

Half-Quadratic Algorithm for ℓ_p - ℓ_q Problems with Applications to TV- ℓ_1 Image Restoration and Compressive Sensing

Raymond H. Chan¹(✉) and Hai-Xia Liang²(✉)

¹ Department of Mathematics, The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, People's Republic of China

`rchan@math.cuhk.edu.hk`

² Mathematics and Physics Teaching Centre, Xi'an Jiaotong-Liverpool University, No. 111 Ren'ai Road, Suzhou Industrial Park, Suzhou, Jiangsu Province, People's Republic of China

`haixia.liang@xjtlu.edu.cn`

Abstract. In this paper, we consider the ℓ_p - ℓ_q minimization problem with $0 < p, q \leq 2$. The problem has been studied extensively in image restoration and compressive sensing. In the paper, we first extend the half-quadratic algorithm from ℓ_1 -norm to ℓ_p -norm with $0 < p < 2$. Based on this, we develop a half-quadratic algorithm to solve the ℓ_p - ℓ_q problem. We prove that our algorithm is indeed a majorize-minimize approach. From that we derive some convergence results of our algorithm, e.g. the objective function value is monotonically decreasing and convergent. We apply the proposed approach to TV- ℓ_1 image restoration and compressive sensing in magnetic resonance (MR) imaging applications. The numerical results show that our algorithm is fast and efficient in restoring blurred images that are corrupted by impulse noise, and also in reconstructing MR images from very few k -space data.

Keywords: ℓ_p regularization · Half-quadratic · Majorize-minimize algorithm · Impulse noise · Compressive sensing · Magnetic resonance imaging

1 Introduction

In this paper, we consider the following ℓ_p - ℓ_q minimization problem

$$\min_{\mathbf{u}} \left\{ \frac{\lambda}{p} \|\Psi \mathbf{u}\|_p^p + \frac{1}{q} \|\mathbf{A} \mathbf{u} - \mathbf{f}\|_q^q \right\} \quad (1)$$

where $0 < p, q \leq 2$ and $\mathbf{u} \in \mathbb{R}^n$ is an image represented by a vector by concatenating the columns. Here, Ψ can be a sparsifying operator such as a wavelet transform or a regularization operator such as the discrete gradient operator;

The research was supported in part by HKRGC Grant CUHK 400510 and CUHK DAG 2060408.

and A can be a sampling operator or a blurring operator. Problem (1) has been studied extensively in image processing and compressive sensing. For example, if $p = 1$, $q = 2$, and Ψ is the discrete gradient operator, then (1) is the TV- ℓ_2 minimization problem. It has been successfully applied to image restoration [20, 37, 39] because of its good property in preserving edges. TV- ℓ_1 model (i.e. $q = 1$) has also been successfully applied to applications such as impulse noise removal [26, 41], image cartoon-texture decomposition [45], feature selection [45], multiscale decomposition [46], and computer vision [10].

When A is a sampling operator, model (1) with $0 \leq p \leq 1$ and $q = 2$ has received a lot of attention lately because of the introduction of compressive sensing techniques ($\|\mathbf{u}\|_0$ is defined to be the number of nonzero entries in \mathbf{u}). The techniques allow high resolution images and signals to be reconstructed from a small amount of data [6, 7, 17]. There, the linear constrained minimization problems are considered. Unfortunately, as $p = 0$, the constrained minimization problem is NP-hard [1]. For this reason, different approaches are used to approximate the ℓ_0 -norm [3, 14, 23, 29, 30, 34], or alternatively one solves the ℓ_1 -norm [5, 20, 32, 47] or the non-convex ℓ_p -norm [12–14] problem with $0 < p < 1$. The application of compressive sensing with $p = 1$ and $q = 2$ to magnetic resonance (MR) image reconstruction can be found in [7, 27]. There it was shown that perfect reconstruction of the Shepp-Logan phantom is possible from 22 radial lines or 9% of the k -space data. For real images which are less sparse than the synthetic phantoms, one can obtain improved results by having both a wavelet transform and a discrete gradient in the objective function. However, the ℓ_1 -norm regularized model can not get good results from fewer k -space data. See [27].

Problem (1) with $0 < p < 1$ is a non-convex optimization problem. Theoretical work [15, 38] has justified the non-convex approach as it guarantees perfect reconstruction under a less restrictive condition than that would be needed by ℓ_1 minimization. There are quite a few algorithms for solving the non-convex problem, see [4, 12–14, 35]. The numerical results in [12, 13] show that the perfect MR image can be recovered from 10 radial lines (i.e. 3.8% of the k -space data) for some $0 < p < 1$. In [12], a fast algorithm based on the p -shrinkage reduces the number of lines further to 9. For more details on p -shrinkage, one may consult [43] where the 1/2-thresholding algorithm was studied. In [42], the author analyzed the effectiveness of problem (1) in recovering sparse signals. The results showed that if $p \in [1/2, 1)$, then the smaller the p is, the sparser the ℓ_p -norm regularization solution will be; and if $p \in (0, 1/2]$, there are no significant differences in the sparsity of the solution.

In this paper, we propose a half-quadratic algorithm (HQA) to solve (1) for $0 < p, q \leq 2$. We prove that our algorithm is indeed a majorize-minimize algorithm [11, 24, 25] for solving (1) and from that some convergence results can be obtained immediately. For example, we show that the objective function value is monotonically decreasing and convergent. We also give the convergence rate of the method for $1 \leq p, q \leq 2$. We test our algorithms on two applications: (i) TV- ℓ_1 minimization problem, and (ii) non-convex ℓ_p - ℓ_2 compressive sensing. Problem

(i) is for restoring blurred images that are corrupted by impulse noise, and our algorithm can reach high SNR value in less CPU time than the augmented-Lagrangian method (ALM) in [41] and the fast total variation deconvolution method (FTVDM) in [44]. Problem (ii) is for reconstructing MR images from few k -space data, and our algorithm can get better results with less computational cost than the p -shrinkage algorithm in [12].

The outline of the paper is as follows: In Sect. 2, we first derive our HQA for model (1) and then adapt it to solve the TV- ℓ_1 minimization problem and non-convex ℓ_p - ℓ_2 compressive sensing problem. In Sect. 3, we prove that our HQA is indeed a majorize-minimize algorithm, and hence we derive some convergence results for the algorithm. Comparison with the ALM, FTVDM and the p -shrinkage algorithm [12] are given in Sect. 4. Finally Sect. 5 is on some concluding remarks.

2 The Half-Quadratic Approach for ℓ_p - ℓ_q Minimization Problem

The half-quadratic regularization approach has been used in image processing [11, 19]. In [31], the authors showed the equivalence of the HQ minimization and the gradient linearization iterations. The HQ technique is based on the fact that, if $0 \neq t \in \mathbb{R}$, then

$$|t| = \min_{v>0} \left\{ vt^2 + \frac{1}{4v} \right\} \quad (2)$$

and the minimum value is reached at $v = \frac{1}{2|t|}$. Note that the function $vt^2 + 1/(4v)$ is quadratic in t but not in v and hence the name *half-quadratic*. In this section, we first study the general form of (2) for $0 < p < 2$. Then we derive our HQA to solve (1) and adapt it to the TV- ℓ_1 minimization problem and compressive sensing.

2.1 The Half-Quadratic Algorithm

The following lemma shows us the corresponding formula of (2) for $|\cdot|^p$ with $0 < p < 2$.

Lemma 1. *For any $0 < p < 2$, if $0 \neq t \in \mathbb{R}$, then there exist positive constants α and ξ such that*

$$|t|^p = \min_{v>0} \left\{ vt^2 + \frac{1}{\xi v^\alpha} \right\}. \quad (3)$$

Proof: Let us first assume that $\alpha, \xi > 0$ and define

$$f(v, t) := vt^2 + \frac{1}{\xi v^\alpha}. \quad (4)$$

Then $f(v, t)$ is convex with respect to $v > 0$ for any fixed $t \neq 0$. In addition, $f(v, t) \rightarrow \infty$ as either $v \rightarrow 0$ or $v \rightarrow \infty$. The minimizer of $f(v, t)$ with respect to v is therefore given by solving $f'_v(v, t) = 0$ and is equal to

$$v^* = \left(\frac{\alpha}{\xi t^2} \right)^{\frac{1}{1+\alpha}}. \quad (5)$$

Substituting (5) into (4), the minimum value is

$$f(v^*, t) = \left[\left(\frac{\alpha}{\xi} \right)^{\frac{1}{1+\alpha}} + \frac{1}{\xi} \left(\frac{\xi}{\alpha} \right)^{\frac{\alpha}{1+\alpha}} \right] t^{\frac{2\alpha}{1+\alpha}}.$$

Since we want the minimum value to be $|t|^p$ for any $t \neq 0$, we set

$$\begin{cases} 2\alpha/(1+\alpha) = p, \\ \left(\frac{\alpha}{\xi} \right)^{\frac{1}{1+\alpha}} + \frac{1}{\xi} \left(\frac{\xi}{\alpha} \right)^{\frac{\alpha}{1+\alpha}} = 1. \end{cases}$$

Solving the system for α and ξ , we have

$$\alpha = \frac{p}{2-p} \quad \text{and} \quad \xi = \frac{2^{\frac{2}{2-p}}}{(2-p) \cdot p^{\frac{p}{2-p}}}. \quad (6)$$

Clearly both α and ξ are positive for any fixed $0 < p < 2$. \square

Remarks

(a) As an example, for $p = 1/2$, we have $\xi = 2^{8/3}/3$ and $\alpha = 1/3$. For $p = 1$, we have $\alpha = 1$ and $\xi = 4$, and hence (3) reduces to (2). Note that we would like to have $\alpha > 0$ so that the functional $f(v, t)$ is convex with respect to v for any fixed $t \neq 0$. By (5) and (6), we have

$$v^* = \frac{p}{2} |t|^{p-2}. \quad (7)$$

(b) From the above lemma, we know that for fixed $\alpha, \xi > 0$ of (6) and any $t \neq 0$, the minimum of (3) is reached at the stationary point of $f(v, t)$ w.r.t v , which is an interior point of the open, convex, feasible set \mathbb{R}^+ .

Next we apply (3) to solve (1). To simplify the discussions, we first consider the case where $0 < p, q < 2$, and leave the case where p and/or $q = 2$ later.

Case 1: $0 < p, q < 2$. Notice that Lemma 1 holds only for $t \neq 0$ as negative power of $|t|$ appears in v^* , see (7). Hence in order to apply (3), we need to smooth (1) first. In the following, we denote $|\rho|_\epsilon := \sqrt{\rho^2 + \epsilon}$ for any $\rho \in \mathbb{R}$ and $\epsilon > 0$. The smoothed ℓ_p - ℓ_q problem of (1) is

$$\min_{\mathbf{u}} \left\{ \frac{\lambda}{p} \|\Psi \mathbf{u}\|_{p,\beta}^p + \frac{1}{q} \|\mathbf{A} \mathbf{u} - \mathbf{f}\|_{q,\gamma}^q \right\} =: \min_{\mathbf{u}} \{ \Phi_{\beta,\gamma}(\mathbf{u}) \} \quad (8)$$

where $\|\Psi \mathbf{u}\|_{p,\beta}^p := \sum_{i=1}^n |\Psi_i \mathbf{u}|_\beta^p$ and $\|A \mathbf{u} - \mathbf{f}\|_{q,\gamma}^q := \sum_{i=1}^n |A_i \mathbf{u} - f_i|_\gamma^q$, with β and γ being small positive numbers, and Ψ_i and A_i are the i th rows of Ψ and A respectively. Applying (3) to (8), problem (8) becomes

$$\begin{aligned} & \min_{\mathbf{u}} \left\{ \sum_{i=1}^n \left[\frac{\lambda}{p} \min_{v_i > 0} \left(v_i |\Psi_i \mathbf{u}|_\beta^2 + \frac{1}{\xi_p v_i^{\alpha_p}} \right) + \frac{1}{q} \min_{w_i > 0} \left(w_i |A_i \mathbf{u} - f_i|_\gamma^2 + \frac{1}{\xi_q w_i^{\alpha_q}} \right) \right] \right\} \\ &= \min_{\mathbf{u}, \mathbf{v} > \mathbf{0}, \mathbf{w} > \mathbf{0}} \left\{ \sum_{i=1}^n \left[\frac{\lambda}{p} \left(v_i |\Psi_i \mathbf{u}|_\beta^2 + \frac{1}{\xi_p v_i^{\alpha_p}} \right) + \frac{1}{q} \left(w_i |A_i \mathbf{u} - f_i|_\gamma^2 + \frac{1}{\xi_q w_i^{\alpha_q}} \right) \right] \right\} \\ &=: \min_{\mathbf{u}, \mathbf{v} > \mathbf{0}, \mathbf{w} > \mathbf{0}} \{ \mathcal{L}(\mathbf{u}, \mathbf{v}, \mathbf{w}) \}, \end{aligned} \quad (9)$$

where $\mathbf{v}, \mathbf{w} > \mathbf{0}$ mean that all the components of \mathbf{v}, \mathbf{w} are greater than 0. Here ξ_i and α_i , $i = p, q$ are scalars given by (6).

To solve (9), we apply the alternating minimization procedure, namely,

$$\mathbf{v}^{k+1} = \arg \min_{\mathbf{v} > \mathbf{0}} \mathcal{L}(\mathbf{u}^k, \mathbf{v}, \mathbf{w}^k), \quad (10)$$

$$\mathbf{w}^{k+1} = \arg \min_{\mathbf{w} > \mathbf{0}} \mathcal{L}(\mathbf{u}^k, \mathbf{v}^{k+1}, \mathbf{w}), \quad (11)$$

$$\mathbf{u}^{k+1} = \arg \min_{\mathbf{u}} \mathcal{L}(\mathbf{u}, \mathbf{v}^{k+1}, \mathbf{w}^{k+1}). \quad (12)$$

By (7), we know that (10) and (11) have explicit component minimizers

$$v_i^{k+1} = \frac{p}{2} |\Psi_i \mathbf{u}^k|_\beta^{p-2} \quad \text{and} \quad w_i^{k+1} = \frac{q}{2} |A_i \mathbf{u}^k - f_i|_\gamma^{q-2}. \quad (13)$$

Note that $\mathcal{L}(\mathbf{u}, \mathbf{v}^{k+1}, \mathbf{w}^{k+1})$ is continuously differentiable in \mathbf{u} . Hence \mathbf{u}^{k+1} in (12) is the solution of

$$0 = \nabla_{\mathbf{u}} \mathcal{L}(\mathbf{u}, \mathbf{v}^{k+1}, \mathbf{w}^{k+1}) = \lambda \Psi^\top D_\beta(\mathbf{u}^k) \Psi \mathbf{u} + A^\top D_\gamma(\mathbf{u}^k) (A \mathbf{u} - \mathbf{f}), \quad (14)$$

where $D_\beta(\mathbf{u}^k)$ and $D_\gamma(\mathbf{u}^k)$ are diagonal matrices with their i th diagonal entries being

$$\frac{2}{p} v_i^{k+1} = |\Psi_i \mathbf{u}^k|_\beta^{p-2} \quad \text{and} \quad \frac{2}{q} w_i^{k+1} = |A_i \mathbf{u}^k - f_i|_\gamma^{q-2} \quad (15)$$

respectively. Equation (14) provides us an iterative scheme for finding the minimizer of (8).

Case 2: p and/or $q = 2$. In that case, the corresponding term in (1) is quadratic and differentiable. So there is no need to apply the half-quadratic technique (3) to the term. However, one can easily check by differentiation of (1) that (14) and (15) are still valid. More precisely, if $p = 2$, then differentiation of the first term in (1) gives $\lambda \Psi^\top \Psi \mathbf{u}$. But by (15), $D_\beta(\mathbf{u}^k) \equiv I$, the identity matrix and hence the first term of (14) also reduces to $\lambda \Psi^\top \Psi \mathbf{u}$. Similarly, if $q = 2$, then $D_\gamma(\mathbf{u}^k) \equiv I$ and (14) still holds. In particular, if $p = q = 2$, then (14) reduces to the least-squares problem as expected, and the minimizer \mathbf{u} can be obtained in one iteration. In the following discussions, we will exclude this trivial case $p = q = 2$.

Thus combining Case 1 and Case 2, we see that (14) holds for $0 < p, q \leq 2$. We summarize our half-quadratic algorithm (HQA) for the smoothed ℓ_p - ℓ_q problem (8) below:

- (1) Initialize \mathbf{u}^0 ;
- (2) For $l = 0, 1, \dots$ until convergence, find \mathbf{u}^{k+1} by solving (cf (14))

$$\left(\lambda \Psi^\top D_\beta(\mathbf{u}^k) \Psi + A^\top D_\gamma(\mathbf{u}^k) A \right) \mathbf{u} = A^\top D_\gamma(\mathbf{u}^k) \mathbf{f}, \quad (16)$$

where $D_\beta(\cdot)$ and $D_\gamma(\cdot)$ are diagonal matrices given in (15).

To find the solution to the ℓ_p - ℓ_q problem (1), we can use a continuation method and apply HQA to a sequence of $\{\beta^l, \gamma^l\}$ going to zero. We will discuss the implementation in more details in the section on numerical tests, see **Algorithm 1** in Sect. 4.

2.2 Half-Quadratic Algorithm for TV- ℓ_1 and Compressive Sensing

Let us consider HQA (16) for two specific examples: TV- ℓ_1 image restoration and compressive sensing. The TV- ℓ_1 minimization problem is of the form:

$$\min_{\mathbf{u}} \left\{ \lambda \|\nabla \mathbf{u}\|_1 + \|A\mathbf{u} - \mathbf{f}\|_1 \right\}, \quad (17)$$

where $\|\nabla \mathbf{u}\|_1 := \sum_{i=1}^n \sqrt{[(G_1)_i \mathbf{u}]^2 + [(G_2)_i \mathbf{u}]^2}$ with $(G_j)_i$ representing the i th row of the finite difference operator in the x_j -coordinate. The smoothed version of (17) is

$$\min_{\mathbf{u}} \left\{ \sum_{i=1}^n [\lambda |\nabla u_i|_\beta + |A_i \mathbf{u} - f_i|_\gamma] \right\}, \quad (18)$$

where $|\nabla u_i|_\beta := \sqrt{[(G_1)_i \mathbf{u}]^2 + [(G_2)_i \mathbf{u}]^2 + \beta}$ and $\beta, \gamma \rightarrow 0$. Letting $p = q = 1$, $\Psi = G$ in (15) and (16), we see that (16) should be replaced by:

$$\left\{ \lambda \sum_{j=1}^2 [G_j^\top D_\beta(\mathbf{u}^k) G_j] + A^\top D_\gamma(\mathbf{u}^k) A \right\} \mathbf{u} = A^\top D_\gamma(\mathbf{u}^k) \mathbf{f}, \quad (19)$$

where $D_\beta(\mathbf{u}^k)$ and $D_\gamma(\mathbf{u}^k)$ are diagonal matrices with their i th diagonal entries being $|\nabla u_i^k|_\beta^{-1}$ and $|A_i \mathbf{u}^k - f_i|_\gamma^{-1}$ respectively.

Next we consider HQA for MR image reconstruction problem. In [27], a regularization term combining a discrete gradient ∇ and an orthogonal wavelet W [8] is considered for $0 < p \leq 1$:

$$\min_{\mathbf{u}} \left\{ \|\nabla \mathbf{u}\|_p^p + \delta \|W\mathbf{u}\|_p^p : R F \mathbf{u} = \mathbf{f} \right\}. \quad (20)$$

Here R is a selection matrix (a diagonal matrix) and F is the Fourier transform. As mentioned in [12], it is sufficient to use $\delta = 1$ in (20); and for gradient-sparse

images, we can simply take $\delta = 0$. Problem (20) is related to the minimization problem

$$\min_{\mathbf{u}} \left\{ \frac{\lambda}{p} \left(\|\nabla \mathbf{u}\|_p^p + \delta \|W\mathbf{u}\|_p^p \right) + \frac{1}{2} \|RF\mathbf{u} - \mathbf{f}\|_2^2 \right\}, \quad (21)$$

where λ is the Lagrange multiplier. As before, since the data fitting term is quadratic, we keep it intact, and we smooth only the p -norm terms. Hence we have the following smoothed problem:

$$\min_{\mathbf{u}} \left\{ \frac{\lambda}{p} \sum_{i=1}^n \left(|\nabla u_i|_\beta^p + \delta |W_i \mathbf{u}|_\gamma^p \right) + \frac{1}{2} \|RF\mathbf{u} - \mathbf{f}\|_2^2 \right\},$$

with $\beta, \gamma \rightarrow 0$. Correspondingly, Eq. (16) should be replaced by:

$$\left[\lambda \left(\sum_{j=1}^2 [G_j^\top D_\beta(\mathbf{u}^k) G_j] + \delta W^\top D_\gamma(\mathbf{u}^k) W \right) + F^* RF \right] \mathbf{u} = F^* R\mathbf{f}, \quad (22)$$

where F^* is inverse Fourier transform, $D_\beta(\mathbf{u}^k)$ and $D_\gamma(\mathbf{u}^k)$ are diagonal matrices with their i th diagonal entries being $|\nabla u_i^k|_\beta^{p-2}$ and $|W_i \mathbf{u}^k|_\gamma^{p-2}$ respectively.

3 Convergence Analysis

In this section, we analyze the convergence of the HQA (16) based on the convergence property of the majorize-minimize algorithm (MMA) in [11, 24, 25, 40] for fixed β, γ . We first show that $\Phi_{\beta, \gamma}(\mathbf{u}^k)$ is monotonically decreasing and convergent for $0 < p, q \leq 2$. Then we show that \mathbf{u}^k is convergent and linearly convergent for $1 \leq p, q \leq 2$.

3.1 Convergence of $\Phi_{\beta, \gamma}(\mathbf{u}^k)$ for $0 < p, q \leq 2$

The MM optimization technique [11, 24, 25] is to solve a minimization problem $\min_{\mathbf{u}} \Phi_{\beta, \gamma}(\mathbf{u})$ by

$$\mathbf{u}^{k+1} = \arg \min_{\mathbf{u}} \{Q(\mathbf{u}, \mathbf{u}^k)\}, \quad (23)$$

where $Q(\mathbf{u}, \mathbf{u}^k)$, called a *tangent majorant function* of $\Phi_{\beta, \gamma}(\mathbf{u})$ at \mathbf{u}^k , must satisfy

$$Q(\mathbf{u}, \mathbf{u}^k) \geq \Phi_{\beta, \gamma}(\mathbf{u}), \quad \forall \mathbf{u} \in \mathbb{R}^n, \quad (24)$$

$$Q(\mathbf{u}, \mathbf{u}^k) = \Phi_{\beta, \gamma}(\mathbf{u}), \quad \text{at } \mathbf{u} = \mathbf{u}^k, \quad (25)$$

$$\nabla_1 Q(\mathbf{u}, \mathbf{u}^k) = \nabla \Phi_{\beta, \gamma}(\mathbf{u}), \quad \text{at } \mathbf{u} = \mathbf{u}^k. \quad (26)$$

Here, $\nabla_1 Q(\mathbf{u}, \mathbf{u}^k)$ denotes the partial derivative with respect to the first vector variable. Convergence analysis of the MMA can be found in [11, 24, 25].

For the smoothed ℓ_p - ℓ_q problem (8), if we define $Q(\mathbf{u}, \mathbf{u}^k) := \mathcal{L}(\mathbf{u}, \mathbf{v}^{k+1}, \mathbf{w}^{k+1})$, then our HQA (12) can be written as

$$\mathbf{u}^{k+1} = \arg \min_{\mathbf{u}} \mathcal{L}(\mathbf{u}, \mathbf{v}^{k+1}, \mathbf{w}^{k+1}) = \arg \min_{\mathbf{u}} \{Q(\mathbf{u}, \mathbf{u}^k)\}, \quad (27)$$

which is of the same form as (23). For $0 < p, q < 2$, substituting (13) into (27), we obtain the explicit form of $Q(\mathbf{u}, \mathbf{u}^k)$:

$$Q(\mathbf{u}, \mathbf{u}^k) = \sum_{i=1}^n \left[\frac{\lambda}{p} \left(\frac{p}{2} |\Psi_i \mathbf{u}^k|^{p-2} |\Psi_i \mathbf{u}|_\beta^2 + \frac{2-p}{2} |\Psi_i \mathbf{u}^k|^p \right) + \frac{1}{q} \left(\frac{q}{2} |A_i \mathbf{u}^k - f_i|_\gamma^{q-2} |A_i \mathbf{u} - f_i|_\gamma^2 + \frac{2-q}{2} |A_i \mathbf{u}^k - f_i|_\gamma^q \right) \right]. \quad (28)$$

We recall that when p or q is equal to 2, there is no need to smooth the corresponding term as it is already differentiable. Hence if we use the convention that $\beta = 0$ (or respectively $\gamma = 0$) whenever $p = 2$ (or respectively $q = 2$), then (28) holds for all $0 < p, q \leq 2$.

Lemma 2. *Let $0 < p, q \leq 2$. For any fixed $\beta, \gamma > 0$, the HQA for the smoothed ℓ_p - ℓ_q problem (8) is the same as an MMA with tangent majorant function $Q(\mathbf{u}, \mathbf{u}^k)$ defined by (28).*

Proof: Since our HQA (12) is rewritten as (27) with $Q(\mathbf{u}, \mathbf{u}^k)$ given by (28), we only need to prove that (24)–(26) holds for such Q . Substituting $\mathbf{u} = \mathbf{u}^k$ in (28) and using the definition of $\Phi_{\beta, \gamma}(\mathbf{u})$ in (8), we see that $Q(\mathbf{u}^k, \mathbf{u}^k) = \Phi_{\beta, \gamma}(\mathbf{u}^k)$, which is (25). To prove that Q satisfies (24), we use the Young inequality, which states that $(x^a/a + y^b/b) \geq xy$ for all $x, y \geq 0$, $a, b \geq 1$ and $1/a + 1/b = 1$. Let us consider the case where $0 < p < 2$ first, and set

$$x = |\Psi_i \mathbf{u}^k|_\beta^{\frac{(p-2)p}{2}} |\Psi_i \mathbf{u}|_\beta^p, \quad y = |\Psi_i \mathbf{u}^k|_\beta^{\frac{(2-p)p}{2}}, \quad a = \frac{2}{p}, \quad b = \frac{2}{2-p}.$$

Then Young's inequality implies that

$$\frac{p}{2} |\Psi_i \mathbf{u}^k|_\beta^{p-2} |\Psi_i \mathbf{u}|_\beta^2 + \frac{2-p}{2} |\Psi_i \mathbf{u}^k|^p \geq |\Psi_i \mathbf{u}|_\beta^p.$$

Clearly, the inequality becomes a trivial equality when $p = 2$. Similarly, we can show that

$$\frac{q}{2} |A_i \mathbf{u}^k - f_i|_\gamma^{q-2} |A_i \mathbf{u} - f_i|_\gamma^2 + \frac{2-q}{2} |A_i \mathbf{u}^k - f_i|_\gamma^q \geq |A_i \mathbf{u} - f_i|_\gamma^q,$$

for all $0 < q \leq 2$. Then by taking the summation, we immediately have $Q(\mathbf{u}, \mathbf{u}^k) \geq \Phi_{\beta, \gamma}(\mathbf{u})$; and hence (24) holds. Finally by taking the derivatives of $\Phi_{\beta, \gamma}(\mathbf{u})$ and $Q(\mathbf{u}, \mathbf{u}^k)$ with respect to \mathbf{u} , we have

$$\nabla \Phi_{\beta, \gamma}(\mathbf{u}) = \lambda \Psi^\top D_\beta(\mathbf{u}) \Psi \mathbf{u} + A^\top D_\gamma(\mathbf{u})(A\mathbf{u} - \mathbf{f}), \quad (29)$$

$$\nabla_1 Q(\mathbf{u}, \mathbf{u}^k) = \lambda \Psi D_\beta(\mathbf{u}^k) \Psi \mathbf{u} + A^\top D_\gamma(\mathbf{u}^k)(A\mathbf{u} - \mathbf{f}), \quad (30)$$

where $D_\beta(\cdot)$ and $D_\gamma(\cdot)$ are defined as in (14). Substituting $\mathbf{u} = \mathbf{u}^k$ into (29) and (30), we immediately have (26). \square

Based on Lemma 2, we can derive the following fundamental convergence theorem.

Theorem 1. *Let $0 < p, q \leq 2$. For the sequence $\{\mathbf{u}^k\}$ generated by the HQA, we have that $\{\Phi_{\beta,\gamma}(\mathbf{u}^k)\}$ is monotonically decreasing and convergent.*

Proof: The theorem is a direct result of Lemma 2. First, $\{\Phi_{\beta,\gamma}(\mathbf{u}^k)\}$ is bounded from below by 0. In Lemma 2, we have shown that the HQA is an MMA, which implies that

$$\Phi_{\beta,\gamma}(\mathbf{u}^{k+1}) \leq Q(\mathbf{u}^{k+1}, \mathbf{u}^k) \leq Q(\mathbf{u}^k, \mathbf{u}^k) = \Phi_{\beta,\gamma}(\mathbf{u}^k). \quad (31)$$

Here the first inequality and the last equality follow from (24) and (25), while the second inequality holds because \mathbf{u}^{k+1} is a minimizer of $Q(\mathbf{u}, \mathbf{u}^k)$. \square

3.2 Convergence of \mathbf{u}^k for $1 \leq p, q \leq 2$

Note that if $0 < p, q < 1$, the ℓ_p - ℓ_q minimization problem is non-convex. Hence, in the following, we discuss the convergence of \mathbf{u}^k for $1 \leq p, q \leq 2$ only. In order that \mathbf{u}^k is solvable from (16) and hence our HQA will not break down, we need the following assumption:

$$\ker(\Psi^\top \Psi) \cap \ker(A^\top A) = \{\mathbf{0}\}. \quad (32)$$

We remark that this assumption is very general and usually satisfied. For example, in regularization problems, Ψ is usually a difference operator, and hence is a high-pass filter; whereas A is a blurring operator, and hence is a low-pass filter. Therefore, (32) holds. For compressive sensing, Ψ is usually taken to be an orthogonal transform and we have $\Psi^\top \Psi = I$. Hence, $\ker(\Psi^\top \Psi) = \{\mathbf{0}\}$ which implies (32) holds for any A .

In [11, 40], the authors gave the convergence proof of general MMAs when the objective function Φ and its corresponding tangent majorant function Q satisfy Hypotheses 4.1 and 4.2 there. The convergence proof for our HQA will follow closely the proofs there. More precisely, we will show that our $\Phi_{\beta,\gamma}$ defined in (8) and our Q defined in (27) do satisfy the hypotheses, and hence the convergence follows immediately. Let us write out the hypotheses below.

Hypothesis 1. *[Hypothesis 4.1 in [11]]*

1. Φ is twice continuously differentiable and strictly convex.
2. Φ is coercive, i.e., $\lim_{\|\mathbf{u}\|_2 \rightarrow \infty} \Phi(\mathbf{u}) = \infty$.
3. Φ is bounded from below.

Hypothesis 2. *[Hypothesis 4.2 in [11]]*

- (a) *There exists a properly defined function $C : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ such that*
 - (i) $Q(\mathbf{u}, \mathbf{v}) = \Phi(\mathbf{v}) + (\mathbf{u} - \mathbf{v})^\top \nabla \Phi(\mathbf{v}) + \frac{1}{2}(\mathbf{u} - \mathbf{v})^\top C(\mathbf{v})(\mathbf{u} - \mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$.
 - (ii) C is continuous.

(iii) *There exists a constant η such that, for the smallest eigenvalue $\lambda_{\min}(C(\mathbf{v}))$ of $C(\mathbf{v})$, the following inequality holds: $\lambda_{\min}(C(\mathbf{v})) \geq \eta > 0$, for all $\mathbf{v} \in \mathbb{R}^n$.*

(b) $\Phi(\mathbf{u}) \leq Q(\mathbf{u}, \mathbf{v})$ for all $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$.

Lemma 3. *Let $1 \leq p, q \leq 2$ and $\ker(\Psi^\top \Psi) \cap \ker(A^\top A) = \{\mathbf{0}\}$. Then $\Phi_{\beta, \gamma}(\mathbf{u})$ defined in (8) satisfies Hypothesis 1. In particular, $\Phi_{\beta, \gamma}(\mathbf{u})$ has a unique minimizer.*

Proof: By the definition of $\Phi_{\beta, \gamma}$ in (8), it is obvious that $\Phi_{\beta, \gamma}$ is twice continuously differentiable and bounded from below by 0. We thus only need to prove the strict convexity and coercivity.

We start with the strict convexity. Taking derivatives on both sides of (29), we have

$$\nabla^2 \Phi_{\beta, \gamma}(\mathbf{u}) = \lambda \Psi^\top P_\beta(\mathbf{u}) \Psi + A^\top P_\gamma(\mathbf{u}) A,$$

where $P_\beta(\mathbf{u})$ and $P_\gamma(\mathbf{u}) \in \mathbb{R}^{n \times n}$ are the diagonal matrices with their i th diagonal entries being $|\Psi_i \mathbf{u}|_\beta^{p-4} (\beta + (p-1)|\Psi_i \mathbf{u}|^2)$ and $|A_i \mathbf{u} - f_i|_\gamma^{q-4} (\gamma + (q-1)|A_i \mathbf{u} - f_i|^2)$ respectively. Here recall our convention that when p or $q = 2$, the corresponding β or γ should be set to 0 because there is no need to smooth the term. Bounding each diagonal entry from below, we have

$$\nabla^2 \Phi_{\beta, \gamma}(\mathbf{u}) \succeq c_1 \Psi^\top \Psi + c_2 A^\top A,$$

where

$$c_1 =: \begin{cases} \frac{\lambda \beta}{\|\Psi\|_\infty \|\mathbf{u}\|_\infty^{4-p}}, & 1 \leq p < 2, \\ \lambda, & p = 2, \end{cases} \quad c_2 =: \begin{cases} \frac{\gamma}{\|A\|_\infty \|\mathbf{u}\|_\infty + \|\mathbf{f}\|_\infty^{4-q}}, & 1 \leq q < 2, \\ 1, & q = 2. \end{cases}$$

By the assumption (32), we have $\nabla^2 \Phi_{\beta, \gamma}(\mathbf{u}) \succ 0$, and the strict convexity of $\Phi_{\beta, \gamma}(\mathbf{u})$ is proven.

Next we show the coercivity. Note that $f(\cdot) = |\cdot|^p$ is convex for $p \geq 1$, which implies that

$$\frac{1}{n} \sum_{i=1}^n f(x_i) \geq f\left(\frac{1}{n} \sum_{i=1}^n x_i\right). \quad (33)$$

We rewrite $\Phi_{\beta, \gamma}$ in (8) into a summation form and then apply (33). Then we have

$$\begin{aligned} \Phi_{\beta, \gamma}(\mathbf{u}) &= \frac{\lambda}{p} \sum_{i=1}^n |\Psi_i \mathbf{u}|_\beta^p + \frac{1}{q} \sum_{i=1}^n |A_i \mathbf{u} - f_i|_\gamma^q \\ &\geq \frac{\lambda n}{p} \left(\frac{1}{n} \sum_{i=1}^n |\Psi_i \mathbf{u}|_\beta\right)^p + \frac{n}{q} \left(\frac{1}{n} \sum_{i=1}^n |A_i \mathbf{u} - f_i|_\gamma\right)^q. \end{aligned} \quad (34)$$

Next we apply the inequality $\sum_i |a_i| \geq \sqrt{\sum_i |a_i|^2}$ to (34), then we have

$$\begin{aligned} \Phi_{\beta,\gamma}(\mathbf{u}) &\geq \frac{\lambda n^{1-p}}{p} \left(\sqrt{\mathbf{u}^\top \Psi^\top \Psi \mathbf{u} + \beta n} \right)^p + \frac{n^{1-q}}{q} \left(\sqrt{(\mathbf{A}\mathbf{u} - \mathbf{f})^\top (\mathbf{A}\mathbf{u} - \mathbf{f}) + \gamma n} \right)^q \\ &\geq c_3 \left[\left(\sqrt{\mathbf{u}^\top \Psi^\top \Psi \mathbf{u} + \beta n} \right)^p + \left(\sqrt{(\mathbf{A}\mathbf{u} - \mathbf{f})^\top (\mathbf{A}\mathbf{u} - \mathbf{f}) + \gamma n} \right)^q \right], \end{aligned} \quad (35)$$

where $c_3 = \min\{\lambda n^{1-p}/p, n^{1-q}/q\}$. Now, we prove the coercivity of $\Phi_{\beta,\gamma}(\mathbf{u})$ by contradiction. Define

$$\phi_{\beta,\gamma}(\mathbf{u}) = \mathbf{u}^\top (\Psi^\top \Psi + A^\top A) \mathbf{u} - 2\mathbf{f}^\top \mathbf{A}\mathbf{u} + \|\mathbf{f}\|_2^2 + (\beta + \gamma)n.$$

Since $\Psi^\top \Psi + A^\top A \succ 0$, $\sigma^2 =: \lambda_{\min}(\Psi^\top \Psi + A^\top A) > 0$. Thus, if $\|\mathbf{u}\|_2 \rightarrow \infty$, we see that $\lim_{\|\mathbf{u}\|_2 \rightarrow \infty} \frac{\phi_{\beta,\gamma}(\mathbf{u})}{\|\mathbf{u}\|_2^2} \geq \sigma^2$. Hence, $\phi_{\beta,\gamma}(\mathbf{u})$ is coercive, i.e. for any $M_0 > 0$, there exists $M_1 > 0$, for any $\|\mathbf{u}\|_2 \geq M_1$, then we have $\phi_{\beta,\gamma}(\mathbf{u}) > M_0$. Suppose that $\Phi_{\beta,\gamma}(\mathbf{u})$ is non coercive, i.e.

$$\lim_{\|\mathbf{u}\|_2 \rightarrow \infty} \Phi_{\beta,\gamma}(\mathbf{u}) \neq \infty.$$

Thus, for the above M_0 , for any $M_2 \geq M_1$, there exists $\|\mathbf{u}_0\|_2 \geq M_2$, but yet $\Phi_{\beta,\gamma}(\mathbf{u}_0) \leq c_3 \min\{(M_0/2)^p, (M_0/2)^q\}$. Together with (35), we have

$$\begin{aligned} \mathbf{u}_0^\top \Psi^\top \Psi \mathbf{u}_0 + \beta n &\leq \frac{M_0}{2}, \\ (\mathbf{A}\mathbf{u}_0 - \mathbf{f})^\top (\mathbf{A}\mathbf{u}_0 - \mathbf{f}) + \gamma n &\leq \frac{M_0}{2}. \end{aligned}$$

Summing these two inequalities up, we have $\phi_{\beta,\gamma}(\mathbf{u}_0) \leq M_0$, which is a contradiction to the coercivity of $\phi_{\beta,\gamma}(\mathbf{u})$. \square

Regarding Hypothesis 2, in fact, we cannot show that Hypothesis 2 holds for arbitrary vectors \mathbf{v} . We can only show that it holds for $\mathbf{v} = \mathbf{u}^k$, the sequence generated by HQA. However, as we will see later in Theorem 2, it will be enough for us to prove the convergence of HQA.

Lemma 4. *Let $1 \leq p, q \leq 2$ and $\ker(\Psi^\top \Psi) \cap \ker(A^\top A) = \{\mathbf{0}\}$. Then $\Phi_{\beta,\gamma}(\mathbf{u})$ defined in (8) and $Q(\mathbf{u}, \mathbf{u}^k)$ defined in (27) satisfy Hypothesis 2 at $\mathbf{v} = \mathbf{u}^k$. In particular, the coefficient matrix of the linear system (16) is invertible.*

Proof: By definition of $Q(\mathbf{u}, \mathbf{u}^k)$ in (28), $Q(\mathbf{u}, \mathbf{u}^k)$ is quadratic in \mathbf{u} and its Hessian matrix is given by

$$\nabla_1^2 Q(\mathbf{u}, \mathbf{u}^k) = \lambda \Psi^\top D_\beta(\mathbf{u}^k) \Psi + A^\top D_\gamma(\mathbf{u}^k) A, \quad (36)$$

which is independent of \mathbf{u} . Taking the Taylor expansion for $Q(\mathbf{u}, \mathbf{u}^k)$ at \mathbf{u}^k , we have

$$Q(\mathbf{u}, \mathbf{u}^k) = Q(\mathbf{u}^k, \mathbf{u}^k) + \langle \nabla_1 Q(\mathbf{u}^k, \mathbf{u}^k), \mathbf{u} - \mathbf{u}^k \rangle + \frac{1}{2} (\mathbf{u} - \mathbf{u}^k)^\top \nabla_1^2 Q(\mathbf{u}^k, \mathbf{u}^k) (\mathbf{u} - \mathbf{u}^k).$$

Since we have proven that the HQA is indeed an MMA in Lemma 2, we can replace $Q(\mathbf{u}^k, \mathbf{u}^k)$ and $\nabla_1 Q(\mathbf{u}^k, \mathbf{u}^k)$ by $\Phi_{\beta, \gamma}(\mathbf{u}^k)$ and $\nabla \Phi_{\beta, \gamma}(\mathbf{u}^k)$ respectively in the equality above and then we obtain

$$Q(\mathbf{u}, \mathbf{u}^k) = \Phi_{\beta, \gamma}(\mathbf{u}^k) + \langle \nabla \Phi_{\beta, \gamma}(\mathbf{u}^k), \mathbf{u} - \mathbf{u}^k \rangle + \frac{1}{2} (\mathbf{u} - \mathbf{u}^k)^\top \nabla_1^2 Q(\mathbf{u}^k, \mathbf{u}^k) (\mathbf{u} - \mathbf{u}^k).$$

Notice that $\{\Phi_{\beta, \gamma}(\mathbf{u}^k)\}$ is bounded from below by 0 from the definition in (8). In addition, recalling that $\{\Phi_{\beta, \gamma}(\mathbf{u}^k)\}$ is monotonically decreasing and bounded from above by $\Phi_{\beta, \gamma}(\mathbf{u}^0)$ by (31). Therefore, by coercivity, see Hypothesis 1(b), $\{\|\mathbf{u}^k\|_2\}$ must be bounded from above. Denote the bound by M . Recalling the definition of $D_\beta(\mathbf{u}^k)$, $D_\gamma(\mathbf{u}^k)$ in (16), we have

$$\lambda_{\min}(\nabla_1^2 Q_1(\mathbf{u}^k, \mathbf{u}^k)) \geq \lambda_{\min}(c_4 \Psi^\top \Psi + c_5 A^\top A) := \eta, \quad (37)$$

where

$$c_4 = \begin{cases} \frac{\lambda}{\|\Psi\|_\infty M | \beta|}, & 1 \leq p < 2, \\ \lambda, & p = 2, \end{cases} \quad c_5 = \begin{cases} \frac{1}{|M \|A\|_\infty + \|\mathbf{f}\|_\infty \gamma}, & 1 \leq q < 2, \\ 1, & q = 2. \end{cases}$$

By (32), $\eta > 0$. Hypothesis 2(a)(iii) holds.

Hypothesis 2(b) is just (24), and hence is true. Finally notice that the coefficient matrix of the linear system in (16) is precisely $\nabla_1^2 Q(\mathbf{u}^k, \mathbf{u}^k)$ in (36) and hence by (37), it is invertible. \square

Since Hypothesis 2 is only valid for \mathbf{u}^k and not for arbitrary vectors \mathbf{v} , we cannot directly apply the convergence theorems in [11]. However, the proof in [11] can easily be adapted to prove the following two convergence theorems for HQA.

Theorem 2. *Let $1 \leq p, q \leq 2$ and $\ker(\Psi^\top \Psi) \cap \ker(A^\top A) = \{\mathbf{0}\}$. For the sequence $\{\mathbf{u}^k\}$ generated by HQA, we have*

- (a) $\lim_{k \rightarrow \infty} \|\mathbf{u}^k - \mathbf{u}^{k+1}\|_2 = 0$;
- (b) $\{\mathbf{u}^k\}$ converges to the unique minimizer \mathbf{u}^* of $\Phi_{\beta, \gamma}(\mathbf{u})$ from any initial guess \mathbf{u}^0 .

Proof:

- (a) We see from (36) that $Q(\mathbf{u}, \mathbf{u}^k)$ is quadratic in \mathbf{u} . Taking Taylor expansion of $Q(\mathbf{u}, \mathbf{u}^k)$ at \mathbf{u}^{k+1} , we have

$$Q(\mathbf{u}, \mathbf{u}^k) = Q(\mathbf{u}^{k+1}, \mathbf{u}^k) + \langle \nabla_1 Q(\mathbf{u}^{k+1}, \mathbf{u}^k), \mathbf{u} - \mathbf{u}^{k+1} \rangle + \frac{1}{2} (\mathbf{u} - \mathbf{u}^{k+1})^\top \nabla_1^2 Q(\mathbf{u}^k, \mathbf{u}^k) (\mathbf{u} - \mathbf{u}^{k+1}). \quad (38)$$

By (23), we have $\nabla_1 Q(\mathbf{u}^{k+1}, \mathbf{u}^k) = 0$. By taking $\mathbf{u} = \mathbf{u}^k$ in (38) and using (37), we thus have

$$Q(\mathbf{u}^k, \mathbf{u}^k) \geq Q(\mathbf{u}^{k+1}, \mathbf{u}^k) + \frac{\eta}{2} \|\mathbf{u}^k - \mathbf{u}^{k+1}\|_2^2,$$

where $\eta > 0$. Together with (31), we obtain that

$$\Phi_{\beta,\gamma}(\mathbf{u}^k) - \Phi_{\beta,\gamma}(\mathbf{u}^{k+1}) \geq Q(\mathbf{u}^k, \mathbf{u}^k) - Q(\mathbf{u}^{k+1}, \mathbf{u}^k) \geq \frac{\eta}{2} \|\mathbf{u}^k - \mathbf{u}^{k+1}\|_2^2. \quad (39)$$

By Theorem 1, the convergence of $\{\Phi_{\beta,\gamma}(\mathbf{u})\}$ implies that

$$\lim_{k \rightarrow 0} \Phi_{\beta,\gamma}(\mathbf{u}^k) - \Phi_{\beta,\gamma}(\mathbf{u}^{k+1}) = 0$$

Together with (39) and $\eta > 0$, we have $\lim_{k \rightarrow \infty} \|\mathbf{u}^k - \mathbf{u}^{k+1}\|_2 = 0$.

- (b) By the proof for Lemma 4, we know that the sequence $\{\|\mathbf{u}^k\|_2\}$ is bounded from above. Hence it converges to the unique minimizer \mathbf{u}^* if and only if all convergent subsequences of $\{\mathbf{u}^k\}$ converge to \mathbf{u}^* . Let $\{\mathbf{u}^{k_j}\}$ be an arbitrary convergence subsequence of $\{\mathbf{u}^k\}$ that converges to $\bar{\mathbf{u}}$. To finish the proof for the theorem, we only need to prove that $\bar{\mathbf{u}} = \mathbf{u}^*$. Since $Q(\mathbf{u}, \mathbf{u}^{k_j})$ is quadratic in \mathbf{u} , we have

$$\begin{aligned} Q(\mathbf{u}, \mathbf{u}^{k_j}) &= Q(\mathbf{u}^{k_j}, \mathbf{u}^{k_j}) + \langle \nabla_1 Q(\mathbf{u}^{k_j}, \mathbf{u}^{k_j}), \mathbf{u} - \mathbf{u}^{k_j} \rangle \\ &\quad + \frac{1}{2} (\mathbf{u} - \mathbf{u}^{k_j})^\top \nabla_1^2 Q(\mathbf{u}^{k_j}, \mathbf{u}^{k_j}) (\mathbf{u} - \mathbf{u}^{k_j}). \end{aligned}$$

By taking the partial derivative with respect to \mathbf{u} and substituting (26), we then have

$$\nabla_1 Q(\mathbf{u}, \mathbf{u}^{k_j}) = \nabla \Phi_{\beta,\gamma}(\mathbf{u}^{k_j}) + \nabla_1^2 Q(\mathbf{u}^{k_j}, \mathbf{u}^{k_j}) (\mathbf{u} - \mathbf{u}^{k_j}).$$

$\nabla_1 Q(\mathbf{u}, \mathbf{v})$ is continuous since $\Phi_{\beta,\gamma}$ is twice continuously differentiable by Hypothesis 1(a), and $C(\mathbf{v}) = \nabla_1^2 Q(\mathbf{v}, \mathbf{v})$ is continuous by Hypothesis 2(a)(ii). Letting $\mathbf{u} = \mathbf{u}^{k_j+1}$ and using (23), we then have

$$0 = \nabla_1 Q(\mathbf{u}^{k_j+1}, \mathbf{u}^{k_j}) = \nabla \Phi_{\beta,\gamma}(\mathbf{u}^{k_j}) + \nabla_1^2 Q(\mathbf{u}^{k_j}, \mathbf{u}^{k_j}) (\mathbf{u}^{k_j+1} - \mathbf{u}^{k_j}). \quad (40)$$

By (a), we know that $\lim_{j \rightarrow \infty} \|\mathbf{u}^{k_j+1} - \mathbf{u}^{k_j}\|_2 = 0$. This implies that $\lim_{j \rightarrow \infty} \mathbf{u}^{k_j+1} = \bar{\mathbf{u}}$. Taking limits to the both sides of (40), we obtain

$$\begin{aligned} 0 &= \lim_{j \rightarrow \infty} \nabla_1 Q(\mathbf{u}^{k_j+1}, \mathbf{u}^{k_j}) \\ &= \nabla_1 Q(\lim_{j \rightarrow \infty} \mathbf{u}^{k_j+1}, \lim_{j \rightarrow \infty} \mathbf{u}^{k_j}) \\ &= \nabla_1 Q(\bar{\mathbf{u}}, \bar{\mathbf{u}}) = \nabla \Phi_{\beta,\gamma}(\bar{\mathbf{u}}) + \nabla_1^2 Q(\bar{\mathbf{u}}, \bar{\mathbf{u}}) (\bar{\mathbf{u}} - \bar{\mathbf{u}}) = \nabla \Phi_{\beta,\gamma}(\bar{\mathbf{u}}). \end{aligned}$$

By the uniqueness of the minimizer, see Lemma 3, we can conclude that $\bar{\mathbf{u}} = \mathbf{u}^*$. \square

Theorem 3. *Let $1 \leq p, q \leq 2$ and $\ker(\Psi^\top \Psi) \cap \ker(A^\top A) = \{\mathbf{0}\}$. Let \mathbf{u}^* be the unique minimizer of $\Phi_{\beta,\gamma}(\mathbf{u})$ and*

$$\Lambda := 1 - \lambda_{\min} \left(\nabla_1^2 Q(\mathbf{u}^*, \mathbf{u}^*)^{-1} \nabla^2 \Phi_{\beta,\gamma}(\mathbf{u}^*) \right).$$

Then $\Lambda < 1$ and the sequence $\{\Phi_{\beta,\gamma}(\mathbf{u}^k)\}$ has a linear convergence rate of at most Λ while the sequence $\{\mathbf{u}^k\}$ is r -linearly convergent with a convergence rate of at most $\sqrt{\Lambda}$.

To prove Theorem 3, we can follow the proof of Theorem 6.1 in [11].

4 Numerical Results

In this section, we test our algorithm on deblurring images corrupted by impulse noise and restoring MR images from few k -space data. Recall that the HQA given in (16) is for solving the smoothed ℓ_p - ℓ_q problem (8) for a fixed pair of smoothing parameters β and γ . To solve the original ℓ_p - ℓ_q problem (1), we apply the idea of continuation method on HQA for a sequence of $\{\beta^l, \gamma^l\}$ going to zero. We note that continuation methods has been used in solving TV problems before, see [9, 44]. We summarize the HQA for (1) in **Algorithm 1**.

Algorithm 1 The HQA for solving (1) with $0 < p, q \leq 2$:

- (1) Initialize $\beta^0, \gamma^0, \mathbf{u}^0$;
- (2) For $l = 0, 1, \dots$ until stopping criteria are met, do:
 - (a) For $k = 0, 1, \dots$ until stopping criteria are met, do:
 - (i) Initialize $\mathbf{u}^{l,0} = \mathbf{u}^{l-1}$;
 - (ii) Get $\mathbf{u}^{l,k+1}$ by solving

$$\left(\lambda \Psi^\top D_{\beta^l}(\mathbf{u}^{l,k}) \Psi + A^\top D_{\gamma^l}(\mathbf{u}^{l,k}) A \right) \mathbf{u} = A^\top D_{\gamma^l}(\mathbf{u}^{l,k}) \mathbf{f}, \quad (41)$$

where $D_{\beta^l}(\cdot)$ and $D_{\gamma^l}(\cdot)$ are diagonal matrices given as in (14) with $\beta = \beta^l, \gamma = \gamma^l$.

- (b) Set \mathbf{u}^l to be the final solution from part (a).
 - (c) Update β^l, γ^l to $\beta^{l+1}, \gamma^{l+1}$.
-

In all the following tests, we take the stopping criteria for the inner loop as

$$\tau := \|\nabla \Phi_{\beta^l, \gamma^l}(\mathbf{u}^{l,k})\|_2 > 0.2,$$

where $\nabla \Phi_{\beta, \gamma}(\mathbf{u})$ has been given in (29).

4.1 Numerical Results on the TV- ℓ_1 Image Restoration

In this section, we apply **Algorithm 1** to deblur images that are corrupted by impulse noise. Here, (41) is replaced by (19) with $\mathbf{u}^{l,k}$ replacing \mathbf{u}^k . The deblurring problem has been discussed recently in many papers, see for examples [16, 41, 44]. Among all these methods, the FTVDM and the ALM are the most efficient linear algorithms; and according to the numerical results in [41], ALM is the fastest one. Hence in this paper, we compare our HQA with FTVDM and ALM only. The FTVDM and ALM codes we used here are provided by the authors in [41] and we use the same parameters as in [41]. For more details on the algorithms and the parameters, please consult [41, 44].

We test three 256×256 images: Barbara, Bridge and Goldhill. The matrix A is the blurring matrix corresponding to the Gaussian blur generated by the

MATLAB command

```
fspecial('Gaussian', [7 , 7], 5).
```

Then salt-and-pepper noise is added to the blurred image to obtain the observed image \mathbf{f} . The noise levels are taken to be 30 %, 40 %, 50 %, 60 %. For all methods, the regularization parameter λ is set to $1/13, 1/10, 1/8, 1/4$ for noise level 30 %, 40 %, 50 %, 60 % respectively. In our algorithm, we initialize $\mathbf{u}_0 = \text{rand}(\text{size}(\mathbf{f}))$. As in the FTVDM, to speed up the convergence and improve the resolution quality, we take large β, γ at the beginning and reduce them gradually to smaller ones respectively. We set β^l to be $10^{-3}, 10^{-4}, \dots, 10^{-16}$ and $\gamma^l = (\beta^l)^2$. Equation (19) is solved by the conjugate gradient (CG) method. Considering that more iterations for CG are needed with the decreasing of β , therefore we fix the iteration number in the inner loop to be $10 \times l$ at β^l . In all tests, we consider periodic boundary condition for the difference matrix A , as it is the boundary conditions used in the tests in [41]. We compare the accuracy of the methods by the signal-to-noise ratio (SNR) used in [41]. It is defined by

$$\text{SNR} := 10 \log_{10} \frac{\|\mathbf{u} - E(\mathbf{u})\|_2^2}{\|\hat{\mathbf{u}} - \mathbf{u}\|_2^2} (\text{dB}).$$

Here \mathbf{u} and $\hat{\mathbf{u}}$ denote the original image and the restored image respectively, and $E(\mathbf{u})$ is the mean gray-level value of the original image.

First we compare the speed of the three methods. Figures 1, 2 and 3 show the timing comparison of the three algorithms. Each point in the figures show the accumulated CPU time until that iteration and the corresponding SNR. The results show that our method is the fastest amongst the three methods. It is also the most accurate one.

Table 1. SNR of the restored images.

Image	Method	Salt-and-pepper noise			
		30 %	40 %	50 %	60 %
Barbara	ALM	13.93	13.35	12.45	11.37
	HQA	14.24	13.59	12.83	11.72
Bridge	ALM	11.85	10.95	10.13	8.52
	HQA	12.03	11.12	10.27	9.00
Goldhill	ALM	16.08	15.03	13.78	12.05
	HQA	16.50	15.32	14.10	12.72

From Figs. 1, 2 and 3, it is clear that FTVDM is the slowest amongst the three. In order to compare the accuracy of the two faster methods ALM and HQA more precisely, we list in Table 1 the average SNR of the recovered images in five trials by the two methods. To compare the timing fairly, we first run ALM until its stopping criteria [41] is satisfied, say with t_0 CPU seconds. Then we let

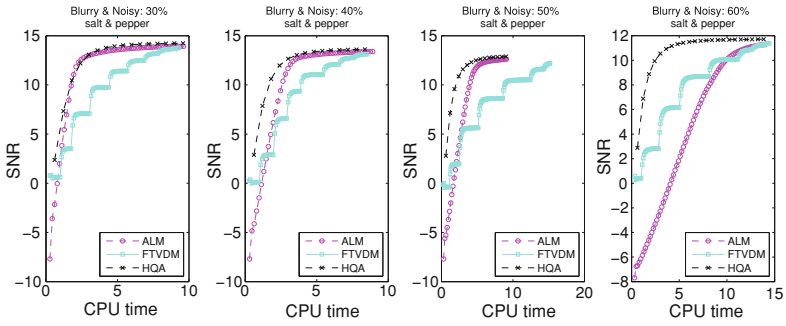


Fig. 1. SNR versus CPU time in seconds for “Barbara” with salt-and-pepper noise.

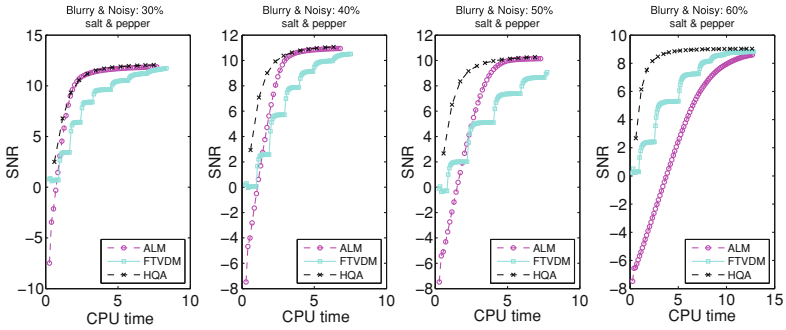


Fig. 2. SNR versus CPU time in seconds for “Bridge” with salt-and-pepper noise.

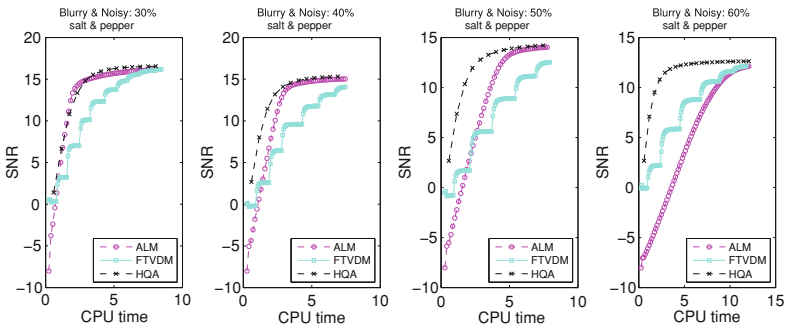


Fig. 3. SNR versus CPU time in seconds for “Goldhill” with salt-and-pepper noise.

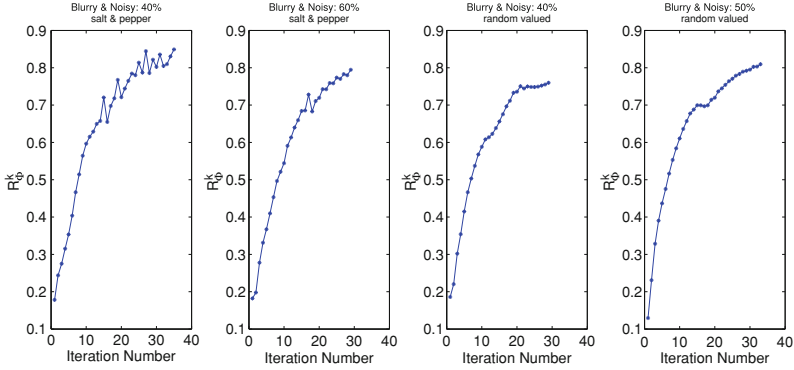


Fig. 4. The ratio $R_{\Phi}^k := \frac{\Phi_{\beta,\gamma}(\mathbf{u}^{k+1}) - \Phi_{\beta,\gamma}(\mathbf{u}^k)}{\Phi_{\beta,\gamma}(\mathbf{u}^k) - \Phi_{\beta,\gamma}(\mathbf{u}^{k-1})}$ versus iteration number for “Barbara”. The ratios are less than 1, illustrating the linear convergence of $\{\Phi_{\beta,\gamma}(\mathbf{u}^k)\}$.

HQA run until the CPU time of the k th iteration is just greater than t_0 . Then we record the SNR of the $(k - 1)$ th iteration as our result for HQA. We see from Table 1 that HQA is more accurate than ALM. Recovered images taking Barbara for example are shown in Fig. 7 for “eyeball” illustration.

We illustrate the convergence rate of $\{\Phi_{\beta,\gamma}(\mathbf{u}^k)\}$ and $\{\mathbf{u}^k\}$ in our HQA as mentioned in Theorem 3. We use the Barbara image as example. Since we do not have the true minimizer, we use

$$R_{\Phi}^k =: [\Phi_{\beta,\gamma}(\mathbf{u}^{k+1}) - \Phi_{\beta,\gamma}(\mathbf{u}^k)] / [\Phi_{\beta,\gamma}(\mathbf{u}^k) - \Phi_{\beta,\gamma}(\mathbf{u}^{k-1})]$$

and

$$R_{\mathbf{u}}^k =: [\|\mathbf{u}^{k+1} - \mathbf{u}^k\|_2 / \|\mathbf{u}^k - \mathbf{u}^{k-1}\|_2]$$

to estimate the convergence rate for $\{\Phi_{\beta,\gamma}(\mathbf{u}^k)\}$ and $\{\mathbf{u}^k\}$ respectively. In Fig. 4, we plot the ratio R_{Φ}^k against the iteration number. We see that the ratios are

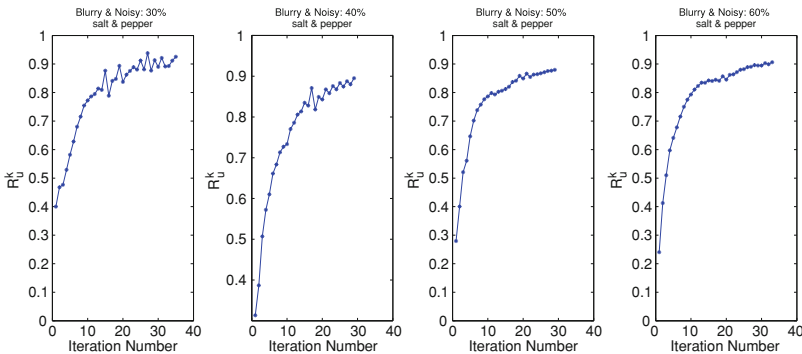


Fig. 5. The ratio $R_{\mathbf{u}}^k = \frac{\|\mathbf{u}^{k+1} - \mathbf{u}^k\|_2}{\|\mathbf{u}^k - \mathbf{u}^{k-1}\|_2}$ versus iteration number for “Barbara”. The ratios are less than 1, illustrating the linear convergence of $\{\mathbf{u}^k\}$.

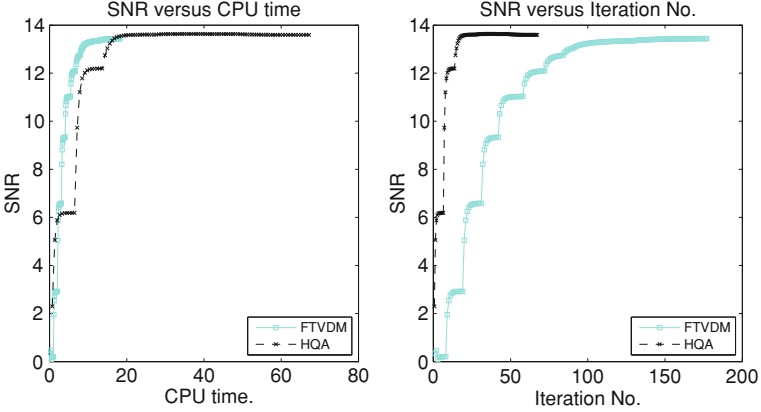


Fig. 6. The figure shows the comparison results of the FTVDM and the HQA for Barbara image with salt-and-pepper noise removal at noise level 40%. For HQA, in (18), $\beta = 10^{-1}, 10^{-2}, \dots, 10^{-17}$, and correspondingly $\gamma = \beta^2$. For FTVDM, $\theta_{\mathbf{w}} = 1, 2^{2/3}, \dots, 2^{10}$, correspondingly, $\theta_{\mathbf{z}} = 1, 2, \dots, 2^{15}$. At each jump, $\beta_k, \gamma_k, \theta_{\mathbf{w}}^k, \theta_{\mathbf{z}}^k$ jump to $\beta_{k+1}, \gamma_{k+1}, \theta_{\mathbf{w}}^{k+1}, \theta_{\mathbf{z}}^{k+1}$ in HQA and FTVDM.

all less than 1, indicating that $\{\Phi_{\beta, \gamma}(\mathbf{u}^k)\}$ is linearly convergent as stated in Theorem 3. In Fig. 5, we plot $R_{\mathbf{u}}^k$ against the iteration number. We see that $R_{\mathbf{u}}^k < c$ (c is a positive constant less than 1), indicating that $\{\mathbf{u}^k\}$ indeed is linearly convergent.

In [44], the original energy functional (17) is a non-differentiable functional of \mathbf{u} , hence auxiliary variables \mathbf{w}, \mathbf{z} and regularization parameters $\theta_{\mathbf{w}}, \theta_{\mathbf{z}}$ are taken to approximate (17). The approximated problem is

$$\min_{\mathbf{w}, \mathbf{z}, \mathbf{u}} \left\{ \lambda (\|\mathbf{w}\|_1 + \frac{\theta_{\mathbf{w}}}{2} \|\mathbf{w} - \nabla \mathbf{u}\|_2^2) + \|\mathbf{z}\|_1 + \frac{\theta_{\mathbf{z}}}{2} \|\mathbf{z} - (B\mathbf{u} - \mathbf{f})\|_2^2 \right\}$$

The parameters $\theta_{\mathbf{w}}$ and $\theta_{\mathbf{z}}$ are upper limited to be $\theta_{\mathbf{w}} = 2^{10}, \theta_{\mathbf{z}} = 2^{15}$ in the approximate problem. To speed up the convergence, $\theta_{\mathbf{w}}$ and $\theta_{\mathbf{z}}$ are both implemented in a continuous scheme; that is, let $\theta_{\mathbf{w}}$ and $\theta_{\mathbf{z}}$ take small values at the beginning and gradually increase their values to 2^{10} and 2^{15} respectively. Specially, a $\theta_{\mathbf{w}}$ -sequence $2^0, 2^{2/3}, 2^{4/3}, \dots, 2^{10}$ is tested. Accordingly, $\theta_{\mathbf{z}}$ is set to be $2^0, 2^1, 2^2, \dots, 2^{15}$.

A similar continuation scheme is also taken in our HQA. We take $\beta = 10^{-1}, 10^{-2}, \dots, 10^{-17}$, correspondingly, $\gamma = \beta^2$ and compare the FTVDM and the HQA. Figure 6 shows their comparison results on the SNR versus CPU time and SNR versus iteration number. The jump shows the improvements in SNR as $\theta_{\mathbf{w}}^k, \theta_{\mathbf{z}}^k, \beta^k, \gamma^k$ change to $\theta_{\mathbf{w}}^{k+1}, \theta_{\mathbf{z}}^{k+1}, \beta^{k+1}, \gamma^{k+1}$.



Fig. 7. Restored images.

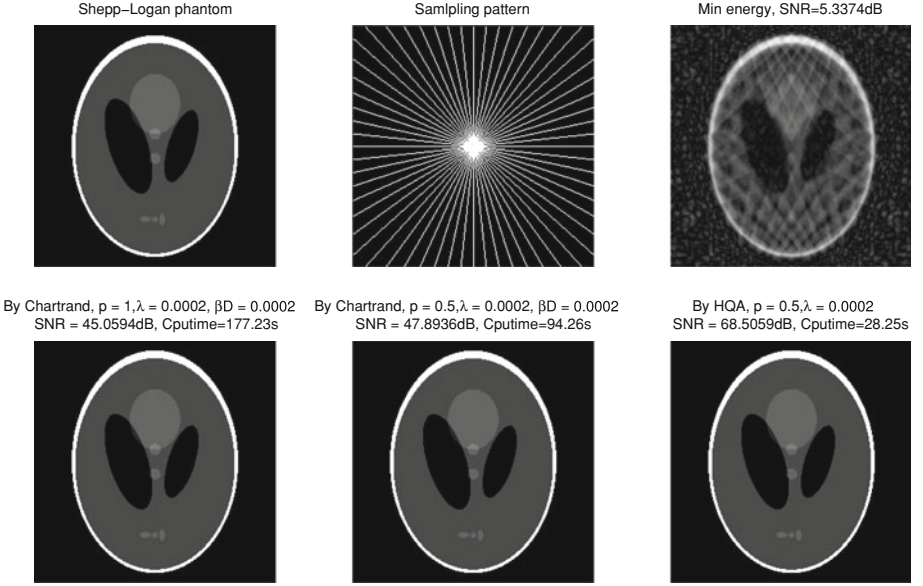


Fig. 8. The figure shows the reconstruction for the 256×256 Shepp-Logan phantom. Top (left): the original Shepp-Logan image; Top (middle): the 30 radial lines on the white pixels (11.32% sampling); Top (right): the backprojection reconstruction with 30 views, which is poor; Bottom (left): the reconstruction from 30 views by p -shrinkage algorithm with $p = 1$; Bottom (middle): the reconstruction by the p -shrinkage algorithm with $p = 1/2$; Bottom (right): the reconstruction by HQA with $p = 1/2$.

4.2 Numerical Results on the MR Image Reconstruction

In this section, we apply **Algorithm 1** to reconstruct MR image from few k -space data. Here (41) is replaced by (22) with $\mathbf{u}^{l,k}$ replacing \mathbf{u}^k . To test the efficiency of our algorithm, we compare our algorithm and the p -shrinkage algorithm by Chartrand in [12]. As in [12], we take $p = 1/2$. In addition, we also give the numerical results by ℓ_1 -norm regularized model for comparison. We test our algorithm on the two images: 256×256 Shepp-Logan phantom and 224×224 Brain image. In all the tests, we set β to be $10^{-4}, 10^{-5}, \dots, 10^{-14}$, and $\gamma = \beta$ correspondingly. Moreover, we just use the simple CG method to solve the corresponding linear system (22).

We begin with the Shepp-Logan phantom. As in [12], because the phantom has a very sparse gradient, we do not use the wavelet regularization, and let $\delta = 0$ in (20). We show the comparison results on the MR image reconstruction from 10 radial lines (3.85% sampling), 22 radial lines (8.36% sampling), and 30 radial lines (11.32% sampling) respectively. In all the three tests, we take $\lambda = 0.0002$. When $p = 1$, the p -shrinkage [12] is actually the soft-thresholding. The results are shown in Figs. 8, 9 and 10. In all three figures, we see that our HQA can reach better reconstruction (at least 13–16 dB better) using less

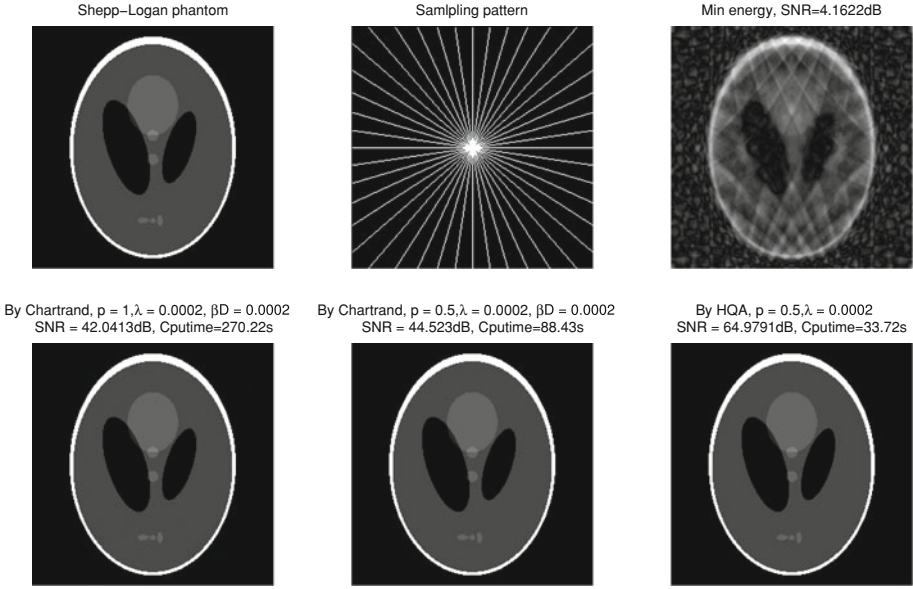


Fig. 9. The figure shows the reconstruction for the 256×256 Shepp-Logan phantom. Top (left): the original Shepp-Logan image; Top (middle): the 22 radial lines on the white pixels (8.36% sampling); Top (right): the backprojection reconstruction with 22 views, which is poor; Bottom (left): the reconstruction from 22 views by p -shrinkage algorithm with $p = 1$; Bottom (middle): the reconstruction by the p -shrinkage algorithm with $p = 1/2$; Bottom (right): the reconstruction by HQA with $p = 1/2$.

computational time (at least 1/2 of the time) than the p -shrinkage algorithm. Among all the results, the ℓ_1 -norm regularization model takes the most time to obtain a suitable reconstruction image, especially when the k -space data are very few.

Next, we apply our algorithm to recover the MR brain image in the presence of noise. We set that the image is corrupted by the white Gaussian noise with noise level $\sigma = 5$. Here, by error and trials, we take $\lambda = 0.002$. Our results show that the recovered images have higher quality by the $\|\nabla \mathbf{u}\|_p^p$ regularization model than by the $\|\nabla \mathbf{u}\|_p^p + \|W\mathbf{u}\|_p^p$. Hence, here, we show the recovered results from (22) with $\delta = 0$. The comparison results are shown in Fig. 11. For the brain image, we take 40 views (16.97% sampling). The results show that our HQA can reach the best reconstruction with the clearest background in the least amount of time.

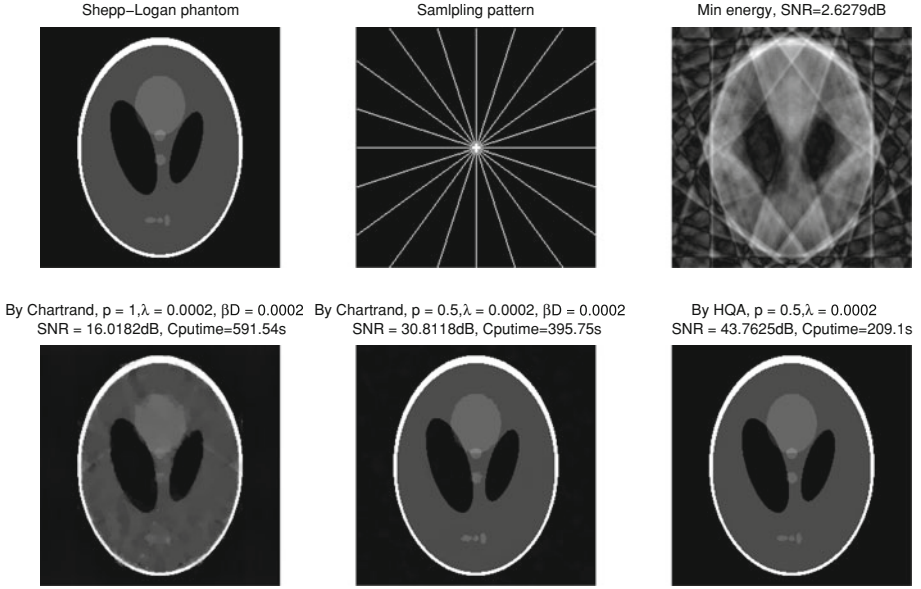


Fig. 10. The figure shows the reconstruction for the 256×256 Shepp-Logan phantom. Top (left): the original Shepp-Logan image; Top (middle): the 10 radial lines on the white pixels (3.85% sampling); Top (right): the backprojection reconstruction with 10 views, which is poor; Bottom (left): the reconstruction from 10 views by p -shrinkage algorithm with $p = 1$; Bottom (middle): the reconstruction by the p -shrinkage algorithm with $p = 1/2$; Bottom (right): the reconstruction by HQA with $p = 1/2$. From the results, we find that it will take more time to reach a good reconstruction from fewer k -space data. By ℓ_1 -norm regularized model, it is still difficult to obtain a good result even with much time, while the ℓ_p -norm regularized models ($0 < p < 1$) do.

5 Conclusion

In this paper, we study the half-quadratic technique for ℓ_p -norm and propose an algorithm for solving ℓ_p - ℓ_q ($0 < p, q \leq 2$) minimization problem. We show that the algorithm for the related minimization problem with regularization smoothing parameters β and γ is equivalent to a majorize-minimize algorithm. Weak convergence result for $0 < p$ or $q < 1$, and linear convergence rate for $1 \leq p, q \leq 2$ are obtained immediately. We will consider the convergence of the original non-smooth problems in our future work. We compare our algorithm with standard ones in the TV- ℓ_1 minimization problem and the MR image reconstruction. The results show that our algorithm can reach better reconstruction results with less computational cost.

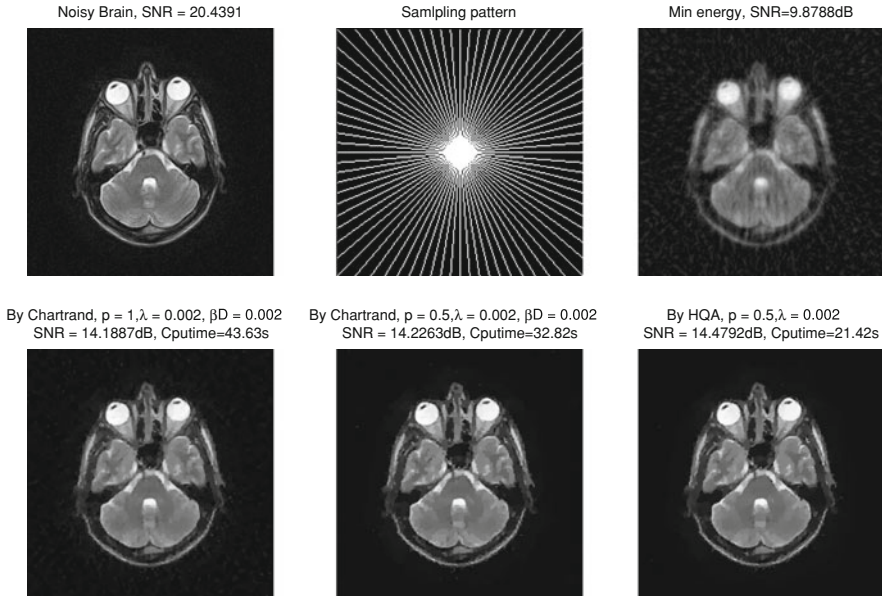


Fig. 11. The figure shows the reconstruction for the 224×224 real brain image. Top (left): the noisy brain image with noise level $\sigma = 5$, which is generated with the Matlab command: “`imnoise(x,'gaussian',0, σ^2)`”; Top (middle): the 40 radial lines on the white pixels (16.97% sampling); Top (right): the backprojection reconstruction with 40 views, which is poor; Bottom (left): the reconstruction from 40 views by p -shrinkage algorithm with $p = 1$; Bottom (middle): the reconstruction by the p -shrinkage algorithm with $p = 1/2$; Bottom (right): the reconstruction by HQA with $p = 1/2$.

Acknowledgement. The authors would like to thank the financial support of project in Nanyang Technological University.

References

1. Amaldi, E., Kann, V.: On the approximability of minimizing nonzero variables or unsatisfied relations in linear system. *Theoretical Comp. Sci.* **209**, 237–260 (1998)
2. Baraniuk, R., Steeghs, P.: Compressive radar imaging. *Radar Conference, 2007 IEEE*, pp. 128–133 (2007)
3. Berger, C.R., Areta, J., Pattipati, K., Willett, P.: Compressed sensing - a look beyond linear programming. In: *33rd International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 3857–3860 (2008)
4. Blake, A., Zisserman, A.: Visual reconstruction and the GNC algorithm. In: *Parallel Architectures and Computer Vision*, pp. 33–48 (1988)
5. Cai, J.F., Osher, S., Shen, Z.: Linearized Bregman iterations for compressed sensing. In: *UCLA CAM Report*, 08-06 (2008)
6. Candè, E.J., Romberg, J.: Signal recovery from random projections. *Proc. SPIE Comput. Imaging III* **5674**, 76–86 (2005)

7. Candè, E.J., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory* **52**, 489–509 (2006)
8. Chan, R.H., Chan, T., Shen, L., Shen, Z.: Wavelet algorithms for high-resolution image reconstruction. *SIAM J. Sci. Comput.* **24**, 1408–1432 (2003)
9. Chan, R.H., Chan, T.F., Zhou, H.M.: Continuation method for total variation denoising problems. In: *UCLA CAM Report 95-18* (1995)
10. Chan, T., Esedoglu, S.: Aspects of total variation regularized L^1 function approximation. *SIAM J. Appl. Math.* **65**, 1817–1837 (2005)
11. Chan, T.F., Mulet, P.: On the convergence of the lagged diffusivity fixed point method in total variation image restoration. *SIAM J. Numer. Anal.* **36**, 354–367 (1999)
12. Chartrand, R.: Fast algorithms for non-convex compressive sensing: MRI reconstruction from very few data. In: *IEEE International Symposium on Biomedical Imaging (ISBI)* (2009)
13. Chartrand, R.: Exact reconstruction of sparse signals via non-convex minimization. *IEEE Signal Process. Lett.* **14**, 707–710 (2007)
14. Chartrand, R., Yin, W.: Iteratively reweighted algorithms for compressive sensing. In: *33rd International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 3869–3872 (2008)
15. Chartrand, R., Staneva, V.: Restricted isometry properties and non-convex compressive sensing. *Inverse Problems* **24**, 1–14 (2008)
16. Dong, Y.Q., Hintermüller, M., Neri, M.: An efficient primal-dual method for L^1 -TV image restoration. *SIAM J. Imaging Sci.* **2**, 1168–1189 (2009)
17. Donoho, D.: Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006)
18. Ekerland, I., Témmam, R.: *Convex Analysis and Variational Problems*. SIAM, Philadelphia (1999)
19. Geman, D., Yang, C.: Nonlinear image recovery with half-quadratic regularization and FFTs. *IEEE Trans. Imaging Proc.* **4**, 932–946 (1995)
20. Goldstein, T., Osher, S.: The split Bregman L^1 regularized problems. *SIAM J. Imaging Sci.* **2**, 323–343 (2009)
21. He, L., Chang, T.C., Osher, S.: MR image reconstruction from sparse radial samples by using iterative refinement procedures. In: *Proceeding of the 13th Annual Meeting of ISMRM*, p. 696 (2006)
22. Hestenes, M.R.: Multiplier and gradient methods. *J. Optim. Theor. Appl.* **4**, 303–320 (1969)
23. Hyder, M., Mahata, K.: An approximate L^0 norm minimization algorithm for compressed sensing. In: *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3365–3368 (2009)
24. Jacobson, M., Fessler, J.: An expanded theoretical treatment of iteration-dependent majorize-minimize algorithms. *IEEE Trans. Image Proc.* **16**, 2411–2422 (2007)
25. Jacobson, M., Fessler, J.: Properties of MM algorithms on convex feasible sets: extended version. Technical Report 353, Communications and Signal Processing Laboratory, Department of EECS, University of Michigan, Ann Arbor, MI, 48109–2122 (2004)
26. Koko, J., Jehan-Besson, S.: An augmented Lagrangian method for $TVg+L^1$ -norm minimization. Research Report LIMOS/RR-09-07 (2009)
27. Lusting, M., Donoho, D., Pauly, J.M.: Sparse MRI: the application of compressed sensing for rapid MR imaging. *Magn. Reson. Med.* **58**, 1182–1195 (2007)

28. Lusting, M., Lee, J.H., Donoho, D.L., Pauly, J.M.: Faster imaging with randomly perturbed undersampled spirals and L1 reconstruction. In: Proceedings of the ISMRM (2005)
29. Mohimani, G.H., Babaic-Zadeh, M., Jutten, C.: Fast sparse representation based on smoothed ℓ^0 norm. In: Davies, M.E., James, C.J., Abdallah, S.A., Plumbley, M.D. (eds.) ICA 2007. LNCS, vol. 4666, pp. 389–396. Springer, Heidelberg (2007)
30. Mancera, L., Portilla, J.: L0-norm-based sparse representation through alternate projections. In: International Conference on Image Processing, pp. 2089–2092. IEEE (2006)
31. Nikolova, M., Chan, R.H.: The equivalence of the half-quadratic minimization and the gradient linearization iteration. IEEE Trans. Image Proc. **16**, 5–18 (2007)
32. Osher, S., Mao, Y., Dong, B., Yin, W.: Fast linearized Bregman iterations for compressed sensing and sparse denoising. In: UCLA CAM Report, 08–37 (2008)
33. Powell, M.J.D.: A method for nonlinear constraints in minimization problems. In: Fletcher, R. (ed.) Optimization, pp. 283–298. Academic Press, New York (1972)
34. Qu, X.B., Cao, X., Guo, D., Hu, C., Chen, Z.: Compressed sensing MRI with combined sparsifying transforms and smoothed L0 norm minimization. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 626–629 (2010)
35. Rao, B.D., Kreutz-Delgado, K.: An affine scaling methodology for best basis selection. IEEE Trans. Signal Process **47**, 187–200 (1999)
36. Rockafellar, R.T.: A dual approach to solving nonlinear programming problems by unconstrained optimization. Math. Program. **5**, 354–373 (1973)
37. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D. **60**, 259–268 (1992)
38. Saab, R., Chartrand, R., Yilmaz, Ö.: Stable sparse approximations via non-convex optimization. In: IEEE International Conference on Acoustics, Speech, and Signal Processing (2008)
39. Vogel, C.R., Oman, M.E.: Fast, robust total variation-based reconstruction of noisy blurred images. IEEE Trans. Image Proc. **7**, 813–824 (1998)
40. Weiszfeld, E.: Sur le point pour lequel la somme des distances de n points donnés est minimum. Tôhoku Math. J. **43**, 355–386 (1937)
41. Wu, C.L., Zhang, J.Y., Tai, X.C.: Augmented Lagrangian method for total variation restoration with non-quadratic fidelity. Inverse Prob. Imaging **5**, 237–261 (2010)
42. Xu, Z.B.: Data modeling: visual psychology approach and $l_{1/2}$ regularization theory. In: Proceedings of the International Congress of Mathematicians, Hyderabad, India (2010)
43. Xu, Z.B., Chang, X.Y., Xu, F.M., Zhang, H.: $L_{1/2}$ regularization: an iterative half thresholding algorithm. Technique Report (2010)
44. Yang, J.F., Zhang, Y., Yin, W.T.: An efficient TVL1 algorithm for deblurring multichannel images corrupted by impulsive noise. SIAM J. Sci. Comput. **31**, 2842–2865 (2009)
45. Yin, W.T., Goldfarb, D., Osher, S.: Image cartoon-texture decomposition and feature selection using the total variation regularized L^1 functional. In: Paragios, N., Faugeras, O., Chan, T., Schnörr, C. (eds.) VLSM 2005. LNCS, vol. 3752, pp. 73–84. Springer, Heidelberg (2005)

46. Yin, W.T., Goldfarb, D., Osher, S.: The total variation regularized L^1 model for multiscale decomposition. *Multis. Model. Simul.* **6**, 190–211 (2006)
47. Yin, W.T., Osher, S., Goldfarb, D., Darbon, J.: Bregman iteration algorithms for ℓ_1 -minimization with applications to compressed sensing. *SIAM J. Imaging Sci.* **1**, 142–168 (2008)