Inverse Problems

PAPER

On the local and global minimizers of ℓ_0 gradient regularized model with box constraints for image restoration

To cite this article: Xue Feng *et al* 2018 *Inverse Problems* **34** 095007

IOP Publishing Inverse Problems

Inverse Problems **34** (2018) 095007 (34pp) https://doi.org/10.1088/1361-6420/aad1c5

On the local and global minimizers of ℓ_0 gradient regularized model with box constraints for image restoration

Xue Feng, Chunlin Wu[1] and Chao Zeng

School of Mathematical Sciences, Nankai University, Tianjin, People's Republic of China

E-mail: xuefeng@mail.nankai.edu.cn and wucl@nankai.edu.cn

Received 6 December 2017, revised 10 June 2018
Accepted for publication 6 July 2018
Published 23 July 2018

Abstract

Recently, nonconvex and nonsmooth models such as those using ℓ_0 'norm' have drawn much attention in the area of image restoration. This work investigates the local and global minimizers of the ℓ_0 gradient regularized model with box constraints. There are four major ingredients. Firstly, we show that the set of local minimizers can be represented by solutions to some quadratic problems, which are independent of the fidelity parameter α. Based on this, every point satisfying the first-order necessary condition is a local minimizer. Secondly, any two local minimizers have different energy values under certain assumptions, implying the uniqueness of the global minimizer. Thirdly, there exists a uniform lower bound for nonzero gradients of the restored images. Finally, we show that the global minimizer set is piecewise constant in terms of α, and when A is of full column rank and α is large enough, the distance between the true image and the restored images is bounded by the noise level. The numerical examples perfectly demonstrate our theoretical analysis.

Keywords: image restoration, ℓ_0 gradient regularization, box constraints, local minimizer, global minimizer

(Some figures may appear in colour only in the online journal)

[1] Author to whom any correspondence should be addressed.

1. Introduction

Image restoration is a typical inverse problem in image processing, including image deblurring, denoising, zooming, padding, etc. Without loss of generality, a real image can be regarded as a matrix $\underline{\mathbf{u}} \in \mathbf{R}^{n\times n}$. The observed image $\mathbf{f} \in \mathbf{R}^{m\times m}$ often contains various degradation such as noise and blur:

$$\mathbf{f} = \mathcal{A}\underline{\mathbf{u}} + \boldsymbol{\eta}, \tag{1}$$

where $\mathcal{A}$ is a linear operator, and $\boldsymbol{\eta} \in \mathbf{R}^{m\times m}$ denotes the noise. Image restoration is often viewed as a linear inverse problem to recover $\underline{\mathbf{u}}$ from $\mathbf{f}$.

In this paper, we consider the following image restoration model by minimizing an energy function with box constraints:

$$\begin{aligned} \min_{\mathbf{u}\in\mathbf{R}^{n\times n}} \quad & \frac{\alpha}{2}\|\mathcal{A}\mathbf{u} - \mathbf{f}\|_F^2 + \sum_{1\leqslant i,j\leqslant n} \varphi(\|(\boldsymbol{\nabla}\mathbf{u})[i,j]\|_2) \\ \text{s.t.} \quad & \underline{b}\cdot \mathbf{1}_{n\times n} \leqslant \mathbf{u} \leqslant \bar{b}\cdot \mathbf{1}_{n\times n}, \end{aligned} \tag{2}$$

where $\alpha > 0$; $\|\cdot\|_F$ represents the Frobenius norm; $\varphi(x) = 0$ if $x = 0$, otherwise $\varphi(x) = 1$; $\boldsymbol{\nabla}\mathbf{u} = (\boldsymbol{\nabla}_x\mathbf{u}, \boldsymbol{\nabla}_y\mathbf{u})$ with $\boldsymbol{\nabla}_x, \boldsymbol{\nabla}_y$ being the forward difference operators with specific boundary conditions; $\underline{b}, \bar{b}$ are bound parameters; $\mathbf{1}_{n\times n}$ is all one matrix in $\mathbf{R}^{n\times n}$. For digital images, since the pixel values lie in a certain range such as $[0, 1]$, box constraints can help to obtain a better restoration [3, 6, 7, 11]. When $\underline{b} = -\infty$ and $\bar{b} = +\infty$, this model becomes an unconstrained problem.

The first term $\|\mathcal{A}\mathbf{u} - \mathbf{f}\|^2$ in the energy function is the fidelity term, and the second one is the regularization term. The regularization term here is a composition of ℓ_0 'norm' and ℓ_2 norm to count the number of nonzero entries in $\boldsymbol{\nabla}\mathbf{u}$, called ℓ_0 gradient regularization. Usually, in image restoration models, the regularization terms are adopted to suppress the noise and preserve key features of the image. Many of the models utilized TV regularization [28], which are convex and thus have well-developed optimization methods to find their solutions. But now, a class of non-convex function based regularization terms, especially the ℓ_0 'norm' based ones, show their better performance on numerical experiments [5, 8, 10, 23, 24].

Model (2) is a general form of some existing models. With Markov random fields theory, Geman in 1984 and Besga in 1986 used it as a prior in MAP energy to restore labeled images, known as Potts prior model [4, 12, 31]. Later, it was applied successfully to reconstruct piecewise constant images and 3D tomographic images [17, 26]. And recently, with some efficient algorithms adopted, researchers showed its advantages in image denoising [25, 33] and image debluring [9, 26, 32, 34]. Except for counting nonzero gradients in (2), ℓ_0 'norm' is used frequently when sparsity is desired, covering very comprehensive fields such as signal processing, dictionary building, compressive sensing, machine learning, classification, morphologic component analysis, subset selection, and so on [1, 2, 11, 15, 19, 29, 37].

Since (2) is nonconvex and nonsmooth, most algorithms can only converge to one of its local minimizers. Thus it is worthy to study the local minimizers of (2), as well as the global minimizers. However, there are scarce results on this. Research on other nonconvex nonsmooth models are still developing [7, 20, 35], and all of their regularization terms are continuous. For ℓ_0 'norm' related models, most existing analyses focus on the sparse signal recovery problem, which is formulated as the minimization of least squares regularized with ℓ_0 'norm' [14, 21, 22, 36]. In [14], the author shows that when minimizing ℓ_0-type problems under constraints, every admissible point is a local minimizer. Specifically, [21] give a thorough investigation on the local and global minimizers of the regularized signal recovery model, including equivalence, uniqueness and so on. Later, in [22, 36], the authors discussed the relationships between the optimal solutions of the regularized signal recovery model and its different equivalent problems. Inspired by these enlightening findings and the importance of the ℓ_0 gradient regularization, we try to extend their results to model (2), and besides, develop some new ideas. In the process, the gradient operator and box constraints in model (2) bring considerable difficulties.

For the simplicity of description, we now rearrange the image data to be a array in $\mathbf{R}^N$ with N equal to the total number of pixels, i.e. $N = n^2$. Correspondingly, the observed image $\mathbf{f}$ is denoted by $f \in \mathbf{R}^M$ with $M = m^2$, and $\mathcal{A}$ is transformed to be a matrix $A \in \mathbf{R}^{M\times N}$. Thus, we can represent model (2) by

$$(\mathrm{P}) \qquad \begin{cases} \min\limits_{u\in\mathbf{R}^N} & \mathcal{F}(u) = \frac{\alpha}{2}\|Au - f\|^2 + \mathcal{R}(\nabla u) \\ \text{s.t.} & \underline{b}\cdot\mathbf{1}_N \leqslant u \leqslant \bar{b}\cdot\mathbf{1}_N, \end{cases} \tag{3}$$

where

$$\mathcal{R}(\nabla u) = \sum_{i=1}^{N}\varphi(\|(\nabla u)[i]\|_2),$$

$(\nabla u) = (\nabla_x u, \nabla_y u)$ is the discrete gradient (see the next section for details), $\mathbf{1}_N$ is all one vector in $\mathbf{R}^N$. Moreover, we denote the feasible region of (P) as

$$X := \{u \in \mathbf{R}^N \ : \ \underline{b}\cdot\mathbf{1}_N \leqslant u \leqslant \bar{b}\cdot\mathbf{1}_N\}. \tag{4}$$

In this paper, we study the local and global minimizers of (P). The main contributions of this paper are summarized below.

- We present an equivalent form of the local minimizer set of (P), by solving a set of quadratic problems which are independent of α. Based on this, we show that every point satisfying the first-order necessary condition of (P) is a local minimizer.
- We study the strict local minimizers of (P). Especially, we show that when A has full column rank and f is not in a zero measure set in $\mathbf{R}^M$, all (strict) local minimizers have different energy values. This also indicates the uniqueness of the global minimizer.
- We establish a uniform lower bound for nonzero gradients of the solutions to (P), implying that the ℓ_0 gradient regularization can generate neat edges.
- Without introducing new equivalent problems, we show that the global minimizer set of (P) with parameter α as the independent variable is piecewise constant in terms of α, and correspondingly, the optimal value is piecewise linear. Moreover, we find that when A has full column rank and α is large enough, the distance between the true image and the solutions of (P) with parameter α is bounded by the noise level.

The assumption that A has full column rank is a necessary condition for (P) to be well-posed, and holds generally for problems in image processing. In addition, most of our conclusions can be extended to more general models(see section 6). These theoretical results may contribute to a better understanding of this model, and can help to analyze the performance of existing algorithms to solve it.

The rest paper is organized as follows. In the next section, we give some basic notations. In section 3, we focus on the local minimizers of (P), especially on the strict local minimizers. In section 4, we show two important properties of the global minimizers: the lower bound for the nonzero gradients and the piecewise-constant dependency on α. In section 5, there are some numerical verifications of our main results. The paper is concluded in section 6.

2. Notations

Here we give some basic notations.

Without loss of generality, we represent an $n \times n$ gray image $\mathbf{u}$ by an $N \times 1$ vector u, where $N = n^2$, expanding in column. The ith entry of u reads as $u[i]$, and the corresponding mapping is

$$u[i] := \begin{cases} \mathbf{u}[i \bmod n, \lceil i/n \rceil], & 1 \leqslant i \bmod n \leqslant n-1, \\ \mathbf{u}[n, \lceil i/n \rceil], & i \bmod n = 0. \end{cases}$$

Then the discrete gradient operator is a mapping $\nabla : \mathbf{R}^N \to \mathbf{R}^N \times \mathbf{R}^N$. For $u \in \mathbf{R}^N$, ∇u is given by

$$(\nabla u)[i] := ((\nabla_x u)[i], (\nabla_y u)[i]),$$

where

$$(\nabla_x u)[i] := \begin{cases} u[i+n] - u[i], & 1 \leqslant \lceil i/n \rceil \leqslant n-1 \\ 0, & \lceil i/n \rceil = n, \end{cases}$$

$$(\nabla_y u)[i] := \begin{cases} u[i+1] - u[i], & 1 \leqslant i \bmod n \leqslant n-1 \\ 0, & i \bmod n = 0. \end{cases}$$

This definition is based on Neumann boundary condition, and our research in this paper also works on other boundary conditions.

Let K be any positive integer. The kth vector in the canonical basis of $\mathbf{R}^K$ is denoted by e_k, i.e. $e_k[l] = \delta_{kl}$, where δ_{kl} is the Kronecker delta. We use $\mathbf{1}_K$ to represent the all one vector in $\mathbf{R}^K$.

Given $u \in \mathbf{R}^K$ and $r > 0$, the open ball at u of radius r with respect to the ℓ_p norm for $1 \leqslant p \leqslant \infty$ reads as

$$B_p(u, r) := \{v \in \mathbf{R}^K : \|v - u\|_p < r\}.$$

To simplify the notation, the ℓ_2 norm in $\mathbf{R}^K$ is denoted by $\|\cdot\| := \|\cdot\|_2$.

We define two totally and strictly ordered index sets as:

$$\mathbb{I}_K := (\{1, \cdots, K\}, <), \quad \mathbb{I}_K^0 := (\{0, 1, \cdots, K\}, <),$$

where the symbol $<$ stands for the natural order of positive integers. That is, $\mathbb{I}_3 = \{1, 2, 3\}$, but not $\{2, 1, 1, 3\}$ or others. Accordingly, any subset $\omega \subseteq \mathbb{I}_K$ is also totally and strictly ordered. We define $\omega[k]$ as the kth element in ω, and then, $\omega[1]$ is the minimum index in ω. The complement of $\omega \subseteq \mathbb{I}_K$ is denoted by $\omega^c = \mathbb{I}_K \backslash \omega \subseteq \mathbb{I}_K$.

A partition of the set $\mathbb{I}_N$ is a grouping of all of its elements into subsets, in such a way that every element is included in one and only one of the subsets. The subsets in a partition are called blocks. To sort all of the blocks, we define the $<$ relation between any two subsets $\bar{\tau}, \tilde{\tau} \subseteq \mathbb{I}_N$ as:

$$\bar{\tau} < \tilde{\tau} \Longleftrightarrow \bar{\tau}[1] < \tilde{\tau}[1].$$

For example, if $\bar{\tau} = \{2, 5\}$ and $\tilde{\tau} = \{3, 4, 7\}$, then we have $\bar{\tau} < \tilde{\tau}$ as $2 < 3$. If $\{\tau_1, \cdots, \tau_l\}$ is a partition of $\mathbb{I}_K$ satisfying $\tau_k < \tau_{k+1}, \forall\, k \in \mathbb{I}_{l-1}$, then we say $(\{\tau_1, \cdots, \tau_l\}, <)$ is a strictly ordered partition.

Definition 2.1. For any $u \in \mathbf{R}^N$, the support of ∇u is defined as

$$\sigma(\nabla u) := \{i \in \mathbb{I}_N : (\nabla u)[i] \neq 0\}.$$

Table 1. The main notations.

(P)	The image restoration model defined in (3)
(Q_ω)	The constrained quadratic problem defined in (5)
$(\mathcal{Q}_\omega)$	The bounded-variable least square problem defined in (16)
(P_α)	Problem (P) with parameter α
$\mathcal{F}(u)$	$\frac{\alpha}{2}\|Au-f\|^2+\mathcal{R}(\nabla u)$
$\mathcal{I}_D(u)$	The indicator function of subset $D\subset\mathbf{R}^N$
X	$\{u\in\mathbf{R}^N\ :\ \underline{b}\cdot\mathbf{1}_N\leqslant u\leqslant\bar{b}\cdot\mathbf{1}_N\}$
C_ω	$\{u\in\mathbf{R}^N\ :\ (\nabla u)[i]=0,\forall\, i\in\omega^c\}$
$\mathbb{I}_K$	$(\{1,\cdots,K\},<)$
$\mathbb{I}_K^0$	$(\{0,1,\cdots,K\},<)$
$\sigma(\nabla u)$	$\{i\in\mathbb{I}_N\ :\ (\nabla u)[i]\neq 0\}$
$\mathcal{S}_\omega$	The block partition of $\mathbb{I}_N$ with respect to ω
u_ω	The block image of u with respect to ω
E_ω	The extension matrix defined in (13)
A_ω	AE_ω
$\mathcal{J}(\alpha)$	The optimal value of (P_α)
$\mathcal{U}^g(\alpha)$	The global minimizer set of (P_α)
$\mathcal{U}^l$	The local minimizer set of (P_α)
$\mathcal{U}^l_k$	$\mathcal{U}^l\cap\{u\in\mathbf{R}^N\ :\ \mathcal{R}(\nabla u)=k\}$
$\mathcal{C}_k$	$\arg\min_{u\in\mathcal{U}^l_k}\|Au-f\|^2$
c_k	$\min_{u\in\mathcal{U}^l_k}\|Au-f\|^2$
$\mathcal{J}_k(\alpha)$	$\frac{c_k}{2}\alpha+k$
$\phi(k,j)$	The intersection of $\mathcal{J}_k(\alpha)$ and $\mathcal{J}_j(\alpha)$

If $u=c\cdot\mathbf{1}_N$ for some $c\in\mathbf{R}$, we have $\sigma(\nabla u)=\emptyset$. More importantly, $\mathcal{R}(\nabla u)=\sharp\sigma(\nabla u),\forall\, u\in\mathbf{R}^N$.

In this work, we shall frequently refer to the following quadratic problem with a given $\omega\subseteq\mathbb{I}_N$:

$$(\mathrm{Q}_\omega)\qquad \begin{cases}\min_{u\in\mathbf{R}^N} & \|Au-f\|^2\\ \text{s.t.} & \underline{b}\cdot\mathbf{1}_N\leqslant u\leqslant\bar{b}\cdot\mathbf{1}_N,\\ & (\nabla u)[i]=0,\quad \forall\, i\in\omega^c.\end{cases}\tag{5}$$

Obviously, problem (Q_ω) is convex and proper, and always admits a solution. Specially, we denote

$$C_\omega:=\{u\in\mathbf{R}^N\ :\ (\nabla u)[i]=0,\forall\, i\in\omega^c\}.\tag{6}$$

Then, the feasible domain of (Q_ω) is $X\cap C_\omega$.

Finally, we recall some definitions about subdifferential [27].

Definition 2.2. Let $\psi:\mathbf{R}^N\to\mathbf{R}\cup\{+\infty\}$ be a proper lower semicontinuous function.

(a) The domain of ψ is defined by $\mathrm{dom}\psi:=\{u\in\mathbf{R}^N\ :\ \psi(u)<+\infty\}$.

(b) For each $u\in\mathrm{dom}\psi$, the Fréchet subdifferential of ψ at u, written $\hat\partial\psi(u)$, is the set of vectors $u^*\in\mathbf{R}^N$ which satisfy

$$\liminf_{\substack{v\to u\\ v\neq u}}\frac{1}{\|v-u\|}[\psi(v)-\psi(u)-\langle u^*,v-u\rangle]\geqslant 0.$$

If $u \notin \mathrm{dom}\psi$, then $\hat{\partial}\psi(u) = \emptyset$.

(c) The subdifferential of ψ at $u \in \mathrm{dom}\psi$, written $\partial\psi(u)$, is defined as follows:

$$\partial\psi(u) := \{u^* \in \mathbf{R}^N \ : \ \exists\, u_n \to u, \psi(u_n) \to \psi(u), \hat{\partial}\psi(u_n) \ni u_n^* \to u^*\}.$$

If $u \notin \mathrm{dom}\psi$, then $\partial\psi(u) = \emptyset$.

The following results, although elementary, is central to this paper.

Proposition 2.3 ([27]). *Let $\psi : \mathbf{R}^N \to \mathbf{R}$ be a proper function.*

(a) (Fermat's rule, p422) If ψ has a local minimum at u, then

$$0 \in \partial\psi(u).$$

(b) (p304) If $\psi = \psi_0 + \psi_1$ with ψ_0 finite at u and ψ_1 smooth on a neighborhood of u, then $\partial\psi(u) = \partial\psi_0(u) + \nabla\psi_1(u)$ where $\nabla\psi_1$ is the derivative of ψ_1.

Proposition 2.4 ([27], p 203, p 310). *Let D be a closed and nonempty subset. We denote by $\mathcal{I}_D$ its indicator function, i.e.*

$$\mathcal{I}_D(u) = \begin{cases} 0, & \text{if} \quad u \in D, \\ +\infty, & \text{otherwise.} \end{cases}$$

If D is convex, then

$$\partial\mathcal{I}_D(u) = \hat{\partial}\mathcal{I}_D(u) = \{u^* \in \mathbf{R}^N \ : \ \langle u^*, v - u\rangle \leqslant 0, \forall\, v \in D\}.$$

The main notations are listed in table 1.

3. The local minimizer

3.1. Representing the local minimizers of (P) by solutions of (Q_ω)

In this subsection, we will discuss the set of local minimizers of (P), and find its relationship with (Q_ω). The discontinuity of $\mathcal{R}(\nabla u)$ plays an important role here.

Lemma 3.1. *The function $\mathcal{R}(\nabla u)$ is lower semicontinuous. Specially, if $\bar{u} \in \mathbf{R}^N$, then there exists an open ball at $\bar{u}$, denoted by $B_\infty(\bar{u}, r_{\bar{u}})$, such that $\forall\, u \in B_\infty(\bar{u}, r_{\bar{u}})$, one of the following two cases holds:*

$$\text{(a)} \quad \sigma(\nabla u) = \sigma(\nabla\bar{u}) \Longleftrightarrow \mathcal{R}(\nabla u) = \mathcal{R}(\nabla\bar{u}) \Longleftrightarrow u \in B_\infty(\bar{u}, r_{\bar{u}}) \cap C_{\sigma(\nabla\bar{u})}, \tag{7}$$

$$\text{(b)} \quad \sigma(\nabla u) \supsetneqq \sigma(\nabla\bar{u}) \Longleftrightarrow \mathcal{R}(\nabla u) \geqslant \mathcal{R}(\nabla\bar{u}) + 1, \tag{8}$$

where $C_{\sigma(\nabla\bar{u})}$ is defined in (6).

Proof. Denote $\bar{\omega} := \sigma(\nabla\bar{u})$. If $\bar{u} = \bar{u}[1] \cdot \mathbf{1}_N$, then $\bar{\omega} = \emptyset$. The lemma is obvious for this case.

If $\bar{\omega} \neq \emptyset$, we denote $\bar{r} := \min\{\|(\nabla\bar{u})[i]\|, i \in \bar{\omega}\}$. Let $r_{\bar{u}} = \frac{\sqrt{2}}{4}\bar{r}$, and take an arbitrary $u \in B_\infty(\bar{u}, r_{\bar{u}})$. Then for any $i \in \bar{\omega}$,

$$
\begin{aligned}
\|(\nabla u)[i]-(\nabla \bar{u})[i]\|^2 &= \|(\nabla(u-\bar{u}))[i]\|^2 \\
&= \{(u-\bar{u})[i+1]-(u-\bar{u})[i]\}^2+\{(u-\bar{u})[i+n]-(u-\bar{u})[i]\}^2 \\
&< (2r_{\bar{u}})^2+(2r_{\bar{u}})^2 \\
&= \bar{r}^2.
\end{aligned}
$$

As $\|\nabla\bar{u}[i]\| \geqslant \bar{r}$, it follows that $\|(\nabla u)[i]\| \geqslant \|(\nabla\bar{u}[i])\| - \|(\nabla u)[i]-(\nabla\bar{u})[i]\| > \bar{r}-\bar{r}=0$. Thus,

$$(\nabla u)[i] \neq 0, \quad \forall\, i \in \bar{\omega},$$

which indicates that $\bar{\omega} \subseteq \sigma(\nabla u)$. Finally, the fact that $\bar{\omega}=\sigma(\nabla\bar{u})$ implies $\sigma(\nabla\bar{u}) \subseteq \sigma(\nabla u)$, followed by

$$\sigma(\nabla\bar{u})=\sigma(\nabla u), \quad \text{or} \quad \sigma(\nabla\bar{u}) \subsetneqq \sigma(\nabla u).$$

Recall that $\mathcal{R}(\nabla u)=\sharp\sigma(\nabla u)$. Meanwhile, if $u \in C_{\sigma(\nabla\bar{u})}$, by (6), one has $\mathcal{R}(\nabla u) \leqslant \mathcal{R}(\nabla\bar{u})$. Then (7) and (8) follows immediately. □

The lemma above helps to discuss the fidelity term and regularization term separately when we are checking if $\hat{u}$ is a local minimizer.

Proposition 3.2. *Given $\omega \subseteq \mathbb{I}_N$, suppose that $\bar{u}$ solves (Q_ω) in (5). Then $\bar{u}$ is a local minimizer of* (P)*, and $\sigma(\nabla\bar{u}) \subseteq \omega$.*

Proof. As $\bar{u}$ solves problem (Q_ω), the second constraint $(\nabla u)[i]=0, \forall\, i \in \omega^c$ in (Q_ω) entails $\sigma(\nabla\bar{u}) \subseteq \omega$. According to lemma 3.1, one has the intersection between the neighborhood $B_\infty(\bar{u}, r_{\bar{u}})$ of $\bar{u}$ and the feasible region X of (P) can be divided into two subsets: $X \cap B_\infty(\bar{u}, r_{\bar{u}}) = B_1 \cup B_2$, where

$$
\begin{aligned}
&B_1 = X \cap B_\infty(\bar{u}, r_{\bar{u}}) \cap \{u \in \mathbf{R}^N \;:\; \sigma(\nabla u)=\sigma(\nabla\bar{u})\}, \\
&B_2 = X \cap B_\infty(\bar{u}, r_{\bar{u}}) \cap \{u \in \mathbf{R}^N \;:\; \sigma(\nabla u) \supsetneqq \sigma(\nabla\bar{u})\}.
\end{aligned} \tag{9}
$$

Take an arbitrary $u \in B_1$. Since $\sigma(\nabla u)=\sigma(\nabla\bar{u}) \subseteq \omega$, it follows that $(\nabla u)[i]=0, \forall\, i \in \omega^c$. Thus, with $u \in X$, u is a feasible point of (Q_ω). As $\bar{u}$ solves (Q_ω), we have $\|A\bar{u}-f\|^2 \leqslant \|Au-f\|^2$. Moreover, applying (7) gives $\mathcal{R}(\nabla u)=\mathcal{R}(\nabla\bar{u})$. Hence, $\forall\, u \in B_1$,

$$\mathcal{F}(\bar{u}) = \frac{\alpha}{2}\|A\bar{u}-f\|^2+\mathcal{R}(\nabla\bar{u}) \leqslant \frac{\alpha}{2}\|Au-f\|^2+\mathcal{R}(\nabla u)=\mathcal{F}(u).$$

Take an arbitrary $u \in B_2$. We have $\mathcal{R}(\nabla u) \geqslant \mathcal{R}(\nabla\bar{u})+1$ according to (8). Because $\mathcal{H}(u) := \|Au-f\|^2$ is continuous, there must exist a neighborhood $\mathcal{O}(\bar{u})$ of $\bar{u}$ such that $\forall\, u \in \mathcal{O}(\bar{u})$, $\|Au-f\|^2 \geqslant \|A\bar{u}-f\|^2-\frac{2}{\alpha}$. Hence, $\forall\, u \in B_2 \cap \mathcal{O}(\bar{u})$, we have $\mathcal{F}(\bar{u}) \leqslant \mathcal{F}(u)$ as well.

Consequently, $\forall\, u \in X \cap B_\infty(\bar{u}, r_{\bar{u}}) \cap \mathcal{O}(\bar{u})$, we have $\mathcal{F}(\bar{u}) \leqslant \mathcal{F}(u)$, which means $\bar{u}$ is a local minimizer of (P). □

For any given $\omega \subseteq \mathbb{I}_N$, we have such a problem (Q_ω), and any solution $\bar{u}$ of (Q_ω) is a local minimizer of (P). In this way, we can obtain some local minimizers of (P). Besides, from proposition 3.2, $\bar{u}$ satisfies $\sigma(\nabla\bar{u}) \subseteq \omega$. However, it is not necessary that $\sigma(\nabla\bar{u})=\omega$. See example 3.3.

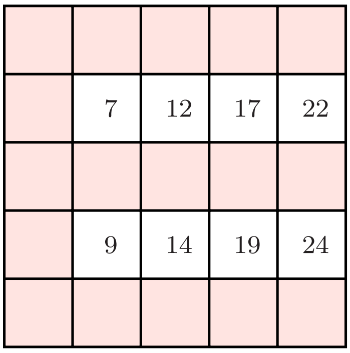


Figure 1. A counter example of $\sigma(\nabla \bar{u}) = \omega$.

Example 3.3. Figure 1 shows an image $\boldsymbol{u} \in \mathbf{R}^{5\times 5}$. Then $N = 25$. Let $\omega = \{7, 9, 12, 14, 17, 19, 22, 24\}$. The pixels in ω^c are colored in pink. For any $u \in \mathbf{R}^N$, employing the second constraint $(\nabla u)[i] = 0, \forall\, i \in \omega^c$ in (Q_ω) yields $u = u[1] \cdot \mathbf{1}_N$, and then $\sigma(\nabla u) = \emptyset$. Thus, it is quite clear that for any solution $\bar{u}$ of (Q_ω), we have $\sigma(\nabla \bar{u}) = \emptyset$. It follows that $\sigma(\nabla \bar{u}) \neq \omega$.

Proposition 3.4. *If $\bar{u}$ is a local minimizer of* (P)*, then $\bar{u}$ solves $(\mathrm{Q}_{\bar{\omega}})$ with $\bar{\omega} := \sigma(\nabla \bar{u})$.*

Proof. Since $\bar{u}$ is a local minimizer of (P), $\bar{u}$ is also a local minimizer of the following problem:

$$\min_{u \in \mathbf{R}^N} \mathcal{F}(u) \quad \text{subject to} \quad \underline{b} \cdot \mathbf{1}_N \leqslant u \leqslant \bar{b} \cdot \mathbf{1}_N; \quad (\nabla u)[i] = 0, \forall\, i \in \bar{\omega}^c. \tag{10}$$

The feasible domain of (10) is $X \cap C_{\bar{\omega}}$. Thus, there exists a neighborhood $\mathcal{O}(\bar{u})$ of $\bar{u}$ such that $\forall\, u \in \mathcal{O}(\bar{u}) \cap X \cap C_{\bar{\omega}}$, $\mathcal{F}(u) \geqslant \mathcal{F}(\bar{u})$.

According to lemma 3.1, $\forall\, u \in C_{\bar{\omega}} \cap B_\infty(\bar{u}, r_{\bar{u}}), \mathcal{R}(\nabla u) = \mathcal{R}(\nabla \bar{u})$. Thus,

$$\forall\, u \in \mathcal{O}(\bar{u}) \cap X \cap C_{\bar{\omega}} \cap B_\infty(\bar{u}, r_{\bar{u}}), \qquad \|Au - f\|^2 \geqslant \|A\bar{u} - f\|^2.$$

It follows that $\bar{u}$ is a local minimizer of $(\mathrm{Q}_{\bar{\omega}})$, whose feasible domain is also $X \cap C_{\bar{\omega}}$. Since $(\mathrm{Q}_{\bar{\omega}})$ is a convex optimization problem, we have $\bar{u}$ solves $(\mathrm{Q}_{\bar{\omega}})$. $\square$

Combining propositions 3.2 and 3.4, we can see that solving problem (Q_ω) for all $\omega \subseteq \mathbb{I}_N$ is equivalent to finding all of the local minimizers of (P). Specially, we have the following theorem.

Theorem 3.5. *Denote the local minimizer set of* (P) *as $\mathcal{U}^l$. Then*

$$\mathcal{U}^l = \bigcup_{\omega \subseteq \mathbb{I}_N} \{u \in \mathbf{R}^N \ : \ u \text{ solves } (\mathrm{Q}_\omega)\ \}.$$

Clearly, there are at most 2^N different problems (Q_ω), which are irrelevant to α. Thus, $\mathcal{U}^l$ is independent of α as well. In other words, for any given $\alpha \geqslant 0$, the local minimizer set of (P) is the same.

Note that any local minimizer of (P) satisfies the first-order necessary condition. Conversely, based on theorem 3.5, we can show the following result.

Theorem 3.6. *Every point satisfying the first-order necessary condition of* (P) *is a local minimizer.*

Proof. By introducing an indicator function, the constrained problem (P) is equivalent to

$$\min_{u\in\mathbf{R}^N} \psi(u) := \frac{\alpha}{2}\|Au-f\|^2 + \mathcal{R}(\nabla u) + \mathcal{I}_X(u),$$

where X is the feasible domain of (P) defined in (4). Let $\bar{u}$ be a point satisfying the first-order necessary condition of (P), i.e. $0 \in \partial\psi(\bar{u})$. Denote $\bar{\omega} = \sigma(\nabla\bar{u})$. Our idea is to show that $\bar{u}$ is also a solution to $(\mathrm{Q}_{\bar{\omega}})$, and thus a local minimizer of (P).

We define:

$$\psi_1(u) := \mathcal{I}_{X\cap C_{\bar{\omega}}}(u), \qquad \psi_2(u) := \mathcal{R}(\nabla u) + \mathcal{I}_X(u),$$

where $C_{\bar{\omega}}$ is defined in (6) so that $X \cap C_{\bar{\omega}}$ is the feasible domain of $(\mathrm{Q}_{\bar{\omega}})$. Next we will prove the result in three steps.

- *step 1: show that* $\partial\psi_1(\bar{u}) = \partial\psi_2(\bar{u})$
  Since $C_{\bar{\omega}}$ and X are both convex, we have $\forall\, u \in C_{\bar{\omega}} \cap X$,

$$\partial\psi_1(u) = \hat{\partial}\psi_1(u) = \{u^* \in \mathbf{R}^N \ : \ \liminf_{\substack{v\neq u, v\to u\\ v\in X\cap C_{\bar{\omega}}}} -\frac{1}{\|v-u\|}\langle u^*, v-u\rangle \geqslant 0\}.$$

Take $u \in X \cap C_{\bar{\omega}} \cap B_\infty(\bar{u}, r_{\bar{u}})$. Then $\sigma(\nabla u) = \bar{\omega}$, and by definition,

$$\begin{aligned}
\hat{\partial}\psi_2(u) &= \{u^* \in \mathbf{R}^N \ : \ \liminf_{\substack{v\neq u\\ v\to u}} \frac{1}{\|v-u\|}[\psi_2(v) - \psi_2(u) - \langle u^*, v-u\rangle] \geqslant 0\,\} \\
&= \{u^* \in \mathbf{R}^N \ : \ \liminf_{\substack{v\neq u, v\to u\\ v\in X\cap B_\infty(u,r_u)}} \frac{1}{\|v-u\|}[\mathcal{R}(\nabla v) - \mathcal{R}(\nabla u) - \langle u^*, v-u\rangle] \geqslant 0\,\} \\
[\,\text{lemma 3.1}\,] &= \{u^* \in \mathbf{R}^N \ : \ \liminf_{\substack{v\neq u, v\to u\\ v\in X\cap B_\infty(u,r_u)\\ \mathcal{R}(\nabla v)=\mathcal{R}(\nabla u)}} -\frac{1}{\|v-u\|}\langle u^*, v-u\rangle \geqslant 0\,\} \\
[\,(7)\,] &= \{u^* \in \mathbf{R}^N \ : \ \liminf_{\substack{v\neq u, v\to u\\ v\in X\cap C_{\bar{\omega}}\cap B_\infty(u,r_u)}} -\frac{1}{\|v-u\|}\langle u^*, v-u\rangle \geqslant 0\,\} \\
&= \partial\psi_1(u).
\end{aligned}$$

Therefore,

$$\begin{aligned}
\partial\psi_2(\bar{u}) &= \{u^* \in \mathbf{R}^N \ : \ \exists\, u_n \to \bar{u}, \psi_2(u_n) \to \psi_2(\bar{u}), \hat{\partial}\psi_2(u_n) \ni u_n^* \to u^*\} \\
&= \{u^* \in \mathbf{R}^N \ : \ \exists\, u_n \in X \cap B_\infty(\bar{u}, r_{\bar{u}}) \to \bar{u}, \mathcal{R}(\nabla u_n) \to \mathcal{R}(\nabla\bar{u}), \hat{\partial}\psi_2(u_n) \ni u_n^* \to u^*\} \\
&= \{u^* \in \mathbf{R}^N \ : \ \exists\, u_n \in X \cap B_\infty(\bar{u}, r_{\bar{u}}) \to \bar{u}, \mathcal{R}(\nabla u_n) = \mathcal{R}(\nabla\bar{u}), \hat{\partial}\psi_2(u_n) \ni u_n^* \to u^*\} \\
[\,(7)\,] &= \{u^* \in \mathbf{R}^N \ : \ \exists\, u_n \in X \cap C_{\bar{\omega}} \cap B_\infty(\bar{u}, r_{\bar{u}}) \to \bar{u}, \hat{\partial}\psi_1(u_n) \ni u_n^* \to u^*\} \\
&= \partial\psi_1(\bar{u}).
\end{aligned}$$

The last equality is due to the definition of subdifferential of ψ_1.

- *step 2: show that $\bar{u}$ is a solution of* $(\mathrm{Q}_{\bar{\omega}})$
  Since $0 \in \partial\psi(\bar{u})$, by proposition 2.3, we have

$$0 \in \alpha A^T(A\bar{u} - f) + \partial\psi_2(\bar{u}).$$

Thus, $0 \in \alpha A^T(A\bar{u} - f) + \partial\psi_1(\bar{u})$. This implies $\bar{u}$ is a stationary point of the following problem:

$$\min_{u\in\mathbf{R}^N} \frac{\alpha}{2}\|Au - f\|^2 + \mathcal{I}_{X\cap C_{\bar{\omega}}}(u),$$

which is equivalent to $(\mathrm{Q}_{\bar{\omega}})$. Following the fact that $(\mathrm{Q}_{\bar{\omega}})$ is a convex problem, one has $\bar{u}$ is a solution of $(\mathrm{Q}_{\bar{\omega}})$.

- *step 3: conclusion*
  Finally, by theorem 3.5, $\bar{u}$ is a local minimizer of (P). □

3.2. Solution to problem (Q_ω)

To further study the local minimizers of (P), even the global minimizers, we reformulate (Q_ω) to be a bounded-variable least square problem in this subsection, which shows good properties [16]. To deal with the second constraint $(\nabla u)[i] = 0, \forall\, i \in \omega^c$ in (Q_ω), which means $u \in C_\omega$ defined in (6), we need to introduce some notations in this subsection.

3.2.1. Block partition, block image and extension matrix. For each index $i \in \mathbb{I}_N$, we define its forward one-neighborhood as:

$$\mathbb{B}_i := \begin{cases} \{i, i+1, i+n\}, & \text{if} \quad 1 \leqslant i \mod n \leqslant n-1 \quad \text{and} \quad 1 \leqslant \lceil i/n \rceil \leqslant n-1, \\ \{i, i+1\}, & \text{if} \quad 1 \leqslant i \mod n \leqslant n-1 \quad \text{and} \quad \lceil i/n \rceil = n, \\ \{i, i+n\}, & \text{if} \quad i \mod n = 0 \quad \text{and} \quad 1 \leqslant \lceil i/n \rceil \leqslant n-1, \\ \{i\}, & \text{if} \quad i = N. \end{cases}$$

Then,

$$(\nabla u)[i] = 0, i \in \mathbb{I}_N \iff u[j] = u[i], \ \forall j \in \mathbb{B}_i.$$
$$u \in C_\omega \iff u[j] = u[i], \ \forall j \in \mathbb{B}_i, i \in \omega^c.$$

The first concept is block partition.

Let $i_0 \in \omega^c$. If there exists $r_1 \in \omega^c$ such that $\mathbb{B}_{i_0} \cap \mathbb{B}_{r_1} \neq \emptyset$, then $u[i_0] = u[j]$, $\forall j \in \mathbb{B}_{i_0} \cup \mathbb{B}_{r_1}$; next, if there exists $r_2 \in \omega^c$ such that $(\mathbb{B}_{i_0} \cup \mathbb{B}_{r_1}) \cap \mathbb{B}_{r_2} \neq \emptyset$, then $u[i_0] = u[j], \forall j \in \mathbb{B}_{i_0} \cup \mathbb{B}_{r_1} \cup \mathbb{B}_{r_2}$; and so on. Hence, there are as many as possible indices $j \in \mathbb{I}_N$ satisfying $u[i_0] = u[j]$ when $u \in C_\omega$. Thus, we define a partition of $\mathbb{I}_N$ as follows.

Definition 3.7. Given $\omega \subseteq \mathbb{I}_N$, $\mathcal{S}_\omega = (\{\tau_1, \cdots, \tau_{n_\omega}\}, <)$ is a strictly ordered partition of $\mathbb{I}_N$ satisfying:

(a) for any $i \in \omega^c$, there exists $k \in \mathbb{I}_{n_\omega}$ such that $\mathbb{B}_i \subseteq \tau_k$;
(b) for any $k \in \mathbb{I}_{n_\omega}$, if $\sharp\tau_k > 1$, then any two elements in τ_k are connected by ω^c: $\forall\, s, t \in \tau_k$, there exist $r_1, \cdots, r_l \in \omega^c$ such that $s \in \mathbb{B}_{r_1}$, $t \in \mathbb{B}_{r_l}$ and $\mathbb{B}_{r_j} \cap \mathbb{B}_{r_{j+1}} \neq \emptyset, \forall j \in \mathbb{I}_{l-1}$.

Then we say $\mathcal{S}_\omega$ is the block partition of $\mathbb{I}_N$ with respect to ω.

Example 3.8. This is an example for block partition figure 2(a) shows an image $\boldsymbol{u} \in \mathbf{R}^{5\times 5}$. Then $N = 25$. Let $\omega^c = \{1, 4, 8, 13, 22, 25\}$, and $(\nabla u)[i] = 0, \forall\, i \in \omega^c$. The pixels in ω^c are colored in brown, and pixels in $\mathbb{B}_i, i \in \omega^c$ are colored in pink. As

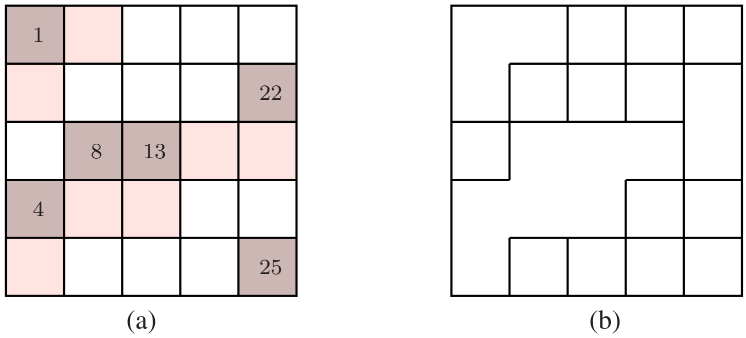


Figure 2. An example of block image. (a) An image u. (b) The block image u_ω.

$\mathbb{B}_4 \cap \mathbb{B}_8 = \{9\}, \mathbb{B}_8 \cap \mathbb{B}_{13} = \{13\}$, we have $u[4] = u[5] = u[8] = u[9] = u[13] = u[14] = u[18]$. Thus, $\mathcal{S}_\omega = (\{\{1,2,6\},\{3\},\{4,5,8,9,13,14,18\},\{7\},\{10\},\{11\},\{12\},\{15\},\{16\},\{17\},\{19\},\{20\},\{21\},\{22,23\},\{24\},\{25\}\},<)$.

For any given $\omega \subseteq \mathbb{I}_N$, the block partition $\mathcal{S}_\omega$ is unique. The number of blocks in $\mathcal{S}_\omega$ is called the block cardinality of $\mathcal{S}_\omega$, denoted by n_ω. Note that n_ω is not equal to $\sharp\omega$ and different ω may correspond to the same block partition.

The second concept is block image.

The following fact can be verified directly from the definition of $\mathcal{S}_\omega$:

$$u \in C_\omega \quad \Longleftrightarrow \quad u[j] = u[\tau_k[1]],\ \forall j \in \tau_k, k \in \mathbb{I}_{n_\omega}. \tag{11}$$

The right term of (11) means all pixels in the same block of $\mathcal{S}_\omega$ share the same value. Therefore, for any $\tau_k \in \mathcal{S}_\omega$, we can use $u[\tau_k[1]]$ to represent all $u[j], j \in \tau_k$. And then the image in $\mathbf{R}^N$ is projected into a lower dimensional space.

Definition 3.9. Given $\omega \subseteq \mathbb{I}_N$, $\mathcal{S}_\omega = (\{\tau_1, \cdots, \tau_{n_\omega}\},<)$ is the block partition with respect to ω. The projection operator $\mathcal{P}_\omega$: $\mathbf{R}^N \to \mathbf{R}^{n_\omega}$ is given by: $u \longmapsto u_\omega := \mathcal{P}_\omega(u)$, where

$$u_\omega[k] = u[\tau_k[1]], \quad \forall k \in \mathbb{I}_{n_\omega}. \tag{12}$$

Then u_ω is referred to as the block image of u with respect to ω.

For convenience, we use u_ω to substitute for $\mathcal{P}_\omega(u)$ in the whole paper. For any given $\omega \subseteq \mathbb{I}_N$, the block image u_ω is unique due to the strict orderedness of the blocks in $\mathcal{S}_\omega$. Figure 2(b) gives the block image u_ω of the image u showed in figure 2(a).

The third concept is extension matrix.

Given a block image in $\mathbf{R}^{n_\omega}$, we need to recover its original image in $\mathbf{R}^N$. That is, for the recovered image, all pixels in the same block of $\mathcal{S}_\omega$ should get the value of this block. Thus, according to the definition of $\mathcal{S}_\omega$, we define the extension matrix $E_\omega \in \mathbf{R}^{N \times n_\omega}$ as

$$(E_\omega)[j,k] := \begin{cases} 1, & j \in \tau_k, \\ 0, & \text{otherwise}. \end{cases} \tag{13}$$

From (13), $(E_\omega)[j,k] = 1$ is equivalent to that the index j belongs to the block τ_k. Notice that every index belongs to only one of the blocks in a partition, and the intersection between any two different blocks in a partition is empty. Then, we obtain that each row in E_ω has one and only one non-zero entry; each column has at least one non-zero entry. Besides, The ith column of E_ω is denoted as $(e_\omega)_i$. The columns of E_ω are orthogonal to each other and E_ω has full column rank.

Lemma 3.10. *Given* $\omega \subseteq \mathbb{I}_N$, E_ω *in* (13) *satisfies* $E_\omega^T E_\omega = \mathrm{diag}(\sharp\tau_1, \cdots, \sharp\tau_{n_\omega})$ *and* $\|E_\omega\|_2 \leqslant \sqrt{N}$.

Proof. Since $\forall\, k, l \in \mathbb{I}_{n_\omega}$,

$$(E_\omega^T E_\omega)[k,l] = (e_\omega)_k^T (e_\omega)_l = \begin{cases} \sharp\tau_k, & \text{if } k = l; \\ 0, & \text{otherwise,} \end{cases}$$

the result follows immediately. □

3.2.2. Bounded-variable least square problem. The relationship between previous notations is summarized below.

Lemma 3.11. *Given* $\omega \subseteq \mathbb{I}_N$ *and* $u \in \mathbf{R}^N$, *the following statements are equivalent:*

(a) $u \in C_\omega$
(b) $u = E_\omega u_\omega$;
(c) $u \in \{E_\omega v \in \mathbf{R}^N : v \in \mathbf{R}^{n_\omega}\}$.

Proof.

$(a) \Leftrightarrow (b)$. By (11), we have

$$\begin{aligned} (\nabla u)[i] = 0,\ \forall\, i \in \omega^c \quad &\Longleftrightarrow \quad u[j] = u[\tau_k[1]] \overset{(12)}{=} u_\omega[k],\ \forall\, j \in \tau_k, k \in \mathbb{I}_{n_\omega} \\ &\Longleftrightarrow \quad u = E_\omega u_\omega. \end{aligned}$$

$(b) \Rightarrow (c)$. This is obvious.

$(c) \Leftrightarrow (b)$. Since $u \in \{E_\omega v \in \mathbf{R}^N : v \in \mathbf{R}^{n_\omega}\}$, there exists $\bar{v} \in \mathbf{R}^{n_\omega}$ such that $u = E_\omega \bar{v}$. Then

$$u[j] = \bar{v}[k] = u[\tau_k[1]] = u_\omega[k], \qquad \forall\, j \in \tau_k, k \in \mathbb{I}_{n_\omega}.$$

Thus $\bar{v} = u_\omega$ and $u = E_\omega u_\omega$. □

For any matrix $A \in \mathbf{R}^{M\times N}$, the ith column in A is denoted by a_i. Then, we denote

$$A_\omega := AE_\omega = (a_{\tau_1[1]} + \cdots + a_{\tau_1[\sharp\tau_1]}, \cdots, a_{\tau_{n_\omega}[1]} + \cdots + a_{\tau_{n_\omega}[\sharp\tau_{n_\omega}]}) \in \mathbf{R}^{M\times n_\omega}. \tag{14}$$

Each column of A_ω is a linear combination of the columns of A. If A has full column rank, then A_ω is of full column rank. The ith column of A_ω is denoted as $(a_\omega)_i$. We set $A_\omega^T := (A_\omega)^T$. If A_ω is invertible, similarly, $A_\omega^{-1} := (A_\omega)^{-1}$. So is E_ω. From lemma 3.7, given $u \in \mathbf{R}^N$, if $u \in C_\omega$, then

$$Au = AE_\omega u_\omega = A_\omega u_\omega. \tag{15}$$

Using lemmas 3.7 and (14), we can eliminate the second constraint from (Q_ω). Then (Q_ω) can be reformulated by a bounded-variable least square problem [13] as follows:

$$(\mathcal{Q}_\omega) \qquad \begin{cases} \min_{v\in\mathbf{R}^{n_\omega}} & \|A_\omega v - f\|^2 \\ \text{s.t.} & \underline{b}\cdot \mathbf{1}_{n_\omega} \leqslant v \leqslant \bar{b}\cdot \mathbf{1}_{n_\omega}. \end{cases}$$

This is a convex optimization problem and always admits a solution. Combining with proposition 3.4, we get the following lemma.

Lemma 3.12. *If $\bar{u} \in \mathbf{R}^N$ is a local minimizer of* (P)*, then $\bar{u}_{\bar{\omega}}$ solves $(\mathcal{Q}_{\bar{\omega}})$ with $\bar{\omega} := \sigma(\nabla \bar{u})$.*

We conclude this subsection as follows:

$$\boxed{\begin{array}{rcl} \bar{u} \text{ solves } (\mathrm{Q}_\omega) & \Longrightarrow & \bar{u}_\omega = \mathcal{P}_\omega(\bar{u}) \text{ solves } (\mathcal{Q}_\omega), \\ \bar{u} = E_w \bar{v} \text{ solves } (\mathrm{Q}_\omega) & \Longleftarrow & \bar{v} \text{ solves } (\mathcal{Q}_\omega). \end{array}} \tag{16}$$

3.3. The strict local minimizer

In this subsection, we study the strict local minimizer of (P). Especially, we will show that any two strict local minimizers correspond to different function values under certain conditions.

Proposition 3.13. *Given $\omega \subseteq \mathbb{I}_N$. If (Q_ω) has a unique solution $\bar{u}$, then $\bar{u}$ is a strict local minimizer of* (P).

Proof. The proof is similar to the one of proposition 3.2. Similarly, we denote $X \cap B_\infty(\bar{u}, r_{\bar{u}}) = B_1 \cup B_2$, with B_1, B_2 defined as (9).

Take any $u \in B_1$. Then u satisfies $\mathcal{R}(\nabla u) = \mathcal{R}(\nabla \bar{u})$, and is a feasible point of (Q_ω). Since $\bar{u}$ is the unique solution of (Q_ω), it follows that $\|A\bar{u} - f\|^2 < \|Au - f\|^2$. Thus, $\mathcal{F}(\bar{u}) < \mathcal{F}(u)$.

For any $u \in B_2$, $\mathcal{R}(\nabla u) \geqslant \mathcal{R}(\nabla \bar{u}) + 1$. Meanwhile, there exists a neighborhood $\tilde{\mathcal{O}}(\bar{u})$ of $\bar{u}$ such that $\forall\, u \in \tilde{\mathcal{O}}(\bar{u})$, $\|Au - f\|^2 < \|A\bar{u} - f\|^2 - \frac{2}{\alpha}$. Thus, $\forall\, u \in B_2 \cap \tilde{\mathcal{O}}(\bar{u})$, we have $\mathcal{F}(\bar{u}) < \mathcal{F}(u)$.

Therefore, for any $u \in X \cap B_\infty(\bar{u}, r_{\bar{u}}) \cap \tilde{\mathcal{O}}(\bar{u})$, we have $\mathcal{F}(\bar{u}) < \mathcal{F}(u)$. Equivalently, $\bar{u}$ is a strict local minimizer of (P). □

If A_ω has full column rank, $(\mathcal{Q}_\omega)$ is a strictly convex problem and has a unique solution. Then, from (16), we can see that the solution of (Q_ω) is unique. However, the reverse direction is not always true. If $\underline{b} \neq -\infty$ or $\bar{b} \neq +\infty$, it is possible that when A_ω is column rank deficient, $(\mathcal{Q}_\omega)$ has a unique solution lying in the boundary of its feasible region.

Proposition 3.14. *Suppose that $\underline{b} = -\infty$ and $\bar{b} = +\infty$. Let $\bar{u} \in \mathbf{R}^N$ be a local minimizer of* (P)*. Denote $\bar{\omega} := \sigma(\nabla \bar{u})$, then the following statements are equivalent:*

(a) $A_{\bar{\omega}}$ has full column rank;
(b) the solution of $(\mathrm{Q}_{\bar{\omega}})$ is unique;
(c) $\bar{u}$ is a strict local minimizer of (P).

Proof.

$(a) \Leftrightarrow (b)$ is obvious.

$(b) \Rightarrow (c)$ is due to proposition 3.13.

$(c) \Rightarrow (a)$. Since $\bar{u}$ is a local minimizer of (P), according to lemma 3.12, we have $\bar{u}_{\bar{\omega}} \in \mathbf{R}^{n_{\bar{\omega}}}$ solves $(\mathcal{Q}_{\bar{\omega}})$. Since $\underline{b} = -\infty$ and $\bar{b} = +\infty$, $(\mathcal{Q}_{\bar{\omega}})$ reads as

$$\min_{v \in \mathbf{R}^{n_{\bar{\omega}}}} \quad \|A_{\bar{\omega}} v - f\|^2. \tag{17}$$

Assume that statement (a) fails, i.e. $\ker A_{\bar{\omega}} \neq \emptyset$. Let $\mathcal{O}(\bar{u})$ be an arbitrary neighborhood of $\bar{u}$. Take $\tilde{v} \in \ker A_{\bar{\omega}}$ whose norm is small enough such that $\tilde{u} = E_{\bar{\omega}}(\bar{u}_{\bar{\omega}} + \tilde{v}) \in \mathcal{O}(\bar{u})$. Then $\bar{u}_{\bar{\omega}} + \tilde{v}$ is a solution of $(\mathcal{Q}_{\bar{\omega}})$, and $\tilde{u}$ solves $(\mathrm{Q}_{\bar{\omega}})$ with $\|A\tilde{u} - f\|^2 = \|A\bar{u} - f\|^2$. Meanwhile, from lemma 3.7, we have $(\nabla \tilde{u})[i] = 0, \forall\, i \in \bar{\omega}^c$. Hence $\mathcal{R}(\nabla \tilde{u}) \leqslant \sharp \bar{\omega} = \mathcal{R}(\nabla \bar{u})$ and then

$\mathcal{F}(\tilde{u}) \leqslant \mathcal{F}(\bar{u})$. This contradicts to the fact that $\bar{u}$ is a strict local minimizer of (P). $\square$

Corollary 3.15. *If A has full column rank, any local minimizer of* (P) *is a strict local minimizer.*

Proof. For any local minimizer $\bar{u}$ of (P), it is a solution of $(\mathrm{Q}_{\bar{\omega}})$ with $\bar{\omega} := \sigma(\nabla \bar{u})$. Since A has full column rank, $A_{\bar{\omega}}$ has full column rank and $\bar{u}$ is the unique solution of $(\mathrm{Q}_{\bar{\omega}})$. Hence, $\bar{u}$ is a strict local minimizer by proposition 3.13. $\square$

Theorem 3.16. *Suppose that A has full column rank. Then, there exists a subset $Z \subset \mathbf{R}^M$, whose Lebesgue measure is zero, such that if $f \in \mathbf{R}^M \backslash Z$, any two local minimizers of* (P) *have different energy values.*

Proof. Let $\bar{u}, \tilde{u}$ be two local minimizers of (P). The main idea is to find the necessary condition for f when $\mathcal{F}(\bar{u}) = \mathcal{F}(\tilde{u})$. The proof is divided into three steps.

- *step 1: find the expressions of $\mathcal{F}(\bar{u})$ and $\mathcal{F}(\tilde{u})$*

Take $\bar{u}$ as an example. Denote $\bar{\omega} := \sigma(\nabla \bar{u})$, then $\bar{u}$ is a solution of $(\mathrm{Q}_{\bar{\omega}})$ and $\bar{u}_{\bar{\omega}}$ is a solution of $(\mathcal{Q}_{\bar{\omega}})$.

Firstly, we give some notations to simplify the presentation. We define the active indices of $\bar{u}_{\bar{\omega}}$ as $\mathbb{A}(\bar{u}_{\bar{\omega}}) = \nu \subseteq \mathbb{I}_{n_{\bar{\omega}}}$ such that

$$\begin{cases} \forall\, i \in \nu, & \bar{u}_{\bar{\omega}}[i] = \underline{b} \text{ or } \bar{b}, \\ \forall\, i \in \nu^c, & \underline{b} < \bar{u}_{\bar{\omega}}[i] < \bar{b}. \end{cases}$$

Let

$$\begin{cases} \bar{u}^{\mathrm{a}}_{\bar{\omega}} & := & (\bar{u}_{\bar{\omega}}[\nu[1]], \cdots, \bar{u}_{\bar{\omega}}[\nu[\sharp\nu]]) \in \mathbf{R}^{\sharp\nu}, \\ \bar{u}^{\mathrm{i}}_{\bar{\omega}} & := & (\bar{u}_{\bar{\omega}}[\nu^c[1]], \cdots, \bar{u}_{\bar{\omega}}[\nu^c[\sharp\nu^c]]) \in \mathbf{R}^{\sharp\nu^c}. \end{cases}$$

We call $\bar{u}^{\mathrm{a}}_{\bar{\omega}}, \bar{u}^{\mathrm{i}}_{\bar{\omega}}$ the active part and inactive part of $\bar{u}_{\bar{\omega}}$ respectively. If $\nu = \emptyset$, then $\bar{u}^{\mathrm{a}}_{\bar{\omega}}$ does not exist and $\bar{u}^{\mathrm{i}}_{\bar{\omega}} = \bar{u}_{\bar{\omega}}$. Furthermore, we define two extension submatrices induced by $\mathbb{A}(\bar{u}_{\bar{\omega}})$ as

$$\begin{cases} E^{\mathrm{a}}_{\bar{\omega}} & := & ((e_{\bar{\omega}})_{\nu[1]}, \cdots, (e_{\bar{\omega}})_{\nu[\sharp\nu]}) \in \mathbf{R}^{M\times\sharp\nu}, \\ E^{\mathrm{i}}_{\bar{\omega}} & := & ((e_{\bar{\omega}})_{\nu^c[1]}, \cdots, (e_{\bar{\omega}})_{\nu^c[\sharp\nu^c]})\mathbf{R}^{M\times\sharp\nu^c}. \end{cases}$$

Then we have

$$\bar{u} = E_{\bar{\omega}} \bar{u}_{\bar{\omega}} = E^{\mathrm{a}}_{\bar{\omega}} \bar{u}^{\mathrm{a}}_{\bar{\omega}} + E^{\mathrm{i}}_{\bar{\omega}} \bar{u}^{\mathrm{i}}_{\bar{\omega}}.$$

Also, we denote $A^{\mathrm{a}}_{\bar{\omega}} := AE^{\mathrm{a}}_{\bar{\omega}}$ and $A^{\mathrm{i}}_{\bar{\omega}} := AE^{\mathrm{i}}_{\bar{\omega}}$ as two transformation submatrices induced by $\mathbb{A}(\bar{u}_{\bar{\omega}})$. Then

$$A\bar{u} = A_{\bar{\omega}} \bar{u}_{\bar{\omega}} = A^{\mathrm{a}}_{\bar{\omega}} \bar{u}^{\mathrm{a}}_{\bar{\omega}} + A^{\mathrm{i}}_{\bar{\omega}} \bar{u}^{\mathrm{i}}_{\bar{\omega}}. \tag{18}$$

For clarity, we define the $M \times M$ matrix $\Pi(A^{\mathrm{i}}_{\bar{\omega}})$ as the orthogonal projection [18] onto the subspace spanned by the columns of $A^{\mathrm{i}}_{\bar{\omega}}$. As $A^{\mathrm{i}}_{\bar{\omega}}$ has full column rank, we have

$$\Pi(A^{\mathrm{i}}_{\bar{\omega}}) = A^{\mathrm{i}}_{\bar{\omega}}[(A^{\mathrm{i}}_{\bar{\omega}})^T A^{\mathrm{i}}_{\bar{\omega}}]^{-1}(A^{\mathrm{i}}_{\bar{\omega}})^T. \tag{19}$$

If $\bar{u}^{\mathrm{a}}_{\bar{\omega}}$ is given, then $\bar{u}^{\mathrm{i}}_{\bar{\omega}}$ solves the following problem

$$\begin{aligned}&\min_{v\in\mathbf{R}^{\sharp\nu^c}} \quad \|A^{\mathrm{i}}_{\bar\omega}v-(f-A^{\mathrm{a}}_{\bar\omega}\bar u^{\mathrm{a}}_{\bar\omega})\|^2,\\ &\text{s.t.} \qquad \underline{b}\cdot\mathbf{1}_{\sharp\nu^c}<v<\bar b\cdot\mathbf{1}_{\sharp\nu^c}.\end{aligned}\tag{20}$$

Since $\bar u^{\mathrm{i}}_{\bar\omega}$ is an interior point of the feasible region of the problem above and $A^{\mathrm{i}}_{\bar\omega}$ has full column rank, we have

$$\bar u^{\mathrm{i}}_{\bar\omega}=[(A^{\mathrm{i}}_{\bar\omega})^TA^{\mathrm{i}}_{\bar\omega}]^{-1}(A^{\mathrm{i}}_{\bar\omega})^T(f-A^{\mathrm{a}}_{\bar\omega}\bar u^{\mathrm{a}}_{\bar\omega})=\Pi(A^{\mathrm{i}}_{\bar\omega})(f-A^{\mathrm{a}}_{\bar\omega}\bar u^{\mathrm{a}}_{\bar\omega}).\tag{21}$$

Denote the identity matrix in $\mathbf{R}^M$ as I. By (18) and (19), we have

$$\begin{aligned}\|A\bar u-f\|^2&=\|A^{\mathrm{i}}_{\bar\omega}\bar u^{\mathrm{i}}_{\bar\omega}+A^{\mathrm{a}}_{\bar\omega}\bar u^{\mathrm{a}}_{\bar\omega}-f\|^2\\&=\|\Pi(A^{\mathrm{i}}_{\bar\omega})(f-A^{\mathrm{a}}_{\bar\omega}\bar u^{\mathrm{a}}_{\bar\omega})+A^{\mathrm{a}}_{\bar\omega}\bar u^{\mathrm{a}}_{\bar\omega}-f\|^2\\&=\|(\Pi(A^{\mathrm{i}}_{\bar\omega})-I)f-(\Pi(A^{\mathrm{i}}_{\bar\omega})-I)A^{\mathrm{a}}_{\bar\omega}\bar u^{\mathrm{a}}_{\bar\omega}\|^2\\&=f^TW(\bar y)f+f^TS(\bar y)+T(\bar y),\end{aligned}\tag{22}$$

where

$$\begin{cases}W(\bar y) & := I-\Pi(A^{\mathrm{i}}_{\bar\omega})\in\mathbf{R}^{M\times M},\\ S(\bar y) & := 2(\Pi(A^{\mathrm{i}}_{\bar\omega})-I)A^{\mathrm{a}}_{\bar\omega}\bar u^{\mathrm{a}}_{\bar\omega}\in\mathbf{R}^{M\times 1},\\ T(\bar y) & := (A^{\mathrm{a}}_{\bar\omega}\bar u^{\mathrm{a}}_{\bar\omega})^T(I-\Pi(A^{\mathrm{i}}_{\bar\omega}))A^{\mathrm{a}}_{\bar\omega}\bar u^{\mathrm{a}}_{\bar\omega}\in\mathbf{R},\end{cases}\tag{23}$$

with $\bar y:=(\bar\omega,\mathbb{A}(\bar u_{\bar\omega}),\bar u^{\mathrm{a}}_{\bar\omega})$. Notice that the definitions of $W(\bar y),S(\bar y),T(\bar y)$ depend on not only $\bar u^{\mathrm{a}}_{\bar\omega}$, but also $A^{\mathrm{a}}_{\bar\omega}$ and $A^{\mathrm{i}}_{\bar\omega}$, which are induced by index sets $\bar\omega$ and $\mathbb{A}(\bar u_{\bar\omega})$. Assume that $\mathbb{A}(\bar u_{\bar\omega})=\emptyset$. Then $A^{\mathrm{i}}_{\bar\omega}=A_{\bar\omega}$, $W(\bar y)=I-\Pi(A_{\bar\omega})$, and $\bar u^{\mathrm{a}}_{\bar\omega},S(\bar y),T(\bar y)$ do not exist. Without loss of generality, we denote $S(\bar y)=0,T(\bar y)=0$ in such a case.

Therefore, we denote

$$Y:=\{y=(\omega,\nu,v)\ :\ \omega\subseteq\mathbb{I}_N,\nu\subseteq\mathbb{I}_{n_{\bar\omega}},v\in\mathbf{R}^{\sharp\nu},v[i]=\underline{b}\text{ or }\bar b,\forall\, i\in\mathbb{I}_{\sharp\nu}\}.\tag{24}$$

Clearly, $\bar y\in Y$, and for any $y\in Y$, we can define $W(y),S(y),T(y)$ as (23). The elements in Y are finite.

Finally, we have

$$\mathcal{F}(\bar u)=\frac{\alpha}{2}[f^TW(\bar y)f+f^TS(\bar y)+T(\bar y)]+\sharp\bar\omega,$$

$$\mathcal{F}(\tilde u)=\frac{\alpha}{2}[f^TW(\tilde y)f+f^TS(\tilde y)+T(\tilde y)]+\sharp\tilde\omega,$$

with $\tilde\omega:=\sigma(\nabla\tilde u),\bar y,\tilde y:=(\tilde\omega,\mathbb{A}(\tilde u_{\tilde\omega}),\tilde u^{\mathrm{a}}_{\tilde\omega})\in Y$.

- *step 2: find the necessary condition for $\mathcal{F}(\bar u)=\mathcal{F}(\tilde u)$ to construct Z*
  It follows from $\mathcal{F}(\bar u)=\mathcal{F}(\tilde u)$ that

$$f^T(W(\bar y)-W(\tilde y))f+f^T(S(\bar y)-S(\tilde y))+T(\bar y)-T(\tilde y)-\frac{2}{\alpha}(\sharp\bar\omega-\sharp\tilde\omega)=0.\tag{25}$$

Moreover, by reduction, we can show that when $\bar u\neq\tilde u$, $W(\bar y)=W(\tilde y)$ implies $S(\bar y)\neq S(\tilde y)$.

Assume that $W(\bar y)=W(\tilde y)$ and $S(\bar y)=S(\tilde y)$.

Similarly, for $\tilde u$, the two extension submatrices induced by $\mathbb{A}(\tilde u_{\tilde\omega})$ are $E^{\mathrm{a}}_{\tilde\omega},E^{\mathrm{i}}_{\tilde\omega}$, and $A^{\mathrm{a}}_{\tilde\omega}=AE^{\mathrm{a}}_{\tilde\omega},A^{\mathrm{i}}_{\tilde\omega}=AE^{\mathrm{i}}_{\tilde\omega}$. Then, one has

$$\bar u=E^{\mathrm{a}}_{\bar\omega}\bar u^{\mathrm{a}}_{\bar\omega}+E^{\mathrm{i}}_{\bar\omega}\bar u^{\mathrm{i}}_{\bar\omega},\qquad \tilde u=E^{\mathrm{a}}_{\tilde\omega}\tilde u^{\mathrm{a}}_{\tilde\omega}+E^{\mathrm{i}}_{\tilde\omega}\tilde u^{\mathrm{i}}_{\tilde\omega}.$$

From

$$0 = W(\bar{y}) - W(\tilde{y}) = [I - \Pi(A^{\mathrm{i}}_{\bar{\omega}})] - [I - \Pi(A^{\mathrm{i}}_{\tilde{\omega}})] = \Pi(A^{\mathrm{i}}_{\tilde{\omega}}) - \Pi(A^{\mathrm{i}}_{\bar{\omega}}), \tag{26}$$

we can see that the subspace spanned by the columns of $A^{\mathrm{i}}_{\tilde{\omega}}$ equals to the subspace spanned by the columns of $A^{\mathrm{i}}_{\bar{\omega}}$. Since $A^{\mathrm{i}}_{\tilde{\omega}}$ and $A^{\mathrm{i}}_{\bar{\omega}}$ have full column rank, there exists an invertible matrix P such that

$$A^{\mathrm{i}}_{\bar{\omega}} = A^{\mathrm{i}}_{\tilde{\omega}} P \Longleftrightarrow AE^{\mathrm{i}}_{\bar{\omega}} = AE^{\mathrm{i}}_{\tilde{\omega}} P \Longleftrightarrow E^{\mathrm{i}}_{\bar{\omega}} = E^{\mathrm{i}}_{\tilde{\omega}} P. \tag{27}$$

The last step is due to that A has full column rank. Denote $v = A^{\mathrm{a}}_{\bar{\omega}} \bar{u}^{\mathrm{a}}_{\bar{\omega}} - A^{\mathrm{a}}_{\tilde{\omega}} \tilde{u}^{\mathrm{a}}_{\tilde{\omega}}$. Then, there exists x_1 such that $v = A^{\mathrm{a}}_{\bar{\omega}} x_1$.

Furthermore, we have

$$0 = S(\bar{y}) - S(\tilde{y}) = 2(\Pi(A^{\mathrm{i}}_{\bar{\omega}}) - I)A^{\mathrm{a}}_{\bar{\omega}} \bar{u}^{\mathrm{a}}_{\bar{\omega}} - 2(\Pi(A^{\mathrm{i}}_{\tilde{\omega}}) - I)A^{\mathrm{a}}_{\tilde{\omega}} \tilde{u}^{\mathrm{a}}_{\tilde{\omega}} = 2(\Pi(A^{\mathrm{i}}_{\bar{\omega}}) - I)v,$$

which indicates that $v = \Pi(A^{\mathrm{i}}_{\bar{\omega}})v$. That is, v belongs to the subspace spanned by the columns of $A^{\mathrm{i}}_{\bar{\omega}}$. Thus, there exists x_2 such that $v = A^{\mathrm{i}}_{\bar{\omega}} x_2$, and then we have $A^{\mathrm{a}}_{\bar{\omega}} x_1 - A^{\mathrm{i}}_{\bar{\omega}} x_2 = 0$. Since $A_{\bar{\omega}}$ has full column rank, we have $x_1 = x_2 = 0$. It follows that

$$A^{\mathrm{a}}_{\bar{\omega}} \bar{u}^{\mathrm{a}}_{\bar{\omega}} = A^{\mathrm{a}}_{\tilde{\omega}} \tilde{u}^{\mathrm{a}}_{\tilde{\omega}} \Longleftrightarrow AE^{\mathrm{a}}_{\bar{\omega}} \bar{u}^{\mathrm{a}}_{\bar{\omega}} = AE^{\mathrm{a}}_{\tilde{\omega}} \tilde{u}^{\mathrm{a}}_{\tilde{\omega}} \Longleftrightarrow E^{\mathrm{a}}_{\bar{\omega}} \bar{u}^{\mathrm{a}}_{\bar{\omega}} = E^{\mathrm{a}}_{\tilde{\omega}} \tilde{u}^{\mathrm{a}}_{\tilde{\omega}}. \tag{28}$$

Note that $\bar{u}^{\mathrm{i}}_{\bar{\omega}} = [(A^{\mathrm{i}}_{\bar{\omega}})^T A^{\mathrm{i}}_{\bar{\omega}}]^{-1} (A^{\mathrm{i}}_{\bar{\omega}})^T (f - A^{\mathrm{a}}_{\bar{\omega}} \bar{u}^{\mathrm{a}}_{\bar{\omega}})$ and $\tilde{u}^{\mathrm{i}}_{\tilde{\omega}} = [(A^{\mathrm{i}}_{\tilde{\omega}})^T A^{\mathrm{i}}_{\tilde{\omega}}]^{-1} (A^{\mathrm{i}}_{\tilde{\omega}})^T (f - A^{\mathrm{a}}_{\tilde{\omega}} \tilde{u}^{\mathrm{a}}_{\tilde{\omega}})$. Then

$$\begin{aligned}
& E^{\mathrm{i}}_{\bar{\omega}} \bar{u}^{\mathrm{i}}_{\bar{\omega}} = E^{\mathrm{i}}_{\bar{\omega}} [(A^{\mathrm{i}}_{\bar{\omega}})^T A^{\mathrm{i}}_{\bar{\omega}}]^{-1} (A^{\mathrm{i}}_{\bar{\omega}})^T (f - A^{\mathrm{a}}_{\bar{\omega}} \bar{u}^{\mathrm{a}}_{\bar{\omega}}) \\
[\,(27)\,] \quad & \qquad = E^{\mathrm{i}}_{\tilde{\omega}} P [(A^{\mathrm{i}}_{\tilde{\omega}} P)^T A^{\mathrm{i}}_{\tilde{\omega}} P]^{-1} (A^{\mathrm{i}}_{\tilde{\omega}} P)^T (f - A^{\mathrm{a}}_{\bar{\omega}} \bar{u}^{\mathrm{a}}_{\bar{\omega}}) \\
[\,(28)\,] \quad & \qquad = E^{\mathrm{i}}_{\tilde{\omega}} [(A^{\mathrm{i}}_{\tilde{\omega}})^T A^{\mathrm{i}}_{\tilde{\omega}}]^{-1} (A^{\mathrm{i}}_{\tilde{\omega}})^T (f - A^{\mathrm{a}}_{\tilde{\omega}} \tilde{u}^{\mathrm{a}}_{\tilde{\omega}}) \\
& \qquad = E^{\mathrm{i}}_{\tilde{\omega}} \tilde{u}^{\mathrm{i}}_{\tilde{\omega}}.
\end{aligned}$$

Combining the equality above with (28) gives

$$\bar{u} = E^{\mathrm{a}}_{\bar{\omega}} \bar{u}^{\mathrm{a}}_{\bar{\omega}} + E^{\mathrm{i}}_{\bar{\omega}} \bar{u}^{\mathrm{i}}_{\bar{\omega}} = E^{\mathrm{a}}_{\tilde{\omega}} \tilde{u}^{\mathrm{a}}_{\tilde{\omega}} + E^{\mathrm{i}}_{\tilde{\omega}} \tilde{u}^{\mathrm{i}}_{\tilde{\omega}} = \tilde{u},$$

which contradicts to the fact that $\bar{u} \neq \tilde{u}$. Therefore, if $W(\bar{y}) = W(\tilde{y})$, then $S(\bar{y}) \neq S(\tilde{y})$.

Finally, we define

$$\tilde{Y} := \{y_1 \in Y \ : \ \forall y_2 \in Y, \text{ if } W(y_1) = W(y_2), \text{then } S(y_1) \neq S(y_2)\} \subseteq Y,$$

and

$$\begin{aligned}
Z := \bigcup_{y_1, y_2 \in \tilde{Y}} \{f \in \mathbf{R}^M \ : \ & f^T(W(y_1) - W(y_2))f + f^T(S(y_1) - S(y_2)) \\
& + T(y_1) - T(y_2) - \frac{2}{\alpha}(\sharp\omega_1 - \sharp\omega_2) = 0\},
\end{aligned} \tag{29}$$

with $y_1 = (\omega_1, \nu_1, v_1), y_2 = (\omega_2, \nu_2, v_2)$. Therefore, (25) means

$$f \in Z.$$

- *step 3: find the Lebesgue measure of Z*

Take two arbitrary $y_1 \neq y_2 \in \tilde{Y}$. Denote $\mathbf{W} := W(y_1) - W(y_2) \in \mathbf{R}^{M \times M}, \mathbf{S} := S(y_1) - S(y_2) \in \mathbf{R}^{M \times 1}, \mathbf{T} := T(y_1) - T(y_2) - \frac{2}{\alpha}(\sharp\omega_1 - \sharp\omega_2) \in \mathbf{R}$. We now discuss the solution set of

$$f^T \mathbf{W} f + f^T \mathbf{S} + \mathbf{T} = 0.$$

If $\mathbf{W} \neq 0$, then the solution set is a quadric, which is a generalization of conic sections [30]. If $\mathbf{W} = 0$ and $\mathbf{S} \neq 0$, then the solution set is an $M - 1$ dimensional hyperplane in $\mathbf{R}^M$. In both cases, the solution set is closed and has zero Lebesgue measure in $\mathbf{R}^M$. Due to the finiteness of $\tilde{Y}$, Z is a finite union of such closed solution sets in $\mathbf{R}^M$ whose dimensions are not larger than $M-1$. Thus, Z has zero measure in $\mathbf{R}^M$.

Conclusively, if $f \in \mathbf{R}^M \backslash Z$, $\mathcal{F}(\bar{u}) = \mathcal{F}(\tilde{u})$ implies $\bar{u} = \tilde{u}$. □

Theorem 3.16 tells that when A has full column rank, different (strict) local minimizers of (P) correspond to different energy values. This property fails for f in a negligible subset of $\mathbf{R}^M$. In other words, it holds for almost all f. Besides, the assumption of full column rank of A is reasonable. In practical application of image processing, $A \in \mathbf{R}^{M\times M}$ is usually a sparse matrix transformed from a blur kernel, and A is invertible under general conditions.

The following corollary is obvious.

Corollary 3.17. *Suppose that A has full column rank. Then, there exists a subset $Z \subset \mathbf{R}^M$, whose Lebesgue measure is zero, such that if $f \in \mathbf{R}^M \backslash Z$, then* (P) *has a unique solution.*

4. The global minimizer

In this section, we focus on global minimizers of (P), and present their two important properties: the uniform lower bound for the ℓ_2 norms of nonzero discrete gradients, and the piecewise-constant dependency on α. Since a global minimizer is also a local minimizer of (P), previous analysis on local minimizers serves the study of the global minimizers.

4.1. The uniform lower bound for the ℓ_2 norm of nonzero discrete gradients

Theorem 4.1. *Suppose that $\bar{u}$ solves* (P). *Then, for any $i \in \sigma(\nabla\bar{u})$, we have*

$$\|(\nabla\bar{u})[i]\| \geqslant \theta, \qquad \text{where} \quad \theta := \min\{\frac{\sqrt{5}-1}{2\sqrt{\alpha N}\|A\|_2}, \frac{\sqrt{2}|\underline{b}-\bar{b}|}{2}\}. \tag{30}$$

Proof. Let $\bar{\omega} := \sigma(\nabla\bar{u})$. The block partition of $\mathbb{I}_N$ with respect to $\bar{\omega}$ is $\mathcal{S}_{\bar{\omega}} = (\{\tau_1, \cdots, \tau_{n_{\bar{\omega}}}\}, <)$. By lemma 3.12, $\bar{u}_{\bar{\omega}}$ is a solution of $(\mathcal{Q}_{\bar{\omega}})$. We define the active indices of $\bar{u}_{\bar{\omega}}$ as $\mathbb{A}(\bar{u}_{\bar{\omega}}) \subseteq \mathbb{I}_{n_{\bar{\omega}}}$ such that

$$\begin{cases} \forall\, i \in \mathbb{A}(\bar{u}_{\bar{\omega}}), & \bar{u}_{\bar{\omega}}[i] = \underline{b} \text{ or } \bar{b}, \\ \forall\, i \in (\mathbb{A}(\bar{u}_{\bar{\omega}}))^c, & \underline{b} < \bar{u}_{\bar{\omega}}[i] < \bar{b}. \end{cases}$$

Take an arbitrary $i \in \bar{\omega}$. We suppose that

$$i \in \tau_k, \quad i+1 \in \tau_s, \quad i+n \in \tau_t, \tag{31}$$

with $k, s, t \in \mathbb{I}_{n_{\bar{\omega}}}$. Thus $\bar{u}[i] = \bar{u}_{\bar{\omega}}[k], \bar{u}[i+1] = \bar{u}_{\bar{\omega}}[s], \bar{u}[i+n] = \bar{u}_{\bar{\omega}}[t]$. Let

$$\bar{u}_{\bar{\omega}}[s] = \bar{u}_{\bar{\omega}}[k] + \Delta_s, \quad \bar{u}_{\bar{\omega}}[t] = \bar{u}_{\bar{\omega}}[k] + \Delta_t,$$

where $\Delta_s, \Delta_t \in \mathbf{R}$. Then $\|(\nabla\bar{u})[i]\|^2 = (\Delta_s)^2 + (\Delta_t)^2 \neq 0$. Our goal is to find the lower bound for the ℓ_2 norm of $(\Delta_s, \Delta_t)^T$.

There are four cases involved: (a) $k \neq s, k \neq t, s \neq t$; (b) $k = s \neq t$; (c) $k = t \neq s$; (d) $k \neq s = t$. It suffices to prove the theorem when $k \neq s, k \neq t, s \neq t$, since the proofs in the other three situations are similar to the one in the following Case 3. To show the theorem when $k \neq s, k \neq t, s \neq t$, we divide our proof into four parts, mainly according to the relation between $\{k, s, t\}$ and the active indices $\mathbb{A}(\bar{u}_{\bar{\omega}})$ of $\bar{u}_{\bar{\omega}}$.

- *Case 1* $\quad \{k, s, t\} \cap \mathbb{A}(\bar{u}_{\bar{\omega}}) = \emptyset$.
  We decompose $\bar{u}_{\bar{\omega}}$ as: $\bar{u}_{\bar{\omega}} = \tilde{v} + \tilde{\Delta}$, where $\tilde{v}, \tilde{\Delta} \in \mathbf{R}^{n_{\bar{\omega}}}$,

$$\tilde{v}[j] := \begin{cases} \bar{u}_{\bar{\omega}}[k], & j = s, t, \\ \bar{u}_{\bar{\omega}}[j], & \text{otherwise}; \end{cases} \quad \text{and} \quad \tilde{\Delta}[j] := \begin{cases} \Delta_s, & j = s, \\ \Delta_t, & j = t, \\ 0, & \text{otherwise}. \end{cases}$$

  As $\bar{u}_{\bar{\omega}}$ solves $(\mathcal{Q}_{\bar{\omega}})$, $\tilde{\Delta}$ is a solution to the following problem

$$\begin{aligned} \min_{\Delta \in \mathbf{R}^{n_{\bar{\omega}}}} \quad & \|A_{\bar{\omega}}(\tilde{v} + \Delta) - f\|^2 \\ \text{s.t.} \quad & \underline{b} \cdot \mathbf{1}_{n_{\bar{\omega}}} \leqslant \tilde{v} + \Delta \leqslant \bar{b} \cdot \mathbf{1}_{n_{\bar{\omega}}}, \\ & \Delta[j] = 0, \quad j \in \mathbb{I}_{n_{\bar{\omega}}} \backslash \{s, t\}. \end{aligned}$$

Under the assumption that $s, t \notin \mathbb{A}(\bar{u}_{\bar{\omega}})$, the above problem is equivalent to

$$\begin{aligned} \min_{(\Delta_1, \Delta_2)^T \in \mathbf{R}^2} \quad & \|\mathrm{A}_{st}(\Delta_1, \Delta_2)^T - (f - A_{\bar{\omega}}\tilde{v})\|^2 \\ \text{s.t.} \quad & \underline{b} < \tilde{v}[s] + \Delta_1 < \bar{b}, \\ & \underline{b} < \tilde{v}[t] + \Delta_2 < \bar{b}, \end{aligned} \tag{32}$$

where $\mathrm{A}_{st} := ((a_{\bar{\omega}})_s, (a_{\bar{\omega}})_t) \in \mathbf{R}^{M \times 2}$. Since $(\Delta_s, \Delta_t)^T$ lies in the interior of the feasible region of (32), we obtain that

$$\mathrm{A}_{st}^T \mathrm{A}_{st}(\Delta_s, \Delta_t)^T = \mathrm{A}_{st}^T \tilde{f}, \quad \tilde{f} := f - A_{\bar{\omega}}\tilde{v} \in \mathbf{R}^M. \tag{33}$$

On the other hand, let $\tilde{u} := E_{\bar{\omega}}\tilde{v}$. Then $\tilde{u}$ is a feasible point of (P), and

$$\frac{\alpha}{2}\|A\bar{u} - f\|^2 + \mathcal{R}(\nabla\bar{u}) = \mathcal{F}(\bar{u}) \leqslant \mathcal{F}(\tilde{u}) = \frac{\alpha}{2}\|A\tilde{u} - f\|^2 + \mathcal{R}(\nabla\tilde{u}). \tag{34}$$

Meanwhile, since $\tilde{u} = E_{\bar{\omega}}\tilde{v}$, applying lemma 3.7 gives $(\nabla\tilde{u})[j] = 0, \forall j \in \bar{\omega}^c$. Furthermore, from (31), we have $\tilde{u}[i] = \tilde{v}[k], \tilde{u}[i+1] = \tilde{v}[s], \tilde{u}[i+n] = \tilde{v}[k]$. The fact that $\tilde{v}[k] = \tilde{v}[s] = \tilde{v}[t]$ implies $\tilde{u}[i] = \tilde{u}[i+1] = \tilde{u}[i+n]$. Thus, $(\nabla\tilde{u})[i] = 0$. Note that $i \in \bar{\omega}$. Then $\mathcal{R}(\nabla\tilde{u}) \leqslant \sharp\bar{\omega} - 1 = \mathcal{R}(\nabla\bar{u}) - 1$. Substituting it into (34) yields

$$\begin{aligned} \frac{2}{\alpha} &\leqslant \|A\tilde{u} - f\|^2 - \|A\bar{u} - f\|^2 \\ [\ (15)\ ] \quad &= \|A_{\bar{\omega}}\tilde{v} - f\|^2 - \|A_{\bar{\omega}}\bar{u}_{\bar{\omega}} - f\|^2 \\ [\ (\bar{u}_{\bar{\omega}} = \tilde{v} + \tilde{\Delta})\ ] \quad &= 2(\Delta_s, \Delta_t)\mathrm{A}_{st}^T\tilde{f} - \|\mathrm{A}_{st}(\Delta_s, \Delta_t)^T\|^2 \\ &\leqslant 2(\Delta_s, \Delta_t)\mathrm{A}_{st}^T\tilde{f} \\ [\ (33)\ ] \quad &= 2\|\mathrm{A}_{st}(\Delta_s, \Delta_t)^T\|^2 \\ &\leqslant 2\|\mathrm{A}_{st}\|_2^2\|(\Delta_s, \Delta_t)^T\|^2. \end{aligned} \tag{35}$$

Recall that $A_{\bar{\omega}} = AE_{\bar{\omega}}$ and A_{st} is a submatrix of $A_{\bar{\omega}}$. From lemma 3.10, we have

$$\|\mathrm{A}_{st}\|_2 \leqslant \|A_{\bar{\omega}}\|_2 = \|AE_{\bar{\omega}}\|_2 \leqslant \|A\|_2\|E_{\bar{\omega}}\|_2 \leqslant \sqrt{N}\|A\|_2. \tag{36}$$

Combining (35) and (36) leads to

$$\frac{2}{\alpha} \leqslant 2N\|A\|_2^2\|(\Delta_s, \Delta_t)^T\|^2.$$

Finally, we get $\|(\Delta_s, \Delta_t)^T\| \geqslant \tilde{\theta}$, where

$$\tilde{\theta} := \frac{1}{\sqrt{\alpha N}\|A\|_2}. \tag{37}$$

- *Case 2* $\sharp\,\{k, s, t\} \cap \mathbb{A}(\bar{u}_{\bar{\omega}}) = 1$.
  There are three subcases included.

Case (2.1). Suppose that $t \in \mathbb{A}(\bar{u}_{\bar{\omega}})$, $k, s \notin \mathbb{A}(\bar{u}_{\bar{\omega}})$. We decompose $\bar{u}_{\bar{\omega}}$ as: $\bar{u}_{\bar{\omega}} = \check{v} + \check{\Delta}$, where $\check{v}, \check{\Delta} \in \mathbf{R}^{n_{\bar{\omega}}}$,

$$\check{v}[j] := \begin{cases} \bar{u}_{\bar{\omega}}[t], & j = k, s, \\ \bar{u}_{\bar{\omega}}[j], & \text{otherwise}; \end{cases} \quad \text{and} \quad \check{\Delta}[j] := \begin{cases} -\Delta_t, & j = k, \\ \Delta_s - \Delta_t, & j = s, \\ 0, & \text{otherwise}. \end{cases}$$

Then $\check{\Delta}$ is a solution of the following problem

$$\begin{aligned} \min_{\Delta \in \mathbf{R}^{n_{\bar{\omega}}}} \quad & \|A_{\bar{\omega}}(\check{v} + \Delta) - f\|^2 \\ \text{s.t.} \quad & \underline{b} \cdot \mathbf{1}_{n_{\bar{\omega}}} \leqslant \check{v} + \Delta \leqslant \bar{b} \cdot \mathbf{1}_{n_{\bar{\omega}}}, \\ & \Delta[j] = 0, \quad j \in \mathbb{I}_{n_{\bar{\omega}}} \backslash \{k, s\}. \end{aligned} \tag{38}$$

Denote A_{ks} as $((a_{\bar{\omega}})_k, (a_{\bar{\omega}})_s) \in \mathbf{R}^{M\times 2}$, then $\|\mathrm{A}_{ks}\|_2 \leqslant \sqrt{N}\|A\|_2$. Since $k, s \notin \mathbb{A}(\bar{u}_{\bar{\omega}})$, $(-\Delta_t, \Delta_s - \Delta_t)^T$ is a solution of the following unconstrained problem:

$$\min_{(\Delta_1, \Delta_2)^T \in \mathbf{R}^2} \quad \|\mathrm{A}_{ks}(\Delta_1, \Delta_2)^T - (f - A_{\bar{\omega}}\check{v})\|^2.$$

Therefore,

$$\mathrm{A}_{ks}^T\mathrm{A}_{ks}(-\Delta_t, \Delta_s - \Delta_t)^T = \mathrm{A}_{ks}^T\check{f}, \quad \check{f} := f - A_{\bar{\omega}}\check{v} \in \mathbf{R}^M. \tag{39}$$

Meanwhile, $\check{u} = E_{\bar{\omega}}\check{v})$ is a feasible point of (P), and $\mathcal{F}(\bar{u}) \leqslant \mathcal{F}(\check{u})$. Together with $\check{v}[k] = \check{v}[s] = \check{v}[t]$, we have $\mathcal{R}(\nabla\bar{u}) \geqslant \mathcal{R}(\nabla\check{u}) + 1$. Then, proceeding as (35) yields

$$\frac{2}{\alpha} \leqslant 2(-\Delta_t, \Delta_s - \Delta_t)\mathrm{A}_{ks}^T\check{f} \overset{(39)}{\leqslant} 2\|\mathrm{A}_{ks}\|_2^2\|(-\Delta_t, \Delta_s - \Delta_t)^T\|^2 \leqslant 2N\|A\|_2^2\|(-\Delta_t, \Delta_s - \Delta_t)^T\|^2.$$

Thus

$$\|(-\Delta_t, \Delta_s - \Delta_t)^T\| \geqslant \frac{1}{\sqrt{\alpha N}\|A\|_2} = \tilde{\theta}. \tag{40}$$

Thanks to $(-\Delta_t, \Delta_s - \Delta_t)^T = H(\Delta_s, \Delta_t)^T$, where $H := [0, -1; 1, -1]$, we have

$$\|(-\Delta_t, \Delta_s - \Delta_t)^T\| \leqslant \|H\|_2\|(\Delta_s, \Delta_t)^T\| = \frac{2}{\sqrt{5} - 1}\|(\Delta_s, \Delta_t)^T\|. \tag{41}$$

Combining (40) and (41) yields $\|(\Delta_s, \Delta_t)^T\| \geqslant \check{\theta}$, where $\check{\theta} := \frac{\sqrt{5}-1}{2}\tilde{\theta}$. Clearly, $\check{\theta} < \tilde{\theta}$.

Case (2.2). Suppose that $s \in \mathbb{A}(\bar{u}_{\bar{\omega}}),\ k, t \notin \mathbb{A}(\bar{u}_{\bar{\omega}})$. The proof is similar to the proof in Case (2.1).

Case (2.3). Suppose that $k \in \mathbb{A}(\bar{u}_{\bar{\omega}}),\ s, t \notin \mathbb{A}(\bar{u}_{\bar{\omega}})$. The proof is similar to the proof in Case 1.

- *Case 3* $\quad \sharp\, \{k, s, t\} \cap \mathbb{A}(\bar{u}_{\bar{\omega}}) = 2$.

Since there are two indices in $\{k, s, t\}$ belonging to $\mathbb{A}(\bar{u}_{\bar{\omega}})$, we need to consider if the values of $\bar{u}_{\bar{\omega}}$ in the two indices are equal. Thus, there are six subcases here.

Case (3.1). Suppose that $k, t \in \mathbb{A}(\bar{u}_{\bar{\omega}})$, $s \notin \mathbb{A}(\bar{u}_{\bar{\omega}})$, and $\bar{u}_{\bar{\omega}}[k] = \bar{u}_{\bar{\omega}}[t]$. Now, we have $\Delta_t = 0$. Thus, we decompose $\bar{u}_{\bar{\omega}}$ as $\bar{u}_{\bar{\omega}} = \hat{v} + \hat{\Delta}$, where $\hat{v}, \hat{\Delta} \in \mathbf{R}^{n_{\bar{\omega}}}$,

$$\hat{v}[j] := \begin{cases} \bar{u}_{\bar{\omega}}[k], & j = s, \\ \bar{u}_{\bar{\omega}}[j], & \text{otherwise}; \end{cases} \quad \text{and} \quad \hat{\Delta}[j] := \begin{cases} \Delta_s, & j = s, \\ 0, & \text{otherwise}. \end{cases}$$

Then $\hat{\Delta}$ is a solution of the following problem

$$\begin{aligned} \min_{\Delta \in \mathbf{R}^{n_{\bar{\omega}}}} \quad & \|A_{\bar{\omega}}(\hat{v} + \Delta) - f\|^2 \\ \text{s.t.} \quad & \underline{b} \cdot \mathbf{1}_{n_{\bar{\omega}}} \leqslant \hat{v} + \Delta \leqslant \bar{b} \cdot \mathbf{1}_{n_{\bar{\omega}}}, \\ & \Delta[j] = 0, \quad j \in \mathbb{I}_{n_{\bar{\omega}}} \backslash \{s\}. \end{aligned}$$

Denote A_s as $((a_{\bar{\omega}})_s) \in \mathbf{R}^{M \times 1}$, then $\|\mathrm{A}_s\|_2 \leqslant \sqrt{N}\|A\|_2$. It follows that

$$\Delta_s \in \arg\min_{\Delta_1 \in \mathbf{R}} \quad \|\mathrm{A}_s \Delta_1 - (f - A_{\bar{\omega}}\hat{v})\|^2.$$

Thus,

$$\mathrm{A}_s^T \mathrm{A}_s \Delta_s = \mathrm{A}_s^T \hat{f}, \quad \hat{f} := f - A_{\bar{\omega}}\hat{v} \in \mathbf{R}^M. \tag{42}$$

Similarly, $\hat{u} = E_{\bar{\omega}}\hat{v}$ is a feasible point of (P), and $\mathcal{F}(\bar{u}) \leqslant \mathcal{F}(\hat{u})$. The fact that $\hat{v}[k] = \hat{v}[s] = \hat{v}[t]$ implies $\mathcal{R}(\nabla \bar{u}) \geqslant \mathcal{R}(\nabla \hat{u}) + 1$, and then

$$\frac{2}{\alpha} \leqslant 2\Delta_s \mathrm{A}_s^T \hat{f} \overset{(42)}{\leqslant} 2\|\mathrm{A}_s\|_2^2 |\Delta_s|^2 \leqslant 2N\|A\|_2^2 |\Delta_s|^2.$$

Hence,

$$|\Delta_s| \geqslant \frac{1}{\sqrt{\alpha}N\|A\|_2} = \tilde{\theta},$$

which indicates $\|(\Delta_s, \Delta_t)^T\| \geqslant \tilde{\theta}$.

Case (3.2). Suppose that $k, s \in \mathbb{A}(\bar{u}_{\bar{\omega}})$, $t \notin \mathbb{A}(\bar{u}_{\bar{\omega}})$, and $\bar{u}_{\bar{\omega}}[k] = \bar{u}_{\bar{\omega}}[s]$. Then we have $\Delta_s = 0$. The proof is similar to the proof in Case (3.1).

Case (3.3). Suppose that $s, t \in \mathbb{A}(\bar{u}_{\bar{\omega}}), k \notin \mathbb{A}(\bar{u}_{\bar{\omega}})$, and $\bar{u}_{\bar{\omega}}[s] = \bar{u}_{\bar{\omega}}[t]$. Then we have $\Delta_s = \Delta_t$. The proof is also similar to the proof in Case (3.1), and we have $\|(\Delta_s, \Delta_t)^T\| \geqslant 2\tilde{\theta}$.

Case (3.4). Suppose that $k, t \in \mathbb{A}(\bar{u}_{\bar{\omega}})$, $s \notin \mathbb{A}(\bar{u}_{\bar{\omega}})$, and $\bar{u}_{\bar{\omega}}[k] \neq \bar{u}_{\bar{\omega}}[t]$. Then we have $|\Delta_t| = |\underline{b} - \bar{b}|$. Thus, $\|(\Delta_s, \Delta_t)^T\| \geqslant \ddot{\theta}$, where

$$\ddot{\theta} := |\underline{b} - \bar{b}|. \tag{43}$$

Case (3.5). Suppose that $k, s \in \mathbb{A}(\bar{u}_{\bar{\omega}}), t \notin \mathbb{A}(\bar{u}_{\bar{\omega}})$, and $\bar{u}_{\bar{\omega}}[k] \neq \bar{u}_{\bar{\omega}}[s]$. Then we have $|\Delta_s| = |\underline{b} - \bar{b}|$. Thus, $\|(\Delta_s, \Delta_t)^T\| \geqslant \ddot{\theta}$.

Case (3.6). Suppose that $s, t \in \mathbb{A}(\bar{u}_{\bar{\omega}}), k \notin \mathbb{A}(\bar{u}_{\bar{\omega}})$, and $\bar{u}_{\bar{\omega}}[s] \neq \bar{u}_{\bar{\omega}}[t]$. Since $s, t \in \mathbb{A}(\bar{u}_{\bar{\omega}})$ and $\bar{u}_{\bar{\omega}}[s] \neq \bar{u}_{\bar{\omega}}[t]$, we have

$$\begin{aligned}\|(\Delta_s, \Delta_t)^T\|^2 &= (\bar{u}_{\bar{\omega}}[k] - \bar{u}_{\bar{\omega}}[s])^2 + (\bar{u}_{\bar{\omega}}[k] - \bar{u}_{\bar{\omega}}[t])^2 \\ &= (\bar{u}_{\bar{\omega}}[k] - \underline{b})^2 + (\bar{u}_{\bar{\omega}}[k] - \bar{b})^2 \\ &\geqslant \frac{(\underline{b} - \bar{b})^2}{2},\end{aligned}$$

Therefore $\|(\Delta_s, \Delta_t)^T\| \geqslant \dot{\theta}$, where

$$\dot{\theta} := \frac{\sqrt{2}|\underline{b} - \bar{b}|}{2}.$$

Clearly, $\dot{\theta} < \ddot{\theta}$.

- *Case 4* $\quad \sharp\,\{k, s, t\} \cap \mathbb{A}(\bar{u}_{\bar{\omega}}) = 3$.
  In this case, we have $|\Delta_s| = |\underline{b} - \bar{b}|$ or $|\Delta_t| = |\underline{b} - \bar{b}|$. Otherwise, we would have $|\Delta_s| = |\Delta_t| = 0$, which means $\|(\Delta_s, \Delta_t)^T\| = 0$. Thus, $\|(\Delta_s, \Delta_t)^T\| \geqslant \ddot{\theta}$, where $\ddot{\theta}$ is defined in (43).

In summary of all cases above, we have $\|(\nabla u)[i]\| \geqslant \theta$, where $\theta = \min\{\check{\theta}, \dot{\theta}\}$ is independent of *f*. This completes the proof. □

By theorem 4.1, the following corollary is immediate.

Corollary 4.2. *The lower bound θ in theorem 4.1 is a decreasing function of α.*

In the above theorem, we have proved that for any $f \in \mathbf{R}^M$, the solutions of (P) have a uniform lower bound for the ℓ_2 norm of nonzero discrete gradients. This conclusion suggests that our regularization applied to (P) yields restorations with neat edges. However, this property can not be extended to local minimizers of (P). Here we provide an example. Suppose that A has full column rank. From theorem 3.5, we know $u_f = (A^TA)^{-1}Af$ is a local minimizer of (P). If the uniform lower bound $\vartheta > 0$ for local minimizers exists, it is less than $\|(\nabla u_f)[i]\|, i \in \mathbb{I}_N$ satisfying $(\nabla u_f)[i] \neq 0$. Let $f \to 0$, then $\vartheta \to 0$, which leads to a contradiction.

4.2. Piecewise-constant dependency on the parameter α

In actual application, given an observed image *f*, we shall adjust the parameter α in (P) frequently to obtain a perfect restoration. The solution of (P) is greatly influenced by α, and the goal in this subsection is to find the relationship between them.

Let $\alpha \in [0, +\infty)$. For better presentation, we define (P_α) as follows:

$$(\mathrm{P}_\alpha) \qquad \begin{cases} \min_{u \in \mathbf{R}^N} & \mathcal{F}_\alpha(u) := \frac{\alpha}{2}\|Au - f\|^2 + \mathcal{R}(\nabla u) \\ \text{s.t.} & \underline{b} \cdot \mathbf{1}_N \leqslant u \leqslant \bar{b} \cdot \mathbf{1}_N. \end{cases} \tag{44}$$

It is worth to stress that there is no restriction on A in this subsection. Based on (P_α), we denote

$$\mathcal{J}(\alpha) := \text{the optimal value of } (\mathrm{P}_\alpha), \tag{45}$$

$$\mathcal{U}^g(\alpha) := \{u \in \mathbf{R}^N \ : \ u \ \text{ solves } (\mathrm{P}_\alpha)\}, \tag{46}$$

$$\mathcal{U}^l \ := \{u \in \mathbf{R}^N \ : \ u \ \text{ is a local minimizer of } (\mathrm{P}_\alpha)\}. \tag{47}$$

Clearly, $\mathcal{J}(\alpha) = \mathcal{F}_\alpha(u), \forall\, u \in \mathcal{U}^g(\alpha)$. Note that $\mathcal{U}^l$ is independent of α according to theorem 3.5. That is, whatever α is, the local minimizer set of (P_α) is the same. Since a global minimizer of (P_α) is a local minimizer, we have $\mathcal{U}^g(\alpha) \subseteq \mathcal{U}^l$.

Next we need to find the elements of $\mathcal{U}^g(\alpha)$ from $\mathcal{U}^l$. Checking whether a local minimizer is a global minimizer requires us to compare its energy value with the optimal value. The energy function $\mathcal{F}_\alpha(u)$ of (P_α) consists of two terms: the fidelity term $\|Au - f\|^2$ and the regularization term $\mathcal{R}(\nabla u)$. Since $\forall\, u \in \mathbf{R}^N, (\nabla u)[N] = 0, \mathcal{R}(\nabla u)$ has only N values: $0, \cdots, N-1$. Hence, we can partition $\mathcal{U}^l$ into at most N subsets according to the regularization term values of its elements:

$$\mathcal{U}^l_k := \mathcal{U}^l \cap \{u \in \mathbf{R}^N \ : \ \mathcal{R}(\nabla u) = k\}, \quad \forall\, k \in \mathbb{I}^0_{N-1}.$$

It is possible that there exists $0 < k_0 \leqslant N-1$ such that $\mathcal{U}^l_{k_0} = \emptyset$. Besides, $\mathcal{U}^l_0 = \{u \in \mathbf{R}^N : u$ solves $(\mathrm{Q}_\emptyset)\} \neq \emptyset$ according to propositions 3.2 and 3.4.

It follows from $\mathcal{U}^g(\alpha) \subseteq \mathcal{U}^l$ that

$$\begin{aligned}\mathcal{J}(\alpha) &= \min_{u \in \mathcal{U}^l} \mathcal{F}_\alpha(u) \\ &= \min_{k \in \mathbb{I}^0_{N-1}} \min_{u \in \mathcal{U}^l_k} \frac{\alpha}{2}\|Au - f\|^2 + k \\ &= \min_{k \in \mathbb{I}^0_{N-1}} \frac{\alpha}{2}(\min_{u \in \mathcal{U}^l_k} \|Au - f\|^2) + k. \end{aligned} \tag{48}$$

We highlight the local minimizers which have the minimal fidelity term value in each $\mathcal{U}^l_k$, and define the following sequences:

for $k \in \mathbb{I}^0_{N-1}$ satisfying $\mathcal{U}^l_k \neq \emptyset$,

$$\mathcal{C}_k := \arg\min_{u \in \mathcal{U}^l_k} \|Au - f\|^2, \qquad c_k := \min_{u \in \mathcal{U}^l_k} \|Au - f\|^2;$$

for $k = -1$, or $k \in \mathbb{I}^0_{N-1}$ satisfying $\mathcal{U}^l_k \neq \emptyset$,

$$\mathcal{C}_k := \emptyset, \qquad\qquad c_k := +\infty.$$

We can see that $\mathcal{C}_0 = \mathcal{U}^l_0 \neq \emptyset$, and $c_0 < +\infty$. Each $\mathcal{C}_k$ is formed by all of the local minimizers whose fidelity term values are c_k and regularization term values are k. For any $k \in \mathbb{I}^0_{N-1}$, we have $\mathcal{C}_k \subseteq \mathcal{U}^l_k$, and $\forall\, u \in \mathcal{C}_k, \mathcal{F}_\alpha(u) = \frac{\alpha}{2}c_k + k$. Thus, for any $k \in \mathbb{I}^0_{N-1}$, we define

$$\mathcal{J}_k(\alpha) := \frac{c_k}{2}\alpha + k = \mathcal{F}_\alpha(u), \quad \forall\, u \in \mathcal{C}_k. \tag{49}$$

Each $\mathcal{J}_k(\alpha), \alpha \in [0, +\infty)$ is a linear function, with $\frac{c_k}{2}$ as its slope. Thus, combining (48) and (49) gives

$$\mathcal{J}(\alpha) = \min_{k \in \mathbb{I}^0_{N-1}} \mathcal{J}_k(\alpha), \qquad \mathcal{U}^g(\alpha) \subseteq \bigcup_{k \in \mathbb{I}^0_{N-1}} \mathcal{C}_k. \tag{50}$$

Remark 4.3. It can be checked by theorem 3.16 that if A has full column rank and $f \in \mathbf{R}^M \backslash Z$ with Z defined in (29), then $\sharp\mathcal{C}_k = 1$.

Now, we have determined the optimal value and optimal solutions from the local minimizers of (P_α). And then we need to solve the above optimization problem to obtain $\mathcal{U}^g(\alpha)$. For better understanding, we present an example at first.

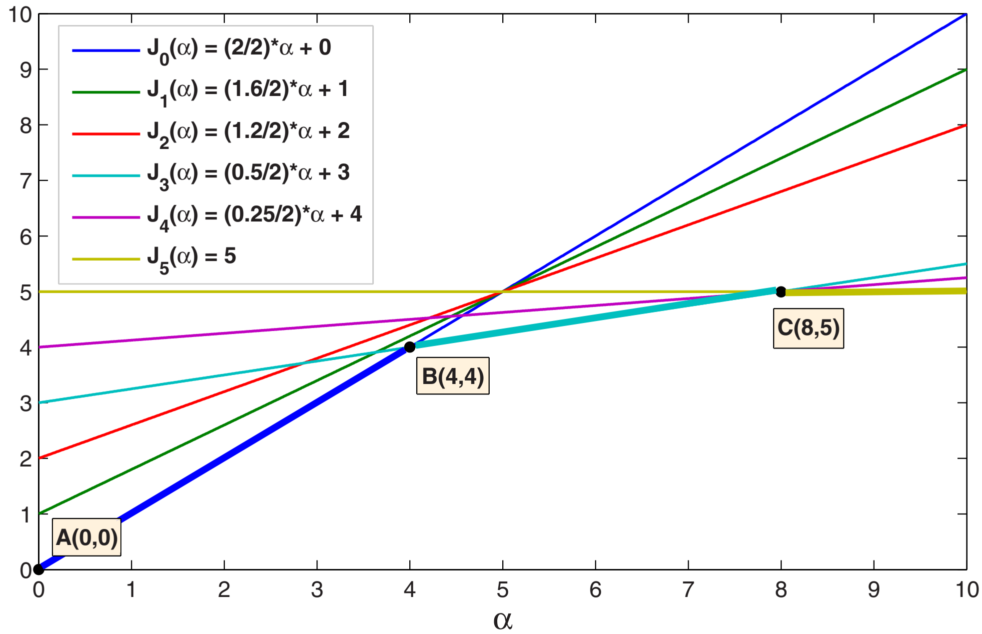


Figure 3. Plots of $\mathcal{J}_k, k \in \mathbb{K}$ in example 4.4. The segments marked as bold is the plot of $\mathcal{J}(\alpha)$.

Example 4.4. We set $N = 6$ and $c_0 = 2, c_1 = 1.6, c_2 = 1.2, c_3 = 0.5, c_4 = 0.25, c_5 = 0$. Then $\mathcal{J}_k(\alpha)$ for $k \in \mathbb{I}_5^0$ can be obtained by (49). We show the plots of $\mathcal{J}_k(\alpha), k \in \mathbb{I}_5^0$ in figure 3. Comparing all these lines, it is straightforward to see that $\mathcal{J}(\alpha)$ is a piecewise linear function:

$$\mathcal{J}(\alpha) = \begin{cases} \mathcal{J}_0(\alpha), & \alpha \in [0, 4), \\ \mathcal{J}_3(\alpha), & \alpha \in [4, 8), \\ \mathcal{J}_5(\alpha), & \alpha \in [8, +\infty). \end{cases}$$

From figure 3, we can see that $\mathcal{J}(\alpha)$ has several turning points: $\mathbf{A}(0,0), \mathbf{B}(4,4), \mathbf{C}(8,5)$. In terms of the abscissa values of these turning points, the positive axis $[0, +\infty)$ is partitioned into some subintervals. In each subinterval, $\mathcal{J}(\alpha)$ coincides with one of $\{\mathcal{J}_k(\alpha)\}$, e.g. $\mathcal{J}_{k_0}(\alpha)$; along the starting point of this subinterval towards the right, once $\mathcal{J}_{k_0}(\alpha)$ intersects with another line, the intersection is a turning point of $\mathcal{J}(\alpha)$. Notice that there are three lines intersecting at $\alpha = 8$, and the line with minimal slope coincides with $\mathcal{J}(\alpha)$ when $\alpha \geqslant 8$.

According to example 4.4, $\mathcal{J}(\alpha), \alpha \in [0, +\infty)$ is a piecewise linear function, and the key point is to find the turning points of $\mathcal{J}(\alpha)$.

Let $k_0 \in \mathbb{I}_{N-1}$. If there exists k_1 satisfying $0 \leqslant k_1 < k_0$ and $c_{k_1} \leqslant c_{k_0}$, then we have

$$\mathcal{J}_{k_1}(\alpha) = \frac{\alpha}{2} c_{k_1} + k_1 < \frac{\alpha}{2} c_{k_0} + k_0 = \mathcal{J}_{k_0}(\alpha),$$

which implies $\mathcal{J}_{k_0}(\alpha)$ could never equal to $\mathcal{J}(\alpha)$. Hence, the optimal value $\mathcal{J}(\alpha)$ is the minimum among $\mathcal{J}_k(\alpha)$ whose indices are in

$$\mathbb{K} := (\{k \in \mathbb{I}_{N-1}^0 \ : \ c_k < c_j, \ \forall -1 \leqslant j < k\}, <).$$

Definitely, $0 \in \mathbb{K}$. The largest index in $\mathbb{K}$ is denoted by K, i.e. $K := \max\{k : k \in \mathbb{K}\}$. Then $\{c_k\}_{k\in\mathbb{K}}$ is strictly decreasing and $c_K = \min\{c_k, k \in \mathbb{I}_{N-1}^0\}$. Consequently, we have $\forall\, \alpha \in [0, +\infty)$,

$$\mathcal{J}(\alpha) = \min_{k\in\mathbb{K}} \mathcal{J}_k(\alpha), \quad \mathcal{K}(\alpha) := \arg\min_{k\in\mathbb{K}} \mathcal{J}_k(\alpha), \quad \mathcal{U}^g(\alpha) = \bigcup_{k\in\mathcal{K}(\alpha)} \mathcal{C}_k. \tag{51}$$

Note that the turning points are always generated from intersections of some lines in $\{\mathcal{J}_k(\alpha), k \in \mathbb{K}\}$. For any $k, j \in \mathbb{K}, k \neq j$,

$$\mathcal{J}_k(\alpha) = \mathcal{J}_j(\alpha) \quad \Longleftrightarrow \quad \alpha = \phi(k,j), \quad \text{where } \phi(k,j) := \frac{2(j-k)}{c_k - c_j}.$$

Let $k \in \mathbb{K}$. Given $\alpha \geqslant 0$, $\mathcal{J}(\alpha) = \mathcal{J}_k(\alpha)$ if and only if $\mathcal{J}_k(\alpha) \leqslant \mathcal{J}_j(\alpha), \forall j \in \mathbb{K}, j \neq k$, which is equivalent to $\frac{\alpha}{2}c_k + k \leqslant \frac{\alpha}{2}c_j + j, \forall j \in \mathbb{K}, j \neq k$. Since $\{c_k\}_{k\in\mathbb{K}}$ is strictly decreasing, we have if $j > k$, then $c_j < c_k$ and

$$\mathcal{J}_k(\alpha) \leqslant (<) \mathcal{J}_j(\alpha), \forall j \in \mathbb{K}, j > k \quad \Longleftrightarrow \quad \alpha \leqslant (<) \min_{j\in\mathbb{K}, j>k} \phi(k,j); \tag{52}$$

if $j < k$, then $c_j > c_k$ and

$$\mathcal{J}_k(\alpha) \leqslant (<) \mathcal{J}_j(\alpha), \forall j \in \mathbb{K}, j < k \quad \Longleftrightarrow \quad \alpha \geqslant (>) \max_{j\in\mathbb{K}, j<k} \phi(k,j). \tag{53}$$

Based on above, we give the following definition.

Definition 4.5. The parameters $\{\alpha_t\}, \{k_t\}$ are defined iteratively, initialized with $\alpha_0 = 0, \Lambda_0 = \{0\}, k_0 = 0$:

$$\forall\, t = 1, 2, \cdots \qquad \begin{cases} \alpha_t & := \min_{j\in\mathbb{K}, j>k_{t-1}} \phi(k_{t-1}, j), \\ \Lambda_t & := \{ j \ : \ j \in \mathbb{K}, j > k_{t-1}, \phi(k_{t-1}, j) = \alpha_t \}, \\ k_t & := \max\{ j \ : \ j \in \Lambda_t \}. \end{cases} \tag{54}$$

Owing to the finiteness of $\mathbb{K}$, the iteration above ends in finite steps. The number of elements in $\{\alpha_t\}$ is denoted by T.

Without loss of generality, we denote $k_{-1} := -1$ and $\alpha_{T+1} := +\infty$. The index t refers to the index of the turning point. For $t \in \mathbb{I}_T^0$, α_t is the abscissa of a turning point; Λ_t includes all of the indices of $\mathcal{J}_k(\alpha)$ which intersect with $\mathcal{J}(\alpha)$ at this turning point; among $\{\mathcal{J}_k(\alpha), k \in \Lambda_t\}$, $\mathcal{J}_{k_t}(\alpha)$ has the minimal slope and coincides with $\mathcal{J}(\alpha)$ from this turning point to the next one. Furthermore, $\{(\ \alpha_t, \mathcal{J}_{k_t}(\alpha_t)\ ), t \in \mathbb{I}_T^0\}$ are all of the turning points of $\mathcal{J}(\alpha)$ (see theorem 4.6). For example 4.4, we can verify that, $T = 2$ and $(\alpha_0, \mathcal{J}_{k_0}(\alpha_0)) = (0, 0), (\alpha_1, \mathcal{J}_{k_1}(\alpha_1)) = (4, 4), (\alpha_2, \mathcal{J}_{k_2}(\alpha_2)) = (8, 5)$.

Theorem 4.6. *For the problem* (P_α)*, its optimal value* $\mathcal{J}(\alpha)$ *and global minimizer set* $\mathcal{U}^g(\alpha)$ *satisfy:* $\forall\, t \in \mathbb{I}_T^0$,

$$\mathcal{J}(\alpha) = \mathcal{J}_{k_t}(\alpha), \quad \forall\, \alpha \in [\alpha_t, \alpha_{t+1}), \qquad \mathcal{U}^g(\alpha) = \begin{cases} \mathcal{C}_{k_t}, & \alpha \in (\alpha_t, \alpha_{t+1}), \\ \bigcup_{k\in\Lambda_t} \mathcal{C}_k, & \alpha = \alpha_t. \end{cases} \tag{55}$$

Proof. The proof is divided into three steps.

- *step 1: show that* $\{k_t\}_{t\in\mathbb{I}_T}$ *strictly increases, and* $k_T = K$
  This is obvious from the definition of $\{k_t\}$.

- *step 2: show that* $\{\alpha_t\}_{t\in\mathbb{I}_T}$ *strictly increases, and satisfies* $\forall\, t\in\mathbb{I}_T, \alpha_t=\phi(k_{t-1},k_t)$*, and*

$$\max_{j\in\mathbb{K},j<k_t}\phi(k_t,j)=\alpha_t=\min_{j\in\mathbb{K},j>k_{t-1}}\phi(k_{t-1},j)$$

Let $t\in\mathbb{I}_T$. According to the definition of α_t, we have

$$\alpha_t=\phi(k_{t-1},k_t)=\frac{2(k_t-k_{t-1})}{c_{k_{t-1}}-c_{k_t}}<\frac{2(k_{t+1}-k_{t-1})}{c_{k_{t-1}}-c_{k_{t+1}}}. \tag{56}$$

Here we give an inequality which can be easily verified: if $a,b,c,d>0$ and $a<c, b<d, \frac{b}{a}<\frac{d}{c}$, then

$$\frac{d-b}{c-a}>\frac{b}{a}. \tag{57}$$

Since $\{k_t\}$ is strictly increasing and $\{c_k\}$ is strictly decreasing, we have $k_t-k_{t-1}<k_{t+1}-k_{t-1}$ and $c_{k_{t-1}}-c_{k_t}<c_{k_{t-1}}-c_{k_{t+1}}$. Thus, from (56), we obtain

$$\alpha_{t+1}=\phi(k_t,k_{t+1})=\frac{2(k_{t+1}-k_t)}{c_{k_t}-c_{k_{t+1}}}=2\frac{(k_{t+1}-k_{t-1})-(k_t-k_{t-1})}{(c_{k_{t-1}}-c_{k_{t+1}})-(c_{k_{t-1}}-c_{k_t})}>\frac{2(k_t-k_{t-1})}{c_{k_{t-1}}-c_{k_t}}=\alpha_t,$$

which indicates that $\{\alpha_t\}$ is strictly increasing.

Assume that there exists $j\in\{j\in\mathbb{K}:k_{t-1}<j<k_t\}$, then

$$\alpha_t=\frac{2(k_t-k_{t-1})}{c_{k_{t-1}}-c_{k_t}}\leqslant\frac{2(j-k_{t-1})}{c_{k_{t-1}}-c_j}.$$

Also, we introduce an inequality like (57): for $a,b,c,d>0$ with $a>c, b>d, \frac{b}{a}\leqslant\frac{d}{c}$, then

$$\frac{d-b}{c-a}\leqslant\frac{b}{a}. \tag{58}$$

Hence, for any $j\in\{j\in\mathbb{K}:k_{t-1}<j<k_t\}$,

$$\phi(k_t,j)=\frac{2(j-k_t)}{c_{k_t}-c_j}=2\frac{(j-k_{t-1})-(k_t-k_{t-1})}{(c_{k_{t-1}}-c_j)-(c_{k_{t-1}}-c_{k_t})}\leqslant\frac{2(k_t-k_{t-1})}{c_{k_{t-1}}-c_{k_t}}=\alpha_t.$$

As $\alpha_t=\phi(k_t,k_{t-1})$, we have $\alpha_t=\max\{\phi(k_t,j)\ :\ j\in\mathbb{K},k_{t-1}\leqslant j<k_t\}$.

Similarly,

$$\alpha_{t-s}=\max\{\phi(k_{t-s},j)\ :\ j\in\mathbb{K},k_{(t-s-1)}\leqslant j<k_{(t-s)}\},\quad s=1,\cdots,t-1.$$

Since $\{\alpha_t\}$ is strictly increasing, we have $\alpha_t>\alpha_{t-s}, s=1,\cdots,t-1$. Thus

$$\alpha_t=\max_{j\in\mathbb{K},j<k_t}\phi(k_t,j).$$

- *step 3: prove (55)*

  Take any $t\in\mathbb{I}_T^0$. From step 4, $\forall\,\alpha\in(\alpha_t,\alpha_{t+1})$,

$$\alpha<\alpha_{t+1}=\min_{j\in\mathbb{K},j>k_t}\phi(k_t,j)\quad\overset{(52)}{\Longrightarrow}\quad\mathcal{J}_{k_t}(\alpha)<\mathcal{J}_j(\alpha),\forall\,j\in\mathbb{K},\,j>k_t;$$

$$\alpha>\alpha_t=\max_{j\in\mathbb{K},j<k_t}\phi(k_t,j)\quad\overset{(53)}{\Longrightarrow}\quad\mathcal{J}_{k_t}(\alpha)<\mathcal{J}_j(\alpha),\forall\,j\in\mathbb{K},\,j<k_t.$$

Thus, $\mathcal{J}_{k_t}(\alpha)<\mathcal{J}_j(\alpha),\forall\,j\in\mathbb{K},j\neq k_t$. Then, we have

$$\forall\,t\in\mathbb{I}_T^0,\quad\mathcal{J}(\alpha)=\mathcal{J}_{k_t}(\alpha),\qquad\mathcal{U}^g(\alpha)=\mathcal{C}_{k_t},\quad\forall\,\alpha\in(\alpha_t,\alpha_{t+1}). \tag{59}$$

By (51), we have

$$\mathcal{J}(0)=\min_{k\in\mathbb{K}}\mathcal{J}_k(0)=\min_{k\in\mathbb{K}}k=\mathcal{J}_0(0)=0. \tag{60}$$

Take any $t\in\mathbb{I}_T$. Similarly, we have

$$\alpha_t=\min_{j\in\mathbb{K},j>k_{t-1}}\phi(k_{t-1},j)\quad\overset{(52)}{\Longrightarrow}\quad\mathcal{J}_{k_{t-1}}(\alpha_t)\leqslant\mathcal{J}_j(\alpha_t),\forall j\in\mathbb{K},\ j>k_{t-1};$$

$$\alpha_t=\max_{j\in\mathbb{K},j<k_t}\phi(k_t,j)\quad\overset{(53)}{\Longrightarrow}\quad\mathcal{J}_{k_t}(\alpha_t)\leqslant\mathcal{J}_j(\alpha_t),\forall j\in\mathbb{K},\ j<k_t.$$

Since $\alpha_t=\phi(k_{t-1},k_t)$, we have $\mathcal{J}_{k_{t-1}}(\alpha_t)=\mathcal{J}_{k_t}(\alpha_t)\leqslant\mathcal{J}_j(\alpha_t),\forall j\in\mathbb{K}$. Moreover, it follows from the definition of Λ_t that $\forall k\in\Lambda_t,\mathcal{J}_k(\alpha_t)=\mathcal{J}_{k_t}(\alpha_t)<\mathcal{J}_j(\alpha_t),\forall j\in\mathbb{K},j\notin\Lambda_t$. Therefore, we have

$$\forall t\in\mathbb{I}_T^0,\quad\mathcal{J}(\alpha_t)=\mathcal{J}_{k_t}(\alpha_t),\qquad\mathcal{U}^g(\alpha_t)=\bigcup_{k\in\Lambda_t}\mathcal{C}_k. \tag{61}$$

Finally, combining (59) and (61) yields (55). □

From the proof, we can see that $\{k_t\}$ is increasing and $\{c_k\}$ is decreasing. Thus, the elements in $\mathcal{U}^g(\alpha)$ has decreasing fidelity term values and growing regularization term values when $\alpha\to+\infty$. As $\mathcal{J}(\alpha)$ is a piecewise linear function with slopes equal or greater than 0 in each subinterval, the following statement is obvious.

Corollary 4.7. *$\mathcal{J}(\alpha),\alpha\in[0,+\infty)$ is piecewise linear, continuous and increasing, and $\mathcal{U}^g(\alpha),\alpha\in[0,+\infty)$ is piecewise constant.*

Recalling (1), the true image $\underline{u}$ is not always a solution of (P_α). However, we can estimate the distance between $\underline{u}$ and the solutions of (P_α).

Proposition 4.8. *Let $\alpha>\alpha_T$. For (P_α), its global minimizer set $\mathcal{U}^g(\alpha)$ satisfies*

$$\mathcal{U}^g(\alpha)=\arg\min_{\underline{b}\cdot\mathbf{1}_N\leqslant u\leqslant\bar{b}\cdot\mathbf{1}_N}\|Au-f\|^2. \tag{62}$$

Furthermore, if there exists $\delta>0$ such that $A^TA-\delta I$ is positive semi-definite, then for any $u\in X$, where X is the feasible region of (P_α), we have

$$\frac{\delta}{2}\|u-\bar{u}\|^2\leqslant\|Au-f\|^2,\quad\forall\bar{u}\in\mathcal{U}^g(\alpha). \tag{63}$$

Especially, when $u=\underline{u}$, where $\underline{u}$ is the true image, the right term in the inequality above is the noise level.

Proof. From theorem 4.6, we can see that $\mathcal{U}^g(\alpha)=\mathcal{C}_K,\forall\alpha>\alpha_T$. Since

$$\begin{aligned}c_K&=\min\{c_k\ :\ k\in\mathbb{I}_{N-1}^0\}\\&=\min_{k\in\mathbb{I}_{N-1}^0}\min_{u\in\mathcal{U}_k^l}\|Au-f\|^2\\&=\min_{u\in\mathcal{U}^l}\|Au-f\|^2\\ [\text{ theorem 3.5 }]\quad&=\min\{\|Au-f\|^2\ :\ u\ \text{ solves }(\mathrm{Q}_\omega),\ \omega\subseteq\mathbb{I}_N\}.\end{aligned}$$

It follows from the definition of (Q_ω) in (5) that

$$c_K = \min_{\underline{b}\cdot\mathbf{1}_N \leqslant u \leqslant \bar{b}\cdot\mathbf{1}_N} \|Au - f\|^2.$$

Then,

$$\mathcal{U}^g(\alpha) = \mathcal{C}_K = \arg \min_{\underline{b}\cdot\mathbf{1}_N \leqslant u \leqslant \bar{b}\cdot\mathbf{1}_N} \|Au - f\|^2, \quad \forall\, \alpha > \alpha_T. \tag{64}$$

Denote $\mathcal{I}_X(u)$ as the indicator function of X, which is the feasible region of (P_α). Then, we have

$$\mathcal{U}^g(\alpha) = \arg \min_{u\in\mathbf{R}^N} \mathcal{G}(u) := \|Au - f\|^2 + \mathcal{I}_X(u). \tag{65}$$

According to proposition 2.3, $\forall\, \bar{u} \in \mathcal{U}^g(\alpha), 0 \in \partial\mathcal{G}(\bar{u}) = 2A^T(A\bar{u} - f) + \partial\mathcal{I}_X(\bar{u})$, followed by

$$-2A^T(A\bar{u} - f) \in \partial\mathcal{I}_X(\bar{u}).$$

Since X is closed and convex, by proposition 2.4, we have

$$\langle 2A^T(A\bar{u} - f), v - \bar{u}\rangle \geqslant 0, \quad \forall\, v \in X. \tag{66}$$

Denote $\mathcal{H}(u) := \|Au - f\|^2$. Note that $\mathcal{H}(u)$ is strongly convex since $A^TA - \delta I$ is positive semi-definite. Then for any $u \in X$,

$$\begin{aligned} & \mathcal{H}(u) \geqslant \mathcal{H}(\bar{u}) + \langle A^T(A\bar{u} - f), u - \bar{u}\rangle + \frac{\delta}{2}\|u - \bar{u}\|^2 \\ [\,(66)\,] \qquad & \geqslant \mathcal{H}(\bar{u}) + \frac{\delta}{2}\|u - \bar{u}\|^2 \\ & \geqslant \frac{\delta}{2}\|u - \bar{u}\|^2. \end{aligned}$$

Thus, $\frac{\delta}{2}\|u - \bar{u}\|^2 \leqslant \|Au - f\|^2$. □

According to proposition 4.8, when A has full column rank and α is large, $\underline{u}$ is sufficiently close to the solutions of (P_α) as long as the noise level $\|A\underline{u} - f\|^2$ is small enough.

5. Experiments

In this section, we present some experiments to demonstrate our main theoretical analyses, including:

(a) referring to theorem 3.5, the set of local minimizers is equivalent to the set of solutions of (Q_ω), which is independent of α;

(b) referring to theorem 3.16, if A has full column rank, then any two local minimizers of (P) practically have different energy values;

(c) referring to theorem 4.6, for problem (P_α) in (44), its optimal value $\mathcal{J}(\alpha)$ is piecewise linear, and its global minimizer set $\mathcal{U}^g(\alpha)$ is piecewise constant; moreover, if α is large enough, the distance between the true image $\underline{u}$ and the solutions of (P_α) is bounded by the noise level $\|A\underline{u} - f\|^2$;

(d) referring to theorem 4.1, there exists a uniform lower bound for the ℓ_2 norm of nonzero entries in $\nabla\bar{u}$, where $\bar{u}$ is an arbitrary solution of (P): the lower bound is independent of the observed image *f*, and it is decreasing when the parameter α is increasing.

Now, we provide the configurations of our experiments. All of the digits are accurate to sixteen decimal places, and for better readability, they are showed in two decimal places. Since it is an *NP-hard* problem to find the global minimizers of (P), we use images in small size. In part(a), we use a 6×4 test image; in part(b)(c)(d), we use 3×3 images. We adopt a 3×3 Gaussian blurring kernel (fspecial('gaussian')), and preprocess it to be a matrix A. Clearly, A has full column rank, and $A^TA - 0.21I$ is positive semi-definite.

When α and f are given, all local minimizers of (P) are acquired by solving all of the problems in $\{(\mathrm{Q}_\omega) : \omega \subseteq \mathbb{I}_N\}$. There are 2^N problems in it. These bounded-variable least square problems (Q_ω) are transformed into $(\mathcal{Q}_\omega)$ and then solved by a combinational search. Due to the fact that different ω may have the same block partition and other factors, some (Q_ω) have the same solution. Thus, the number of local minimizers of (P) is usually less than 2^N. Moreover, the global minimizers of (P) are acquired by a search among all local minimizers.

(a) In this part, we give an example with 6×4 image. We denote $\alpha = 100, \underline{b} = -\infty$, $\bar{b} = +\infty$ and

$$\underline{\boldsymbol{u}} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & 1 & 1 \end{bmatrix}, \qquad \boldsymbol{f} = \mathcal{A}\underline{\boldsymbol{u}} + \boldsymbol{\eta} = \begin{bmatrix} 1.00 & 0.89 & 0.11 & 0.01 \\ 1.02 & 0.90 & 0.11 & 0.01 \\ 0.87 & 0.85 & 0.20 & 0.11 \\ 0.12 & 0.22 & 0.81 & 0.88 \\ 0.00 & 0.09 & 0.89 & 1.01 \\ -0.01 & 0.14 & 0.91 & 1.02 \end{bmatrix}, \tag{67}$$

where $\boldsymbol{\eta} \in \mathbf{R}^{3\times3}$ is a random noise with $\|\boldsymbol{\eta}\|^2 = 0.0038$.

After calculation, we obtain 1651 657 local minimizers, which are independent of α. Among all of the local minimizers, according to the objective function value, we show the global minimizer $\bar{\boldsymbol{u}}$, the best non-global minimizer $\tilde{\boldsymbol{u}}$, the worst non-global minimizer $\hat{\boldsymbol{u}}$ as follows:

$$\bar{\boldsymbol{u}} = \begin{bmatrix} 1.00 & 1.00 & 0.01 & 0.01 \\ 1.00 & 1.00 & 0.01 & 0.01 \\ 1.00 & 1.05 & 0.01 & 0.01 \\ 0.01 & 0.01 & 1.00 & 1.00 \\ 0.01 & 0.01 & 1.00 & 1.00 \\ 0.01 & 0.01 & 1.00 & 1.00 \end{bmatrix}, \quad \tilde{\boldsymbol{u}} = \begin{bmatrix} 1.01 & 1.01 & 0.01 & 0.01 \\ 1.01 & 1.00 & 0.01 & 0.01 \\ 0.96 & 1.07 & 0.01 & 0.01 \\ 0.01 & 0.01 & 1.00 & 1.00 \\ 0.01 & 0.01 & 1.00 & 1.00 \\ 0.01 & 0.01 & 1.00 & 1.00 \end{bmatrix}, \quad \hat{\boldsymbol{u}} = \begin{bmatrix} 0.507 & 0.507 & 0.507 & 0.507 \\ 0.507 & 0.507 & 0.507 & 0.507 \\ 0.505 & 0.505 & 0.505 & 0.505 \\ 0.505 & 0.505 & 0.505 & 0.505 \\ 0.505 & 0.507 & 0.507 & 0.505 \\ 0.507 & 0.507 & 0.507 & 0.507 \end{bmatrix}.$$

We can see that $\bar{\boldsymbol{u}}, \tilde{\boldsymbol{u}}$ are close to $\underline{\boldsymbol{u}}$, whereas $\hat{\boldsymbol{u}}$ is totally different. Moreover, we compare all the local minimizers with $\bar{\boldsymbol{u}}$. See figure 4. Their regularization term values concentrate on $7, 8, \ldots, 21$, while $\mathcal{R}(\nabla\bar{u}) = 9$. Roughly speaking, the energy values of local minimizers decrease, when their regularization term values increase. Besides, only a small subset of the local minimizers is around the global minimizer $\bar{\boldsymbol{u}}$, but there are still 9408 local minimizers nearby $\bar{\boldsymbol{u}}$ within 0.1 error.

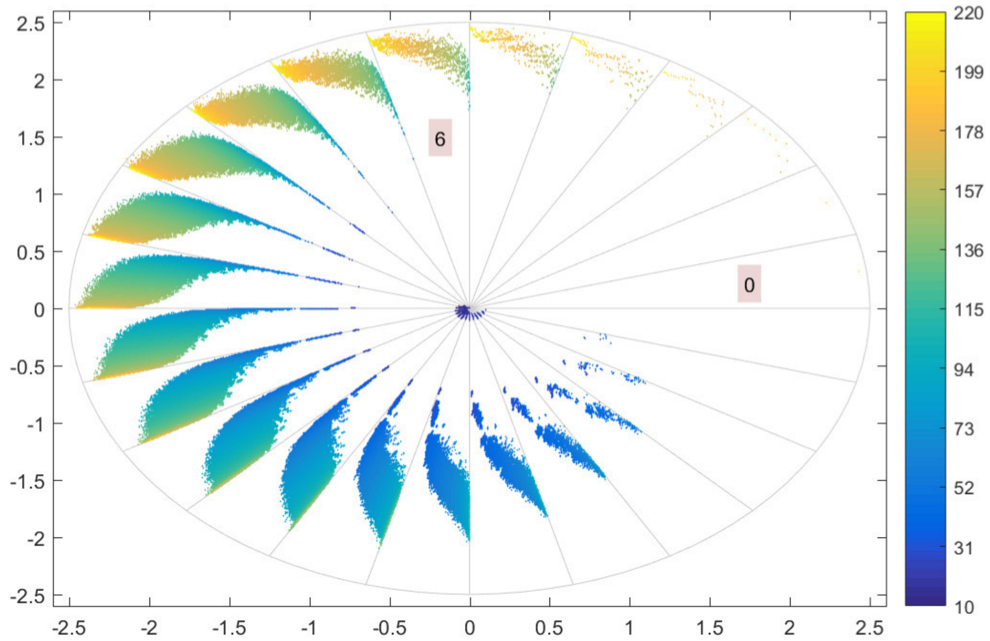


Figure 4. The distribution of all local minimizers in part(a). The global minimizer $\bar{\boldsymbol{u}}$ is centered at the origin. Each other point represents a local minimizer $\boldsymbol{u}$. The distance between the point and the origin equals $\|\boldsymbol{u}-\bar{\boldsymbol{u}}\|$; the color of this point illustrates its energy value $\mathcal{F}(u)$. The whole space is divided into 24 sections, two of which are labelled as 0, 6. Points in the same section have the same regularization term value. That is, if the point is in section labeled 0, then $\mathcal{R}(\nabla u)=0$. The regularization term value increases along the counterclockwise direction.

(b) In this part, we denote $\alpha=888, \underline{b}=0, \bar{b}=1$ and

$$\underline{\boldsymbol{u}}=\begin{bmatrix}1 & 0.5 & 1\\ 0.5 & 0 & 0.5\\ 1 & 0.5 & 1\end{bmatrix}, \qquad \boldsymbol{f}=\mathcal{A}\underline{\boldsymbol{u}}+\boldsymbol{\eta}=\begin{bmatrix}0.82 & 0.60 & 1.04\\ 0.41 & 0.20 & 0.58\\ 0.75 & 0.53 & 0.91\end{bmatrix}, \tag{68}$$

where $\boldsymbol{\eta}\in\mathbf{R}^{3\times3}$ is a random noise with $\|\boldsymbol{\eta}\|^2=0.055$. The noise here is much larger than previous example.

Then, we compute the local minimizers of (P), and there are only 140 local minimizers. We sort these local minimizers in terms of their energy values, and their energy values are shown in table 1. One can see that different local minimizers have different values.

Furthermore, we test other 1000 sets of random $\boldsymbol{f}$. In all tests, different local minimizers have different energy values.

(c) In this part, we also denote $\boldsymbol{f}$ as in (68). The parameter α is a variable here. Recall that the local minimizers of $(\mathrm{P}_\alpha), \alpha\in[0,+\infty)$ is independent of α. Then the local minimizer set of $(\mathrm{P}_\alpha), \alpha\in[0,+\infty)$ is the same as the one of (P_{888}), that is, the 140 local minimizers shown in part (a).

We partition these local minimizers into 9 subsets in terms of their regularization term values. In each subset, we find the ones having the minimal fidelity term value, and obtain $\{C_k\}_{\mathbb{I}_8^0}$, and

Table 2. The values of all local minimizers of (P) when $\alpha = 888$ and f is defined in (68). Each table cell contains the index of a local minimizer, followed by its energy value.

NO. value	NO. value	NO. value	NO. value	NO. value	NO. value	NO. value	NO. value	NO. value	NO. value
1 15.45	15 56.65	29 82.17	43 122.47	57 156.48	71 175.87	85 190.48	99 199.40	113 219.88	127 237.00
2 25.71	16 64.01	30 82.75	44 125.17	58 158.47	72 176.35	86 190.75	100 201.26	114 224.77	128 237.08
3 34.10	17 64.41	31 88.07	45 127.54	59 159.38	73 177.30	87 191.08	101 203.24	115 225.49	129 237.95
4 38.57	18 64.65	32 91.80	46 129.32	60 159.51	74 177.34	88 191.46	102 203.72	116 226.39	130 238.09
5 39.57	19 68.45	33 92.80	47 130.46	61 159.83	75 177.42	89 192.87	103 209.40	117 227.73	131 238.12
6 41.25	20 68.69	34 93.36	48 134.63	62 162.19	76 177.48	90 192.98	104 209.42	118 229.78	132 238.37
7 41.34	21 69.03	35 93.84	49 137.74	63 163.17	77 178.00	91 193.38	105 211.40	119 230.88	133 238.42
8 42.39	22 69.33	36 94.75	50 137.83	64 167.87	78 179.64	92 194.97	106 212.57	120 231.87	134 238.63
9 42.50	23 69.37	37 99.12	51 138.35	65 169.91	79 188.39	93 196.77	107 213.09	121 232.02	135 239.22
10 45.06	24 71.64	38 100.81	52 138.82	66 171.96	80 189.47	94 196.85	108 213.39	122 232.14	136 239.47
11 46.20	25 72.63	39 116.53	53 148.47	67 172.91	81 189.51	95 197.41	109 213.42	123 233.01	137 239.97
12 46.23	26 76.20	40 120.23	54 150.56	68 173.46	82 189.97	96 197.67	110 214.36	124 234.06	138 240.59
13 51.56	27 76.31	41 120.97	55 153.40	69 173.66	83 190.05	97 198.17	111 214.50	125 234.99	139 241.58
14 55.87	28 80.84	42 121.64	56 155.67	70 174.29	84 190.20	98 198.75	112 215.12	126 236.11	140 241.78

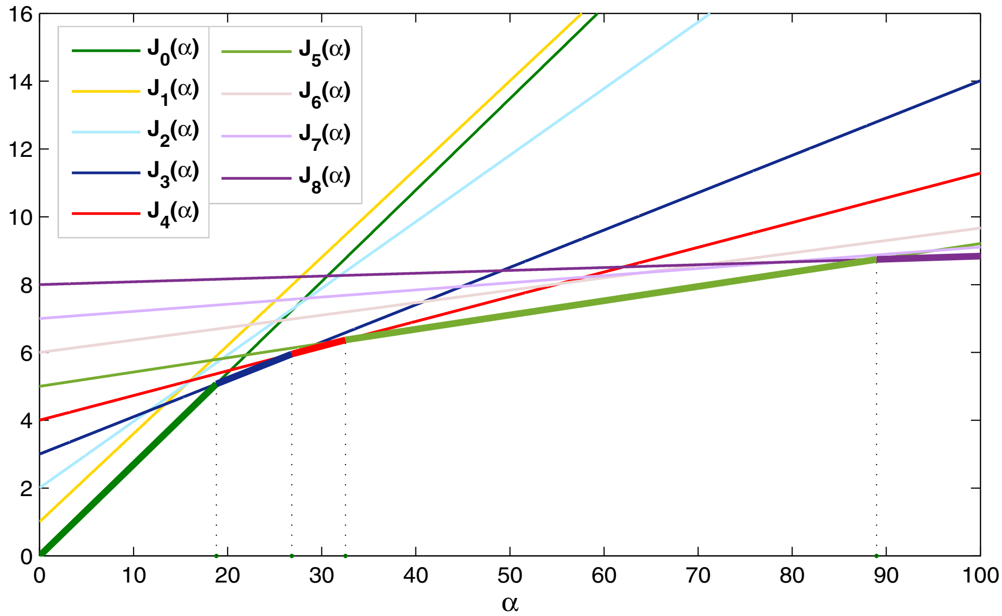


Figure 5. Plots of $\mathcal{J}_k(\alpha), k \in \mathbb{I}_8^0$ defined in (49). The segments marked as bold is the plot of $\mathcal{J}(\alpha)$.

$$\{c_k\}_{\mathbb{I}_8^0} = \{0.54,\ \ 0.52,\ \ 0.39,\ \ 0.22,\ \ 0.15,\ \ 0.08,\ \ 0.07,\ \ 0.04,\ \ 0.02\}.$$

Each $\mathcal{J}_k(\alpha), k \in \mathbb{I}_8^0$ is defined as (49), and their plots are shown in figure 5. Minimizing all of $\mathcal{J}_k(\alpha), k \in \mathbb{I}_8^0$ as (51), we have

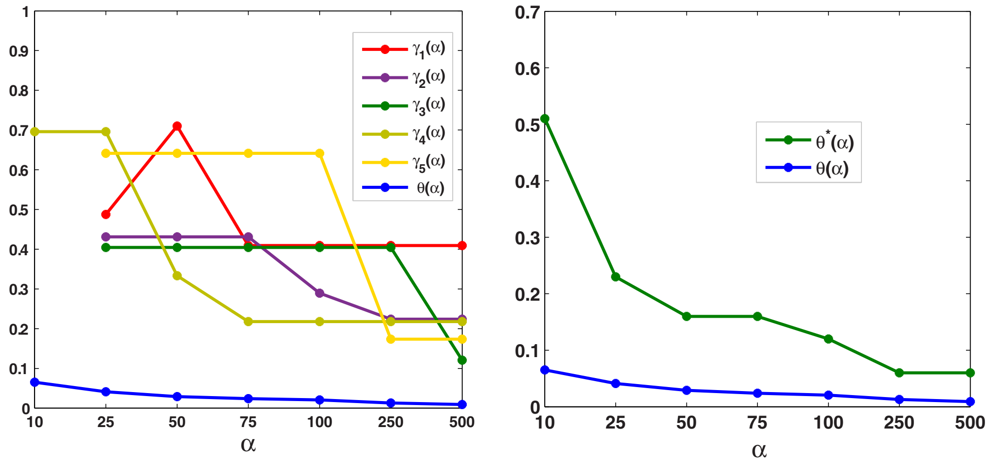
$$\mathcal{J}(\alpha) = \begin{cases} \mathcal{J}_0(\alpha) = 0.54\alpha, & \alpha \in [0, 18.81), \\ \mathcal{J}_3(\alpha) = 0.22\alpha + 3, & \alpha \in [18.81, 26.81), \\ \mathcal{J}_4(\alpha) = 0.15\alpha + 4, & \alpha \in [26.81, 32.53), \\ \mathcal{J}_5(\alpha) = 0.08\alpha + 5, & \alpha \in [32.53, 88.96), \\ \mathcal{J}_8(\alpha) = 0.02\alpha + 8, & \alpha \in [88.96, +\infty), \end{cases} \tag{69}$$

and

$$\mathcal{U}^g(\alpha) = \begin{cases} \{[\ 0.65 \ \ 0.65 \ \ 0.65 \ \ 0.65 \ \ 0.65 \ \ 0.65 \ \ 0.65 \ \ 0.65 \ \ 0.65\ ]^T\}, & \alpha \in [0, 18.81), \\ \{[\ 0.65 \ \ 0.65 \ \ 0.79 \ \ 0.65 \ \ 0 \ \ 0.79 \ \ 0.79 \ \ 0.79 \ \ 0.79\ ]^T\}, & \alpha \in (18.81, 26.81), \\ \{[\ 0.91 \ \ 0.32 \ \ 0.81 \ \ 0.64 \ \ 0 \ \ 0.81 \ \ 0.81 \ \ 0.81 \ \ 0.81\ ]^T\}, & \alpha \in (26.81, 32.53), \\ \{[\ 0.91 \ \ 0.34 \ \ 0.74 \ \ 0.59 \ \ 0 \ \ 0.74 \ \ 1 \ \ 0.74 \ \ 0.74\ ]^T\}, & \alpha \in (32.53, 88.96), \\ \{[\ 0.91 \ \ 0.33 \ \ 0.84 \ \ 0.61 \ \ 0 \ \ 0.50 \ \ 1 \ \ 0.56 \ \ 1\ ]^T\}, & \alpha \in (88.96, +\infty), \end{cases} \tag{70}$$

with $\mathcal{U}^g(\alpha)$ equal to the union of its left limit and right limit on each breakpoint. On the other hand, by (54), one has $T = 4$, and

$$\{k_t\}_{\mathbb{I}_4^0} = \{0,\ 3,\ 4,\ 5,\ 8\}, \qquad \{\alpha_t\}_{\mathbb{I}_4^0} = \{0,\ 18.81,\ 26.81,\ 32.53,\ 88.96\},$$



(a) (b)

Figure 6. Test for the lower bound.

which coincide (69) and (70). Furthermore, we set $\alpha = 10, 20, 30, 40, 90$ respectively, and compute the corresponding global minimizers, in line with the expression of $\mathcal{U}^g(\alpha)$ in (50). Thus, we can see that $\mathcal{J}(\alpha)$ is piecewise linear, and $\mathcal{U}^g(\alpha)$ is piecewise constant.

When $\alpha = 888 > 88.96$, we verify that the distance between $\underline{\boldsymbol{u}}$ in (68) and $\mathcal{U}^g(888)$ is bounded by the noise level:

$$0.5 \times 0.21 \times \|\underline{\boldsymbol{u}} - \mathcal{U}^g(888)\|^2 = 0.008 < 0.055 = \|\boldsymbol{\eta}\|^2.$$

(d) In this part, we need to verify the uniform lower bound of the ℓ_2 norm of nonzero gradients of the solutions to (P), defined in (30). Since this bound is related to parameter α, we rename it as $\theta(\alpha)$:

$$\theta(\alpha) := \min\{\frac{\sqrt{5}-1}{2\sqrt{\alpha N}\|A\|_2}, \frac{\sqrt{2}|\underline{b}-\bar{b}|}{2}\}.$$

To the end, we generate 1000 sets of $\boldsymbol{f}$ from $[0, 1]^{3\times 3}$ randomly, and choose several α: $10, 25, 50, 75, 100, 250, 500$.

Given any $\boldsymbol{f}_k, k = 1, \cdots, 1000$ and any chosen α, we compute the solution of (P), and then figure out the minimal nonzero gradients of the solution, denoted by $\gamma_k(\alpha)$:

$$\gamma_k(\alpha) := \min\{\|(\nabla\bar{u})[i]\| \ : \ i \in \sigma(\nabla\bar{u}), \bar{u} \text{ solves } (\mathrm{P})\}.$$

In figure 6(a), we show the plots of $\gamma_k(\alpha), k = 1, \cdots, 5$. We find that each $\gamma_k(\alpha)$ is different, and they are roughly decreasing when α is increasing. Owing to that the solution set of (P_α) for each f_k is piecewise constant in terms of α, each $\gamma_k(\alpha)$ is actually piecewise constant too. The line in blue represents the theoretical lower bound $\theta(\alpha)$. One can see that $\theta(\alpha)$ is less than all of the $\gamma_k(\alpha), k = 1, \cdots, 5$.

For each α, we compute the minimal nonzero gradients among all of the 1000 $\gamma_k(\alpha)$:

$$\theta^*(\alpha) := \min\{\gamma_k(\alpha), k = 1, \cdots 1000\},$$

and show it in figure 6(b). We can see that it is greater than the theoretical lower bound $\theta(\alpha)$. This fact verifies the reliability of our theoretical lower bound.

6. Conclusions and discussion

The ℓ_0 gradient regularized model has showed its superiority in image restoration. In this paper, we studied the local and global minimizers of this model with box constraints. Compared to previous analyses [21, 22, 36] for sparse signal recovery model, the difficulty caused by the gradient operator in model (3) is resolved by the introduction of the block image, which is defined in section 3.2. Besides, the bounded-variable least square problems also contribute greatly to the whole discussion. The numerical examples were provided to demonstrate our theoretical analyses. Our results give a comprehensive understanding of this image restoration model, and are conductive to the design of new efficient algorithms to solve it.

Furthermore, the results in this paper can be extended in several ways.

- The box constraints in model (3) can be replaced by the more general box constraints: $\underline{\mathbf{b}} \leqslant u \leqslant \bar{\mathbf{b}}$, where $\underline{\mathbf{b}}, \bar{\mathbf{b}} \in \mathbf{R}^N$.
- The regularization term $\mathcal{R}(\nabla u)$ is isotropic in model (3). It can be replaced by anisotropic ones, and the proofs need some modifications.
- When the ℓ_2 norm in the fidelity term is replaced by any other continuous functions, most results, such as theorems 3.5 and 4.6, still hold.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (Grants 11301289 and 11531013), and the Recruitment Programm of Global Young.

ORCID iDs

Xue Feng https://orcid.org/0000-0003-0728-7502

References

[1] Bao C, Dong B, Hou L, Shen Z, Zhang X and Zhang X 2016 Image restoration by minimizing zero norm of wavelet frame coefficients *Inverse Problems* **32** 115004
[2] Baraniuk R G 2007 Compressive sensing (lecture notes) *IEEE Signal Process. Mag.* **24** 118–21
[3] Beck A and Teboulle M 2009 Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems *IEEE Trans. Image Process.* **18** 2419–34
[4] Besag J 1986 On the statistical analysis of dirty pictures *J. R. Stat. Soc.* **48** 259–302
[5] Bian W and Chen X 2015 Linearly constrained non-lipschitz optimization for image restoration *SIAM J. Imaging Sci.* **8** 2294–322
[6] Chan R H, Tao M and Yuan X 2013 Constrained total variation deblurring models and fast algorithms based on alternating direction method of multipliers *SIAM J. Imaging Sci.* **6** 680–97
[7] Chen X, Ng M K and Zhang C 2012 Non-lipschitz ℓ_p-regularization and box constrained model for image restoration *IEEE Trans. Image Process.* **21** 4709–21
[8] Chen X and Zhou W 2010 Smoothing nonlinear conjugate gradient method for image restoration using nonsmooth nonconvex minimization *SIAM J. Imaging Sci.* **3** 765–90
[9] Cheng X, Zeng M and Liu X 2014 Feature-preserving filtering with ℓ_0 gradient minimization *Comput. Graph.* **38** 150–7

[10] Chouzenoux E, Jezierska A, Pesquet J C and Talbot H 2013 A majorize-minimize subspace approach for $\ell_2 - \ell_0$ image regularization *SIAM J. Imaging Sci.* **6** 563–91
[11] Dong B and Zhang Y 2013 An efficient algorithm for ℓ_0 minimization in wavelet frame based image restoration *J. Sci. Comput.* **54** 350–68
[12] Geman S and Geman D 1984 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images *IEEE Trans. Pattern Anal. Mach. Intell.* **6** 721–41
[13] Golub G H 1965 Numerical methods for solving least squares problems *Numer. Math.* **7** 206–16
[14] Hiriart U and Jean B 2013 When only global optimization matters *J. Glob. Optim.* **56** 761–3
[15] Ji S, Xue Y and Carin L 2008 Bayesian compressive sensing *IEEE Trans. Signal Process.* **56** 2346–56
[16] Lawson C L and Hanson R J 1995 *Solving Least Squares Problems* (Philadelphia, PA: SIAM)
[17] Leclerc Y G 1989 Constructing simple stable descriptions for image partitioning *Int. J. Comput. Vis.* **3** 73–102
[18] Meyer C 2000 *Matrix Analysis and Applied Linear Algebra* (Philadelphia, PA: SIAM)
[19] Miller A 2002 *Subset Selection in Regression* (Boca Raton, FL: CRC Press)
[20] Nikolova M 2006 Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares *SIAM J. Multiscale Model. Simul.* **4** 960–91
[21] Nikolova M 2013 Description of the minimizers of least squares regularized with ℓ_0-norm. Uniqueness of the global minimizer *SIAM J. Imaging Sci.* **6** 904–37
[22] Nikolova M 2016 Relationship between the optimal solutions of least squares regularized with ℓ_0 -norm and constrained by k-sparsity *Appl. Comput. Harmon. Anal.* **41** 237–65
[23] Nikolova M, Ng M K and Tam C P 2013 On ℓ_1 data fitting and concave regularization for image recovery *SIAM J. Sci. Comput.* **35** A397–430
[24] Nikolova M, Ng M K, Zhang S and Ching W K 2008 Efficient reconstruction of piecewise constant images using nonsmooth nonconvex minimization *SIAM J. Imaging Sci.* **1** 2–25
[25] Pei S C, Shen C T and Lee T Y 2012 Visual enhancement using constrained ℓ_0 gradient image decomposition for low backlight displays *IEEE Signal Process. Lett.* **19** 813–6
[26] Robini M C, Lachal A and Magnin I E 2007 A stochastic continuation approach to piecewise constant reconstruction *IEEE Trans. Image Process.* **16** 2576–89
[27] Rockafellar R T and Wets R J B 1998 *Variational analysis in Sobolev and BV Spaces* (*MPS-SIAM Series on Optimization* vol 30) pp 324–6
[28] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithms *Physica* D **60** 259–68
[29] Shen L, Xu Y and Zeng X 2016 Wavelet inpainting with the ℓ_0 sparse regularization *Appl. Comput. Harmon. Anal.* **41** 26–53
[30] Shoshichi K and Katsumi N 1996 *Foundations of Differential Geometry, Volume* 2 (New York: Wiley Interscience)
[31] Stan Z L 1995 *Markov Random Field Modeling in Computer Vision* (Berlin: Springer)
[32] Storath M, Weinmann A and Demaret L 2014 Jump-sparse and sparse recovery using potts functionals *IEEE Trans. Signal Process.* **62** 3654–66
[33] Xu L, Lu C, Xu Y and Jia J 2011 Image smoothing via ℓ_0 gradient minimization *ACM Trans. Graph.* **30** 174:1–12
[34] Xu L, Zheng S and Jia J 2013 Unnatural ℓ_0 sparse representation for natural image deblurring *Comput. Vis. Pattern Recognit.* **9** 1107–14
[35] Zeng C and Wu C 2018 On the edge recovery property of noncovex nonsmooth regularization in image restoration *SIAM J. Numer. Anal.* **56** 1168–82
[36] Zhang N and Li Q 2016 On optimal solutions of the constrained ℓ_0 regularization and its penalty problem *Inverse Problems* **33** 025010
[37] Zhang Y, Dong B and Lu Z 2011 ℓ_0 minimization for wavelet frame based image restoration *Math. Comput.* **82** 995–1015