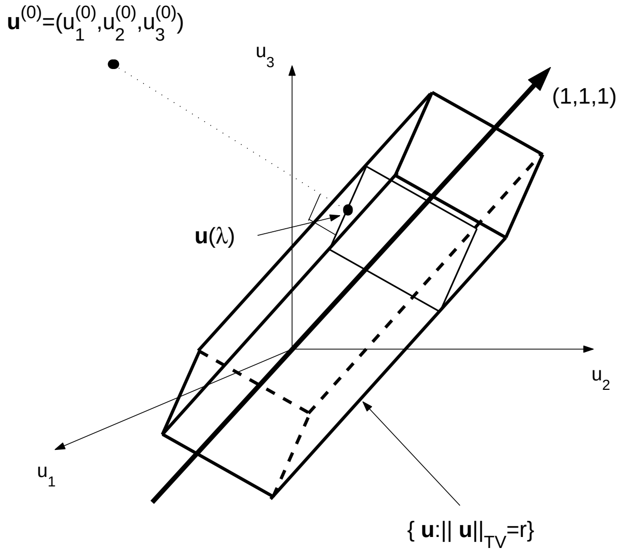


FIG. 4.1. *The level set $\|\mathbf{u}\|_{\mathrm{TV}} = r$ for some $r > 0$ in the 3-dimension case is the shell of a rectangular tube extending to infinity in the directions $\pm\mathbf{1}$. The black dot outside the tube denotes an initial data $\mathbf{u}^{(0)} = (u_1^{(0)}, u_2^{(0)}, u_3^{(0)})^T$ in the phase space. The black dot on the tube denotes a solution $\mathbf{u}(\lambda)$ whose TV-norm is $r$. The solution $\mathbf{u}(\lambda)$ must lie on the intersection of the level set $\|\mathbf{u}(\lambda)\|_{\mathrm{TV}} = r$ and the plane which passes through $\mathbf{u}^{(0)}$ and whose normal vector is $\mathbf{1}$, i.e., $\langle \mathbf{u} - \mathbf{u}^{(0)}, \mathbf{1}\rangle = 0$. This intersection is the rectangle located in the middle of the tube in the figure.*

term among the set $\{\mathbf{u} : \mathbf{u} = \mathbf{u}(\lambda) + c\mathbf{1}, c \in \mathbb{R}\}$, i.e., $F'(c = 0) = 0$. Hence, we have

$$\begin{aligned} \frac{dF(c)}{dc}_{|c=0} &= 0 \\ \langle \mathbf{u}(\lambda) + c\mathbf{1} - \mathbf{u}^{(0)}, \mathbf{1}\rangle_{|c=0} &= 0 \\ \langle \mathbf{u}(\lambda) - \mathbf{u}^{(0)}, \mathbf{1}\rangle &= 0. \end{aligned}$$

The above equation directly implies that the mean is preserved for any $\lambda \geq 0$. □

Next, we observe that the solution $\mathbf{u}(\lambda)$ must move in the direction of steepest descent of the TV-norm as $\lambda$ increases. In Figure 4.2, we show an example of the steepest descent direction of the TV-norm in the affine space

$$\{\mathbf{u} : \langle \mathbf{u}, \mathbf{u}^{(0)} - \mathbf{1}\rangle = 0\} \cap \{\mathbf{u} : \|\mathbf{u}\|_{\mathrm{TV}} = r, r > 0\}$$

where $\mathbf{u} = (u_1, u_2, u_3)^T$. In Figure 4.3, the traces of three different solutions in the phase space corresponding to three different initial data (having the same mean value) are shown. The corresponding pictures in the $u$-$x$ space are shown in Figure 4.4. We see from the figures that the traces follow the stream lines defined by the steepest descent directions with respect to the level sets of the TV-norm. Moreover, once the solution $\mathbf{u}(\lambda)$ reaches the plane $u_i = u_{i+1}$ for some $i$, it remains on this plane indefinitely. All solutions will reach the level set $\|\mathbf{u}\|_{\mathrm{TV}} = 0$ eventually (in finite "time").

Clearly from Figure 4.1, we observe that a necessary condition for optimality is that the hypersphere

$$\{\mathbf{u} \in \mathbb{R}^n : \|\mathbf{u} - \mathbf{u}^{(0)}\|_2 = \|\mathbf{u}(\lambda) - \mathbf{u}^{(0)}\|_2\}$$

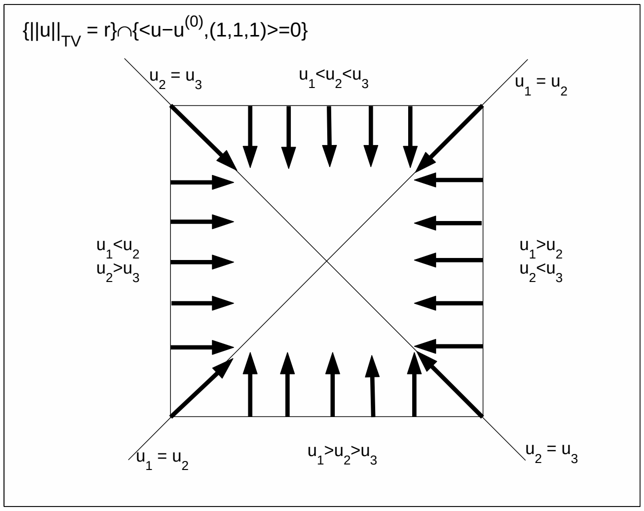


FIG. 4.2. *Steepest descent directions of the TV-norm at the intersection of the level set* $\|\mathbf{u}\|_{\mathrm{TV}} = r$ *and the plane* $\langle \mathbf{u} - \mathbf{u}^{(0)}, \mathbf{1}\rangle = 0$. *The intersection corresponds to the rectangle in the middle of the tube in Figure 4.1. As* $\lambda$ *increases, the solution* $\mathbf{u}(\lambda)$ *must follow the stream lines defined by the steepest descent directions at each level set of* $\|\mathbf{u}\|_{\mathrm{TV}}$.

and the level sets of $\|\cdot\|_{\mathrm{TV}}$ must touch each other at the optimal solution $\mathbf{u}(\lambda)$.

**5. Numerical Demonstration of the Analysis.** In this section, we numerically illustrate some results from our analysis. Our tests employ the following two functions defined on $[0, 1]$, see Figure 5.1,

$$\begin{aligned} f_1(x) &= \chi_{[0.1,0.3)}(x) + \chi_{[0.4,0.5)}(x) + \chi_{[0.6,0.65)}(x) \\ &\quad + \chi_{[0.75,0.775)}(x) + \chi_{[0.875,0.8875)}(x); \\ f_2(x) &= \chi_{[0,0.2)}(x) + (10x-2)\chi_{[0.2,0.4)}(x) + 2\chi_{[0.4,0.6)}(x) \\ &\quad + (5x-1)\chi_{[0.6,0.8)}(x) + (7-5x)\chi_{[0.8,1]}(x). \end{aligned}$$

For each of the above functions, we first construct a noisy vector $\mathbf{u}_l^{(0)}$ of size 4097 by adding $\mathbf{f}_l$ and $\eta_l$ together, i.e.,

$$\mathbf{u}_l^{(0)} := \mathbf{f}_l + \eta_l \qquad \text{for } l = 1, 2.$$

Here, $\mathbf{f}_l$ is obtained by uniformly sampling $f_l(x)$ at 4097 points on $[0, 1]$ and $\eta_l$ is a noise vector of size 4097 whose entries are independent and identically distributed with standard variation $\sigma = 10$, i.e., $\|\eta_l\| = 10$ for $l = 1, 2$. Next, each $\mathbf{u}_l^{(0)}$ is downsampled to form vectors $\mathbf{u}_{l,n}^{(0)}$ of size $n$ for $n = 17, 33, 65, 128, 257, 513$, i.e.,

$$u_{l,n,i}^{(0)} = f_l\left(\frac{i-1}{n-1}\right) + \eta_{l, \frac{4096(i-1)}{n-1}+1} \quad \text{for } i = 1, \ldots, n.$$

See Figure 5.1 for the pictures of $\mathbf{u}_{l,513}^{(0)}$ for $l = 1, 2$.

We observe in Figure 5.2 the two denoised signals. In Figure 5.2 (a) each restored signal feature exhibits some area loss in the sense that the area under the curve is diminished. This is particularly noticeable in the last feature which is defined on a small interval. We expect this type of behavior from the results in Proposition 3.3.

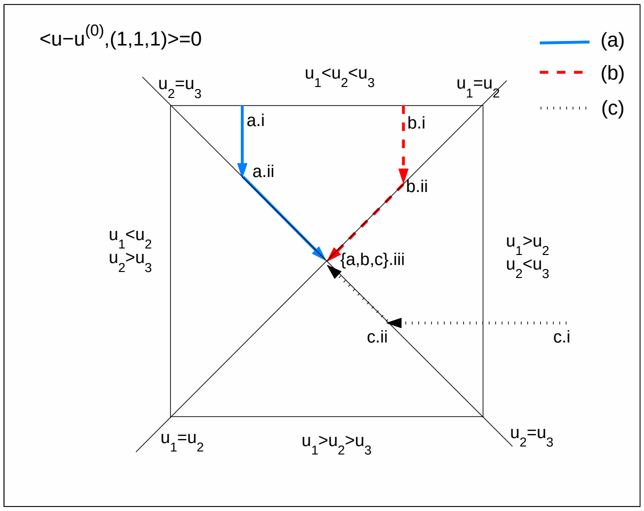


FIG. 4.3. *The 2-dimensional affine space consisting of the plane $\langle \mathbf{u} - \mathbf{u}^{(0)}, \mathbf{1} \rangle = 0$ intersected with the level sets of $\|\mathbf{u}\|_{\mathrm{TV}}$. From this, we may view the trace of solutions. The paths a, b, and, c corresponding to three different initial conditions a.i, b.i, and, c.i are shown (assuming the initial conditions have the same mean value). These paths follow the stream lines defined by the steepest descent directions of the TV-norm. Eventually, all solutions move to the level set $\|\mathbf{u}\|_{\mathrm{TV}} = 0$, i.e., they become a constant vector. The corresponding scenario of these paths in the u-x space is shown in Figure 4.4.*

In Figure 5.2 (b) there is a visible staircasing effect on the restored signal. Again, this effect is understood from our analysis of the one dimensional TV model where lemma 2.3 states that the solutions exhibit a monotonicity preserving property.

In Figure 5.3 we view the two sequences of critical $\lambda$'s that arise as denoising is performed on the two noisy signals $\mathbf{u}^{(0)}_{l,513}$ and $\mathbf{u}^{(0)}_{2,513}$ until each signal becomes constant. Clearly, most of the $\lambda$'s cluster around 0 corresponding to the high frequency components of noise. As the $\lambda$'s are optimally increasing, we see that there is more variation between them corresponding to the discernible features of the signal.

To further illustrate the behavior of the sequence of solutions that our method generates, we consider the following example. The original clean signal is $\mathbf{u}^{(0)} = (0,0,0,0,1,1,1,1,1)^T$ and the observed signal is generated by adding Gaussian noise to the original signal. The sequence of solutions is shown in Figure 5.4. In this example we note that one oscillation is removed after each iteration until the signal becomes a constant vector. Note how denoising is achieved in the exact fashion as outlined in Proposition 3.3. Moreover, once regions merge, they do not separate. We expect this type of behavior from our theory in Lemma 2.3.

## 6. Future Directions.

**6.1. Parameter Marching for the TV Model.** The causality behavior of $\lambda$ suggests that one may consider the variable $u$ in the Euler-Lagrange equation(2.22) as $u = u(x,\lambda)$ and modify (2.22) into a time-dependent problem of the form

$$
\begin{aligned}
F(x,\lambda,u,u_x,u_\lambda,u_{xx},u_{x\lambda},u_{\lambda\lambda},\ldots) &= 0 \qquad \text{in } \Omega \\
\frac{\partial u}{\partial n} &= 0 \qquad \text{on } \partial\Omega
\end{aligned} \tag{6.1}
$$

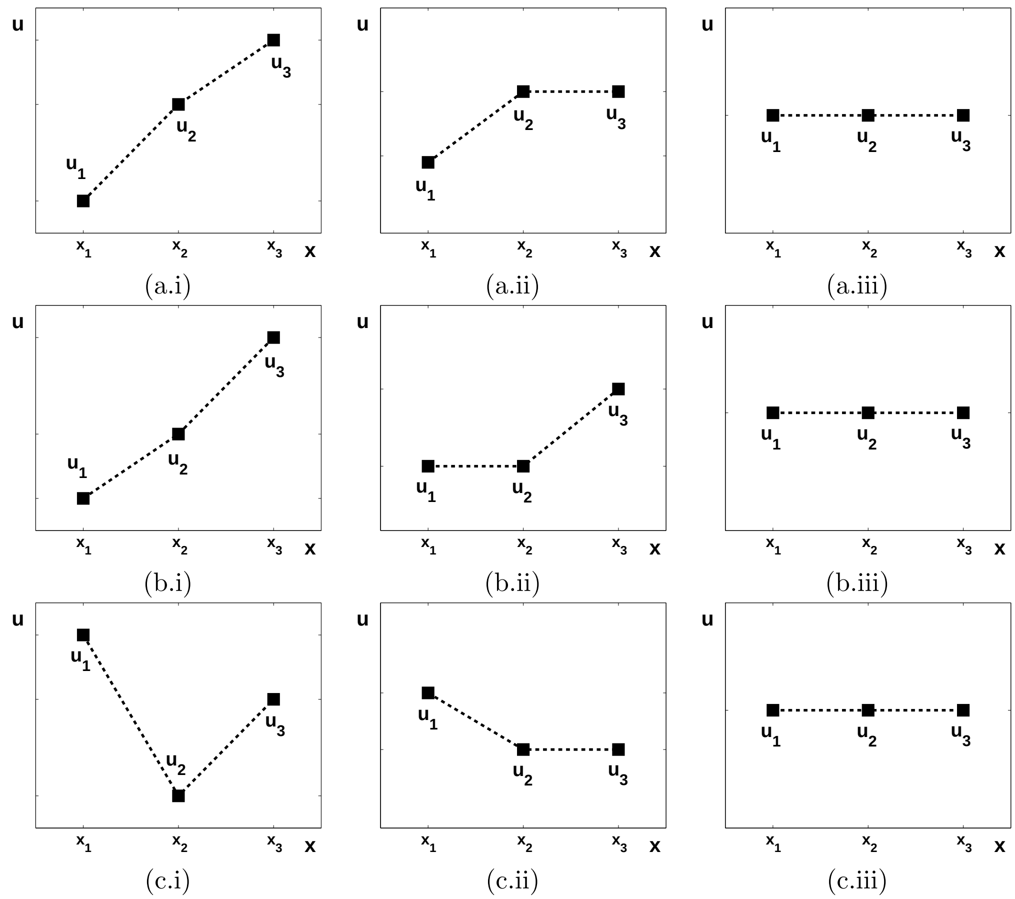


FIG. 4.4. *a.i–a.iii, b.i–b.iii, c.i–c.iii each correspond to traces a,b,c in Figure 4.3 respectively. In a.i the initial data is shown with $u_1 < u_2 < u_3$ and $|u_2 - u_3| < |u_2 - u_1|$. The corresponding initial location in Figure 4.3 is marked with "a.i". In a.ii the two points $u_2$ and $u_3$ have merged together corresponding to the trace hitting the line $u_2 = u_3$ in Figure 4.3, marked with "a.ii". The trace then follows this line until the line $u_1 = u_2 = u_3$ is reached, marked with "a.iii" in Figure 4.3. Similar results can be seen in b.i–b.iii and c.i–c.iii.*

$$u(x,0) = u^{(0)}(x) \qquad \text{in } \Omega$$

which can be solved efficiently by time-marching numerical PDE schemes. Of course such an approach is of practical interest since each iteration generates a meaningful solution to the problem. Intermediate iterates of artificial time marching methods and traditional iterative methods using a fixed $\lambda$ do not possess such a nice property. Moreover, insights into the role of $\lambda$ may be gained through the analysis of PDEs of the form (6.1).

**6.2. Improvements on the TV Model.** The most important feature of the TV model is its preservation of the discontinuities of solutions. However, the TV model does have some deficiencies. Contrast of signals is lost and in 2-d or higher dimensional cases, the geometry of the signal features may also be distorted.

Since the essential mechanisms of noise removal and discontinuity preservation are understood through our analysis of the dynamics of the solutions to 1-D TV de-

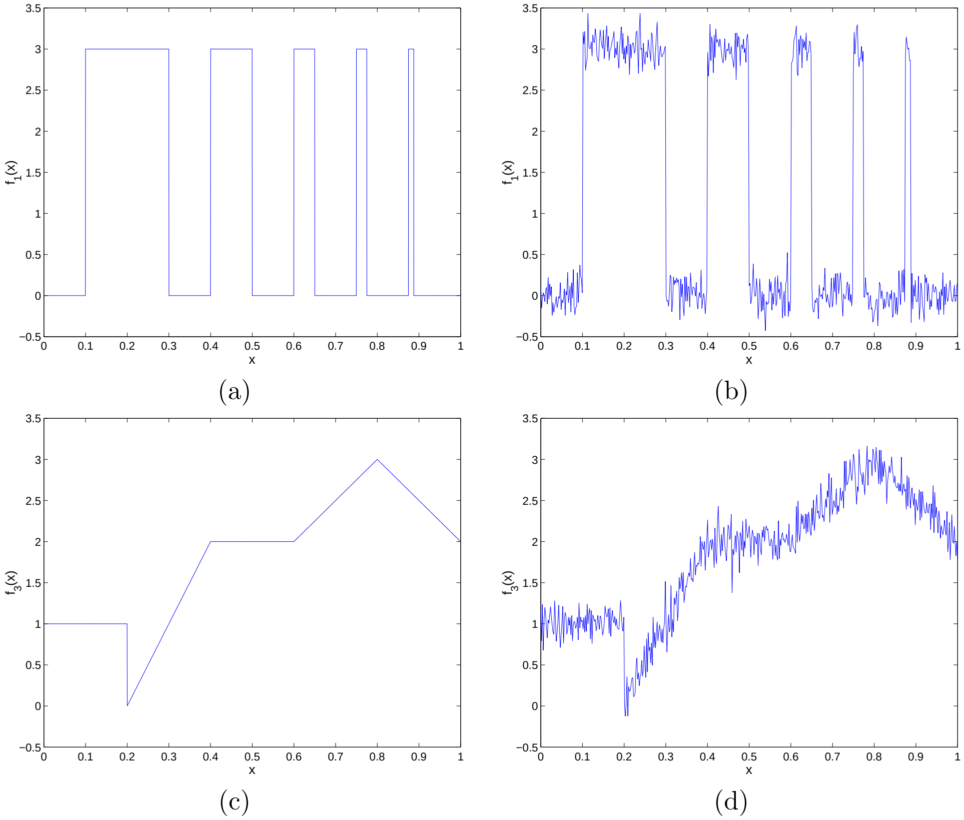


FIG. 5.1. *Test functions and their noisy version of size 513. (a) The noise-free function* $f_1(x)$*. (b) Noisy version of* $f_1(x)$ *with 513 sample points (*$\mathbf{u}_{1,513}^{(0)}$*). (c) The noise-free function* $f_2(x)$*. (d) Noisy version of* $f_2(x)$ *with 513 sample points (*$\mathbf{u}_{2,513}^{(0)}$*).*

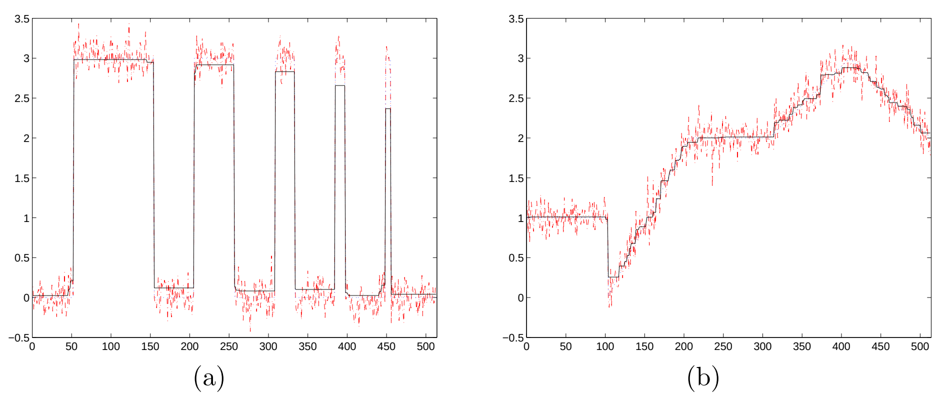


FIG. 5.2. *Denoised signals by the parameter marching algorithm. (a) Denosing* $\mathbf{u}_{1,513}^{(0)}$ *with* $\hat{\lambda}^{(488)} \approx 2.1590$*. Note the expected area loss for regions of the signal defined on smaller intervals. (b) Denosing* $\mathbf{u}_{2,513}^{(0)}$ *with* $\hat{\lambda}^{(451)} \approx 0.7678$*. Here we observe the staircasing effect caused by the monotonicity preserving property. Dashed, dash-dot and solid lines represent original, noisy and denoised signals respectively. These parameters are chosen by inspecting the denoised signals visually.*

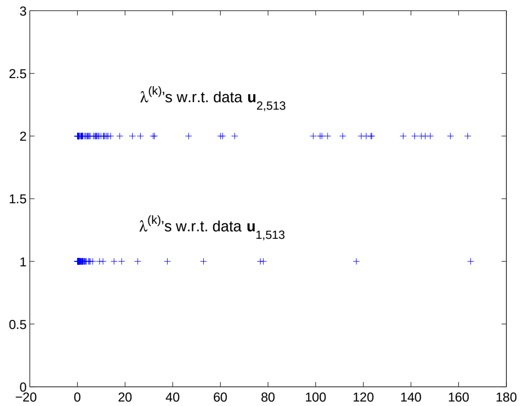


Fig. 5.3. *Optimal sequence of parameters $\{\hat{\lambda}^{(k)}\}$ of two data sets, $\mathbf{u}^{(0)}_{1,513}$ and $\mathbf{u}^{(0)}_{2,513}$. Since the magnitude of oscillations due to noise is smaller than that of the jumps of the original noise-free signals, most of the $\hat{\lambda}^{(k)}$'s are clustered around 0.*

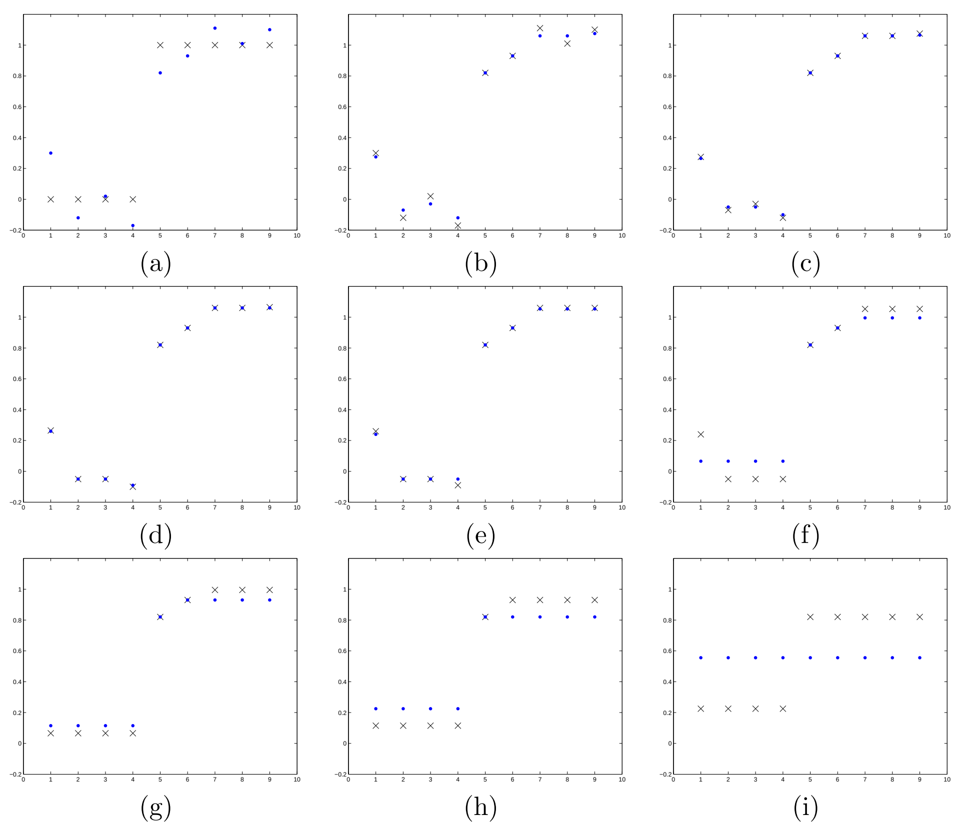


Fig. 5.4. *Sequence of solutions at $\{\hat{\lambda}^{(k)}\}$. (a) Original signal (cross), noise contaminated signal $\mathbf{u}^{(0)}$ (dot). (b)–(i) The solution $\mathbf{u}^{(k)}$ at $\hat{\lambda}^{(k)}$ (cross), the solution $\mathbf{u}^{(k+1)}$ at $\hat{\lambda}^{(k+1)}$ (dot) for $k = 0, 1, 2, \ldots, 7$ respectively. Notice that in each iteration, a pair of adjacent points, which corresponds to the current most obsolete oscillation, merge together, and will not separate in all subsequent iterations. In fact, in (b)–(i), the pairs $(u_7, u_8)$, $(u_2, u_3)$, $(u_8, u_9)$, $(u_3, u_4)$, $(u_1, u_2)$, $(u_6, u_7)$, $(u_5, u_6)$, $(u_4, u_5)$ merge together respectively. This confirms our theory in §2.*

noising problems as $\lambda$ varies, we could easily repair the above-mentioned deficiencies by modifying the dynamics of the solutions. Though it may not be easy, if not impossible, we may derive energy minimization principles from the solutions with modified dynamics. We are currently investigating the effectiveness of imposing monotonicity preserving constraints so that the geometry of desirable features is preserved as the denoising process evolves.

**6.3. Decomposition of Signals.** The existence of critical $\lambda$'s for the TV model suggests that one may decompose a signal according to the subspaces that the $\mathbf{u}^{(k)}$'s live in. In [16], Tadmor et al studied decomposition of signals induced by parameters $\{2^j\lambda_0\}_{j=0,1,2,\ldots}$ where $\lambda_0$ is chosen to capture the smallest oscillations in the noisy signal. In contrast to the results in [16], our analysis suggests that, even in the continuous setting, only the critical $\lambda$'s (which are data dependent and may be infinite in number in the continuous case) are useful for defining a scale space. Currently, such a decomposition and its applications to image compression and denoising are being studied [7].

**6.4. Parameter Marching for General Regularization Problems.** The parameter marching approach may also be applied to a more general class of Tikhonov-type regularization problems such as the deconvolution of blurred images. In this way, a sequence of solutions with respect to various (possibly critical) values of $\lambda$ is efficiently generated in one pass. A user may then select a final solution from this sequence without repeatedly solving the problem with different values of $\lambda$. The dynamics of the solutions of such regularization problems as the parameter $\lambda$ varies may also be thoroughly understood. Thus, the idea of marching the regularization parameter provides a novel framework for solving inverse problems formulated in the Tikhonov regularization setting where finding a reasonable value of $\lambda$ is generally difficult. Last but not least, a precise understanding of the role of $\lambda$ can be achieved in an elegant way.

**Acknowledgements.** The authors would like to thank Tony Chan, Selim Esedoglu and Cho Hong Min for their valuable comments and discussions.

## REFERENCES

[1] R. Acar and C. R. Vogel, *Analysis of bounded variation penalty methods for ill-posed problems*, Inverse Problems, 10 (1994), pp. 1217–1229.
[2] L. Alvarez, F. Guichard, P. Lions, and J. Morel, *Axioms and fundamental equations of image processing*, Arch. Rational Mech. Anal., 123 (1993), pp. 199–257.
[3] Antonin Chambolle, *An algorithm for total variation minimization and applications*, Ceremade technical report 0240, University of Paris IX, Dauphine, France, 2002.
[4] R. H. Chan, T. F. Chan, and W. L. Wan, *Multigrid for differential-convolution problems arising from image processing*, Proc. of the Workshop on Sci. Comput., G. Golub, S. H. Lui, F. Luk and R. Plemmons, eds., Springer-Verlag, 1997, pp. 58–72.
[5] T. F. Chan, G. H. Golub, and P. Mulet, *A nonlinear primal-dual method for total variation-based image restoration*, SIAM J. Sci. Comp., 20:6 (1999), pp. 1964–1977.
[6] T. F. Chan, S. Osher, and J. Shen, *The digital TV filter and nonlinear denoising*, IEEE Trans. Image. Processing, 10 (2001), pp. 231–242.
[7] T. F. Chan and F. Park, *Data dependent multi-scale total variation image decomposition and compression*, in preparation.
[8] T. F. Chan, H. M. Zhou, and R. H. Chan, *Continuation method for total variation denoising problems*, Proceedings to the SPIE Symposium on Advanced Signal Processing: Algorithms, Architectures, and Implementations, San Diego, CA, F. Luk, ed., vol. 2563, 1995, pp. 314–325.

[9] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Third ed., The Johns Hopkins University Press, Baltimore, MD, 1996.

[10] K. Lange, *Optimization Theory*, Lecture notes, Department of Biomathematics, University of California, Los Angeles, CA, 2002.

[11] M. Lysaker, S. Osher, and X. Tai, *Noise removal using smoothed normals and surface fitting*, UCLA CAM Report 03-03, Department of Mathematics, University of California, Los Angeles, CA, 2003.

[12] S. Osher, A. Sole, and L. Vese, *Image decomposition and restoration using total variation minimization and the $H^{-1}$ norm*, UCLA CAM Report 02-57, Department of Mathematics, University of California, Los Angeles, CA, 2002.

[13] E. Radmoser, O. Scherzer, and J. Weickert, *Scale-space properties of regularization methods*, Scale-space Theories in Computer Vision: Second International Conference, Corfu, Greece, 1999, Mads Nielsen et al, eds., Lecture Notes in Computer Science, vol. 1682, 1999, pp. 211–220, Springer.

Proceedings to the SPIE Symposium on Advanced Signal Processing: Algorithms, Architectures, and Implementations, San Diego, CA, F. Luk, ed., vol. 2563, 1995, pp. 314–325.

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

[15] D. M. Strong and T. F. Chan, *Edge-preserving and scale-dependent properties of total variation regularization*, UCLA CAM Report 00-38, Department of Mathematics, University of California, Los Angeles, CA, 2000.

[16] E. Tadmor, S. Nezzar, and L. Vese, *A multiscale image representation using hierarchical $(BV, L^2)$ decompositions*, UCLA CAM Report 03-32, Department of Mathematics, University of California, Los Angeles, CA, 2003.

[17] C. R. Vogel and M. E. Oman, *Iteration methods for total variation denoising*, SIAM J. Sci. Comp., 17 (1996), pp. 227–238.

[18] C. R. Vogel, *A multigrid method for total variation-based image denoising*, K. Bowers and J. Lund, eds., Computation and Control IV, Birkhauser, 1995.