Inverse Problems and Imaging

doi:10.3934/ipi.2024041

MINIMIZING QUOTIENT REGULARIZATION MODEL

Chao Wang[1], Jean-Francois Aujol[2],
Guy Gilboa[3] and Yifei Lou[*4]

[1]Southern University of Science and Technology, Shenzhen 518005, Guangdong Province, China

[2]University of Bordeaux, Bordeaux INP, CNRS, IMB, UMR 5251, F-33400 Talence, France

[3]Technion, Israel Institute of Technology, Haifa, Israel

[4]University of North Carolina at Chapel Hill, NC, 27599 USA

(Communicated by Sung Ha Kang)

Abstract. Quotient regularization models (QRMs) are a class of powerful regularization techniques that have gained considerable attention in recent years, due to their ability to handle complex and highly nonlinear data sets. However, the nonconvex nature of QRM poses a significant challenge in finding its optimal solution. We are interested in scenarios where both the numerator and the denominator of QRM are absolutely one-homogeneous functions, which is widely applicable in the fields of signal processing and image processing. In this paper, we utilize a gradient flow to minimize such QRM in combination with a quadratic data fidelity term. Our scheme involves solving a convex problem iteratively. The convergence analysis is conducted on a modified scheme in a continuous formulation, showing the convergence to a stationary point. Numerical experiments demonstrate the effectiveness of the proposed algorithm in terms of accuracy, outperforming the state-of-the-art QRM solvers.

1. **Introduction.** In this paper, we consider a generalized quotient regularization model (QRM) with a least-squares data fidelity term weighted by a positive constant λ, i.e.,

$$\min_{u\in\Omega} \frac{J(u)}{H(u)} + \frac{\lambda}{2}\|Au-f\|_2^2, \tag{1}$$

where both functionals $J(\cdot), H(\cdot)$ are proper, convex, lower semi-continuous (lsc), and absolutely one-homogeneous on a proper domain $\Omega \subset \mathbb{R}^n$. An absolutely one homogeneous functional $F: u\in\Omega \to \mathbb{R}$ satisfies $F(\alpha u) = |\alpha| F(u), \forall \alpha\in\mathbb{R}, u\in\Omega$.

2020 *Mathematics Subject Classification.* Primary: 65F22, 65Y04; Secondary: 52A41, 49N45.

Key words and phrases. Quotient regularization, gradient flow, fractional programming.

The first author is supported by the Natural Science Foundation of China 12201286, the Shenzhen Science and Technology Program 20231115165836001, and Guangdong Basic and Applied Research Foundation 2024A1515012347. The second and third authors acknowledge the support of the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No777826. The third author acknowledges support by ISF grant 534/19 and ISF 1472/23. The last author is supported by NSF CAREER award 2414705. This work was initiated while J-F. Aujol and Y. Lou were visiting the Mathematical Department of UCLA.

*Corresponding author: Yifei Lou.

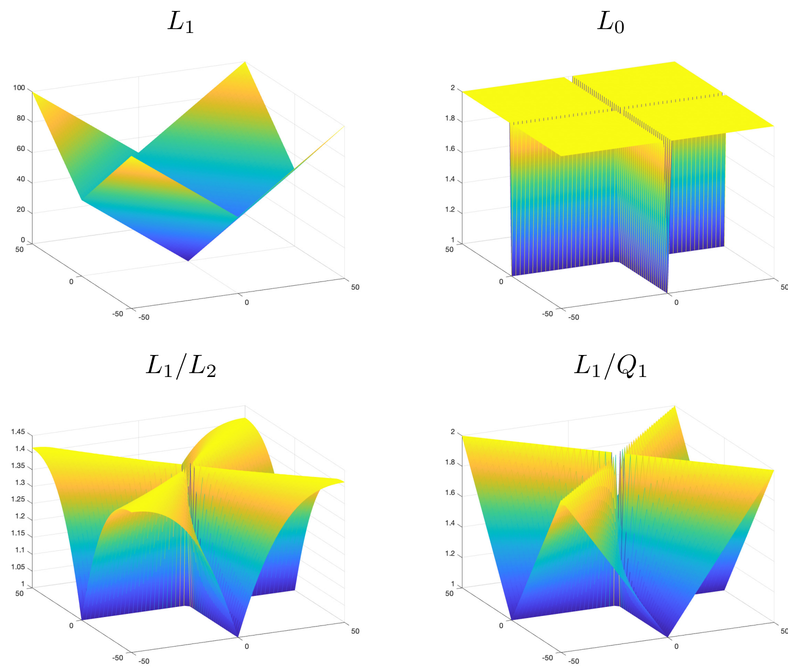

FIGURE 1. A 2D illustration of L_1/L_2 and L_1/Q_1 that give a better approximation to the L_0 norm with a comparison to the convex L_1 norm.

It can be proven that a convex and absolutely one homogeneous function is always non-negative with $J(0) = 0$. We further assume by convention $\frac{J(0)}{H(0)} := 0$, thus it is well-defined at 0. The least-squares misfit between the linear operator A and the measurements f is a standard data fidelity term when the noise $Au - f$ is subject to the Gaussian distribution. For other noise types, the data fidelity term is formulated differently. We give three specific signal and image processing examples that fit into our general model (1).

Example 1. *(L_1/L_2 sparse signal recovery).* The ratio of the L_1 and L_2 norms was prompted as a scale-invariant surrogate to the L_0 norm for sparse signal recovery [15, 17]. Defining $J(0)/H(0) = 0$ aligns with the L_0 norm of the zero vector. Recently, a constrained minimization problem was formulated, i.e.,

$$\min_{u \in \mathbb{R}^n} \frac{\|u\|_1}{\|u\|_2} \quad \text{s.t.} \quad Au = f,$$

for the ease of analyzing the theoretical properties of the L_1/L_2 model [25, 32] as well as deriving a numerical algorithm [29]. Here we adopt the unconstrained formulation [26] that is aligned with our generalized model (1)

$$\min_{u \in \mathbb{R}^n} \frac{\|u\|_1}{\|u\|_2} + \frac{\lambda}{2}\|Au - f\|_2^2. \tag{2}$$

A more general ratio of L_p over L_q (quasi-)norms for $p \in (0, 2)$ and $q \geq 2$ was explored in [9].

Example 2. *(L_1/Q_K sparse signal recovery).* Motivated by the truncated L_1 regularization (a.k.a partial sum) [16, 23] and the L_1/L_2 model, we consider the ratio of the L_1 norm and K-largest truncated L_2 norm as a sparsity-promoting regularization with a given integer K. When $K = 1$, it becomes the L_1 norm over the infinity norm [10, 30]. For $K = n$ (the ambient dimension of u), L_1/Q_K is equivalent to L_1/L_2. In Figure 1, we use a 2D example to illustrate that both L_1/L_2 and L_1/Q_K can promote sparsity by approximating the L_0 norm. Both ratios give a better approximation to the L_0 norm compared to the convex L_1 norm, which is largely attributed to the scale-invariant property of the L_0 norm and the two ratio models. Define $J(u) = \|u\|_1$ and $H(u)$ as the truncated L_2 norm of the K-largest absolute values of entries, denoted as $\|u\|_{(K)}$. As both $J(\cdot)$ and $H(\cdot)$ are absolutely one-homogeneous, we consider the following problem

$$\min_{u\in\mathbb{R}^n} \frac{\|u\|_1}{\|u\|_{(K)}} + \frac{\lambda}{2}\|Au - f\|_2^2, \tag{3}$$

as a special case of (1).

Example 3. *(L_1/L_2 on the gradient for image recovery).* In [27, 28], the L_1/L_2 functional was applied to the image gradient and combined with the least-squares term,

$$\min_{u} \frac{\|\nabla u\|_1}{\|\nabla u\|_2} + \frac{\lambda}{2}\|Au - f\|_2^2. \tag{4}$$

Specifically, Wang et al. [28] demonstrated that this model (4) yields significant improvements in a limited-angle CT reconstruction problem. With an additional H^1-semi norm to (4) for smoothing, a segmentation model was proposed in [31]. A modification of replacing the gradient operator ∇ in (4) by a nonnegative diagonal matrix was explored in [19] for electrical capacitance tomography.

Without the data fitting term, our model (1) reduces to Rayleigh quotient problems, defined by

$$\min_{u\in\Omega} R(u) := \frac{J(u)}{H(u)}. \tag{5}$$

The classic Rayleigh quotient problem in linear eigenvalue analysis [14] is defined by

$$\min_{u\in\mathbb{R}^n} \frac{\langle u, Lu\rangle}{\|u\|_2^2}, \tag{6}$$

with a symmetric matrix $L \in \mathbb{R}^{n\times n}$. Any critical point of (6) is an eigenvector of the matrix L. One can replace the linear mapping Lu in (6) by a nonlinear function, thus leading to a nonlinear eigenproblem. Nossek and Gilboa [22] proposed a continuous flow that minimizes (5) when $J(\cdot)$ is absolutely one homogeneous and $H(\cdot)$ is the square L_2 norm. The convergence proof was later provided in [1]. Under the same setting, a nonlinear power method was proposed in [7] with connections to proximal operators and neural networks. For the case when J is the total variation (TV) and H is the L_1 norm, the Rayleigh quotient (5) approximates the Cheeger cut problem [13, 6]. The quotient minimization (5) also appears in learning parameterized regularizations [4] and filter functions [3].

In this paper, we propose a novel scheme to minimize the general model (1) based on a gradient descent flow for the Rayleigh quotient minimization [11]. We then apply the proposed algorithm to the three specific examples (L_1/L_2, L_1/Q_K, and L_1/L_2 on the gradient). In each case, our algorithm requires minimizing an

L_1-regularized subproblem, which can be solved efficiently using the alternating direction method of the multiplier (ADMM) [5, 12]. Our analysis for the proposed algorithm is towards a slightly modified scheme. We establish a subsequential convergence of the modified scheme under the uniform boundedness of the sequence. With some additional assumptions, the uniform bound can be proven using a continuous flow formulation. In experiments, we demonstrate the efficiency of the proposed algorithm over the relevant methods in the literature. In summary, the novelties of this paper are threefold:

1. We consider a general model (1) that combines the Rayleigh quotient as a regularization with a data fidelity term. Our model has a variety of applications, especially in signal and image reconstruction.
2. We propose a unified algorithm with numerical insights on convergence and the solution's boundedness.
3. Our approach can be adapted to three case studies: (2), (3), and (4). In each case, the proposed scheme outperforms the relevant algorithms in the literature in terms of accuracy.

The rest of the paper is organized as follows. Section 2 describes the proposed algorithms in detail, including numerical formulation and specific closed-form solutions for the three case studies. We provide mathematical analysis on the numerical scheme in Section 3. Extensive experiments are conducted in Section 4 for applications in signal and image recovery. Finally, conclusions and future works are given in Section 5.

2. Proposed algorithms.

Recall that we aim at the minimization problem

$$\min_u G(u) := R(u) + \frac{\lambda}{2}\|Au - f\|_2^2, \tag{7}$$

with $R(u) = J(u)/H(u)$.

Theorem 2.1. *Suppose A is an under-determined matrix, $f \in \mathrm{Im}(A)\backslash\{0\}$, and $R(\cdot)$ has an upper bound, i.e., $R(u) \le M$. For a sufficiently large parameter λ, the optimal solution of (7) can not be 0.*

Proof. As A is an under-determined matrix and $f \in \mathrm{Im}(A)$, there exist infinitely many solutions satisfying $Au = f$, among which we denote $\hat{u}$ to be the least norm solution, that is,

$$\hat{u} = \arg\min_u \|u\|_2 \quad \text{such that} \quad Au = f.$$

It is straightforward that $G(\hat{u}) = R(\hat{u}) \le M$ and $G(0) = \frac{J(0)}{H(0)} + \frac{\lambda}{2}\|f\|_2^2$. If $\lambda > \frac{2M}{\|f\|_2^2}$, we have $G(\hat{u}) < G(0)$, which implies that 0 cannot be the global solution to (7). □

Remark: Note that all the examples listed in the introduction section satisfy the boundedness assumption of $R(\cdot)$. Taking L_1/L_2 for an example, one has $\frac{\|u\|_1}{\|u\|_2} \le \sqrt{n}$ for $u \in \mathbb{R}^n$.

One classic method to minimize $G(u)$ is by using a gradient descent flow, i.e.,

$$u_t \in -\partial G(u), \tag{8}$$

where ∂ indicates the subgradient [21]. We consider the subgradient ∂ here as $J(\cdot), H(\cdot)$ are not necessarily differentiable. The subgradient of G can be expressed as

$$\begin{aligned}\partial G(u) &= \frac{H(u)p - J(u)q}{H^2(u)} + \lambda A^T(Ax - f) \\ &= \frac{p - R(u)q}{H(u)} + \lambda A^T(Au - f),\end{aligned} \tag{9}$$

where $q \in \partial H(u), p \in \partial J(u)$. Plugging the gradient expression (9) into the flow (8) yields

$$u_t = \frac{R(u)}{H(u)}q - \frac{p}{H(u)} - \lambda A^T(Au - f),$$

which can be discretized by the iteration count k,

$$\frac{u^{k+1} - u^k}{dt} = \frac{R(u^k)}{H(u^k)}q^k - \frac{p^{k+1}}{H(u^k)} - \lambda A^T(Au^{k+1} - f). \tag{10}$$

Note that we consider a semi-implicit scheme in (10). A fully explicit discretization requires the subgradient of $J(\cdot)$, which is a set and hence cannot be uniquely determined. If we implicitly discretize the denominator $H(u)$ by $H(u^{k+1})$, the resulting optimization problem does not have a closed-form solution in each iteration, thus computationally inefficient. The update of u^{k+1} is obtained by the following optimization problem,

$$u^{k+1} = \arg\min_u \left\{ \frac{\beta}{2}\|u - u^k\|_2^2 - \frac{R(u^k)}{H(u^k)}\langle q^k, u\rangle + \frac{J(u)}{H(u^k)} + \frac{\lambda}{2}\|Au - f\|_2^2 \right\}, \tag{11}$$

for $\beta = \frac{1}{dt}$, whose optimality condition coincides with (10). Since $J(\cdot)$ is assumed to be convex, the objective function of (11) is strictly convex, ensuring the existence and uniqueness of the optimal solution u^{k+1}.

In what follows, we describe the detailed algorithms for L_1/L_2 and L_1/Q_K in Section 2.1 as well as the gradient model (4) in Section 2.2, all based on the general scheme (11).

2.1. **Quotient regularization for sparse signal recovery.** For $H(u) = \|u\|_2$ and $q \in \partial H(u)$, we get $q = \frac{u}{\|u\|_2}$ if $u \neq 0$; otherwise q is a vector with each element bounded by $[-1, 1]$. As $J(u) = \|u\|_1$, the minimization problem (11) at the kth iteration becomes

$$u^{k+1} = \arg\min_u \left\{ \frac{\beta}{2}\|u - u^k\|_2^2 - \langle h^k, u\rangle + \frac{\|u\|_1}{\|u^k\|_2} + \frac{\lambda}{2}\|Au - f\|_2^2 \right\}, \tag{12}$$

where $h^k = \frac{R(u^k)}{H(u^k)}q^k = \frac{\|u^k\|_1}{\|u^k\|_2^3}u^k$. Note that the scheme (10) is not well-defined if $u^k = 0$, while this turns out not to be restrictive in, as $u^k = 0$ never occurs in our experiments. On the theoretical side, we know from Theorem 2.1 that 0 cannot be the minimizer of the objective function in the minimization problem (7).

To solve for the L_1-regularized minimization (12), we introduce an auxiliary variable y and consider an equivalent problem

$$\min_{u,y} \frac{\beta}{2}\|y - u^k\|_2^2 - \langle h^k, y\rangle + \frac{\|u\|_1}{\|u^k\|_2} + \frac{\lambda}{2}\|Ay - f\|_2^2 \quad \text{s.t.} \quad u = y. \tag{13}$$

The corresponding augmented Lagrangian function is expressed as,

$$\mathcal{L}_k(u, y; \eta) = \frac{\beta}{2}\|y - u^k\|_2^2 - \langle h^k, y\rangle + \frac{\|u\|_1}{\|u^k\|_2} + \frac{\lambda}{2}\|Ay - f\|_2^2 + \frac{\rho}{2}\|u - y + \eta\|_2^2, \tag{14}$$

where η is a dual variable and ρ is a positive parameter. Then ADMM iterates as follows

$$\begin{cases} u_{j+1} = \arg\min_u \mathcal{L}_k(u, y_j; \eta_j) \\ y_{j+1} = \arg\min_y \mathcal{L}_k(u_{j+1}, y; \eta_j) \\ \eta_{j+1} = \eta_j + u_{j+1} - y_{j+1}, \end{cases} \tag{15}$$

where the subscript j represents the inner loop index, as opposed to the superscript k for outer iterations (11). The u-subproblem has a closed-form solution:

$$u_{j+1} = \text{shrink}\left(y_j - \eta_j,\ \frac{1}{\rho\|u^k\|_2}\right).$$

The update of y follows the computation of gradient of $\mathcal{L}_k$ with respect to y:

$$y_{j+1} = (\lambda A^T A + (\beta+\rho)I)^{-1}(\beta u^k + h^k + \lambda A^T f + \rho(u_{j+1} + \eta_j)), \tag{16}$$

which involves solving a large linear system. In the case of sparse signal recovery when the system matrix $A \in \mathbb{R}^{m\times n}$ is under-determined, i.e., $m \ll n$, the closed-form solution of y can be written in an efficient way by the Sherman–Morrison–Woodbury formula:

$$y_{j+1} = \left[\kappa I - \lambda\kappa^2 A^T \left(I + \lambda\kappa AA^T\right)^{-1} A\right]\left[\beta u^k + h^k + \lambda A^T f + \rho(u_{j+1} + \eta_j)\right].$$

where $\kappa = 1/(\beta+\rho)$ and the matrix $I + \lambda\kappa AA^T$ is in m-by-m size, which is much smaller than inverting an $n \times n$ matrix in (16). Using the Choleskey decomposition for $I + \lambda\kappa AA^T$ can further accelerate the computation.

For the L_1/Q_K model (3), $H(u) = \|u\|_{(K)}$ and its subgradient is a random vector bounded by $[-1, 1]$ if $u = 0$. In addition, when $u \neq 0$, one has

$$q_i = \begin{cases} \frac{u_i}{\|u\|_{(K)}} & i \in \Omega_K(u) \\ 0 & \text{Otherwise,} \end{cases}$$

where $q \in \partial H(u)$ and $\Omega_K(u)$ is the index set of the K-largest absolute values of u. As a result, the algorithm for the L_1/Q_K model (3) is the same as (12) except that $h^k = \frac{\|u^k\|_1}{\|u^k\|^3_{(K)}} v^k$ with

$$v_i^k = \begin{cases} u_i^k & i \in \Omega_K(u^k) \\ 0 & \text{Otherwise.} \end{cases} \tag{17}$$

Algorithm 1 presents a unified scheme that minimizes the L_1/L_2 and L_1/Q_K models with the least-squares fit.

2.2. **Quotient regularization for image recovery.** When $J(u) = \|\nabla u\|_1$ and $H(u) = \|\nabla u\|_2$, we get $q = \frac{-\Delta u}{\|\nabla u\|_2}$ if $\nabla u \neq 0$; otherwise q is a vector with each element bounded by $[-1, 1]$. Hence the minimization problem (11) in the k-iteration becomes

$$u^{k+1} = \arg\min_u \left\{\frac{\beta}{2}\|u - u^k\|_2^2 - \langle h^k, u\rangle + \frac{\|\nabla u\|_1}{\|\nabla u^k\|_2} + \frac{\lambda}{2}\|Au - f\|_2^2\right\}, \tag{18}$$

where $h^k = \frac{\|\nabla u^k\|_1}{\|\nabla u^k\|_2^3}\Delta u^k$. The subproblem (18) is a TV regularization with additional linear and least-squares terms, which can be solved by ADMM. In particular, we introduce one auxiliary variable $y = \nabla u$ upon convergence, and formulate the augmented Lagrangian function corresponding to (12) as,

$$\mathcal{L}_k(u, y; \eta) = \frac{\beta}{2}\|u - u^k\|_2^2 - \langle h^k, u\rangle + \frac{\|y\|_1}{\|\nabla u^k\|_2} + \frac{\lambda}{2}\|Au - f\|_2^2 + \frac{\rho}{2}\|\nabla u - y + \eta\|_2^2, \tag{19}$$

Algorithm 1: Proposed algorithm for the models of L_1/L_2 and L_1/Q_K.

Input: a linear operator A, observed data f
Parameters: ρ, λ, β, $\kappa = 1/(\beta+\rho)$, kMax, jMax, $\epsilon \in \mathbb{R}$, and K for the L_1/Q_K model
Initialize: $\eta = 0, k, j = 0$ and u^0
while $k <$ kMax or $\|u^k - u^{k-1}\|_2/\|u^k\|_2 > \epsilon$ **do**
  while $j <$ jMax or $\|u_j - u_{j-1}\|_2/\|u_j\|_2 > \epsilon$ **do**
    $u_{j+1} = \text{shrink}\left(y_j - \eta_j,\ \frac{1}{\rho\|u^k\|_2}\right)$
    $y_{j+1} = \left[\kappa I - \lambda\kappa^2 A^T\left(I + \lambda\kappa AA^T\right)^{-1} A\right]\left[\beta u^k + h^k + \lambda A^T f + \rho(u_{j+1} + \eta_j)\right]$
    $\eta_{j+1} = \eta_j + u_{j+1} - y_{j+1}$
    Assign j by $j+1$
  end while
  Set u^{k+1} as u_j
  Update h^{k+1} by $h^{k+1} = \begin{cases} \frac{\|u^k\|_1}{\|u^k\|_2^3} u^k & \text{for } L_1/L_2 \\ \frac{\|u^k\|_1}{\|u^k\|_{(K)}^3} v^k & \text{for } L_1/Q_K \end{cases}$
  Assign k and j by $k+1$ and 0, respectively
end while
return $u^* = u^k$

where η is a dual variable and ρ is a positive parameter. Then ADMM iterates as follows

$$\begin{cases} u_{j+1} = \arg\min_u \mathcal{L}_k(u, y_j; \eta_j) \\ y_{j+1} = \arg\min_y \mathcal{L}_k(u_{j+1}, y; \eta_j) \\ \eta_{j+1} = \eta_j + \nabla u_{j+1} - y_{j+1}. \end{cases} \tag{20}$$

Taking the derivative of (20) with respect to u, we get

$$u_{j+1} = (\lambda A^T A - \rho\Delta + \beta I)^{-1}(\lambda A^T f + \beta u^k + \rho(y_j - \eta_j) + h^k). \tag{21}$$

For image deblurring or the MRI reconstruction, the inverse in the u-update (21) can be computed efficiently via the fast Fourier transform.

The update for the variable y is given by

$$y_{j+1} = \text{shrink}\left(\nabla u_{j+1} + \eta_j, \frac{1}{\rho\|\nabla u^k\|_2}\right).$$

We summarize the proposed algorithm for minimizing the L_1/L_2 on the gradient in Algorithm 2.

3. Mathematical analysis.

This section is split into two parts. In Section 3.1, we prove the convergence of a modified scheme to the solution of the quotient model (1). To do so, we need a technical uniform bound assumption, which is justified in Section 3.2 based on a continuous formulation of the scheme.

3.1. Convergence of the scheme.

We first show that a fully implicit version of the numerical scheme (11) converges (up to a subsequence) to a solution of our original problem (1) under a reasonable uniform bound assumption. In our analysis, we make use of Lemmas 3.1-3.2 that are related to the subdifferential of one homogeneous convex function (see for instance [8, 7, 11]).

Algorithm 2: Proposed algorithm for the L_1/L_2 model on the gradient.

Input: a linear operator A, observed data f,
Parameters: ρ, λ, β, kMax, jMax, and $\epsilon \in \mathbb{R}$
Initialize: $\eta = 0, k, j = 0$ and u^0
while $k <$ kMax or $\|u^k - u^{k-1}\|_2/\|u^k\|_2 > \epsilon$ **do**
  while $j <$ jMax or $\|u_j - u_{j-1}\|_2/\|u_j\|_2 > \epsilon$ **do**
    $u_{j+1} = (\lambda A^T A - \rho\Delta + \beta I)^{-1}(\lambda A^T f + \beta u^k + \rho(y_j - \eta_j) + h^k)$
    $y_{j+1} = \text{shrink}\left(\nabla u_{j+1} + \eta_j, \frac{1}{\rho\|\nabla u^k\|_2}\right)$
    $\eta_{j+1} = \eta_j + u_{j+1} - y_{j+1}$
    Assign j by $j+1$
  end while
  Set u^{k+1} as u_j
  Update h^{k+1} by $h^k = \frac{\|\nabla u^k\|_1}{\|\nabla u^k\|_2^3}\Delta u^k$
  Assign k and j by $k+1$ and 0, respectively
end while
return $u^* = u^k$

Lemma 3.1. *For a convex one homogeneous function J, we refer to p as a subdifferential of $J(u)$, i.e., $p \in \partial J(u)$, if and only if the following two assertions hold*

(i) $J(u) = \langle p, u\rangle$.
(ii) $J(v) \geq \langle p, v\rangle, \ \forall v$.

Lemma 3.2. *If J is a convex one homogeneous function and $\Omega \subset \mathbb{R}^n$, then there exists $L_J > 0$ such that $\|p\|_2 \leq L_J$ for all $p \in \partial J(0)$.*

It follows from Lemma 3.1 that $\partial J(u) \subset \partial J(0), \forall u$. Hence a direct consequence of Lemma 3.2 is the following (see e.g. [1]):

Lemma 3.3. *If J is a convex one homogeneous function and $\Omega \subset \mathbb{R}^n$, then there exists $L_J > 0$ such that if $u \in \Omega$ and $p \in \partial J(u)$, then $\|p\|_2 \leq L_J$.*

Fully implicit scheme: We recall that the sequence $\{u^k\}$ is defined by Equation (11). In fact, we are going to analyze a slightly different scheme, which is referred to as a fully implicit scheme,

$$u^{k+1} = \arg\min_u \left\{ \frac{\beta}{2}\|u - u^k\|_2^2 - \frac{R(u^k)}{H(u)}\langle q^k, u\rangle + R(u) + \frac{\lambda}{2}\|Au - f\|_2^2 \right\}, \tag{22}$$

where the term $\frac{1}{H(u^k)}$ in (11) has been replaced by $\frac{1}{H(u)}$. We remark that the numerical scheme (11) is much easier to handle with the term $\frac{1}{H(u^k)}$, but the mathematical analysis of (22) happens to be much easier with $\frac{1}{H(u)}$. We establish in Theorem 3.4 that $\|u^{k+1} - u^k\|_2 \to 0$ when $k \to +\infty$.

Theorem 3.4. *For absolutely one-homogeneous functionals $J(\cdot), H(\cdot)$, if the sequence $\{u^k\}$ is defined by (22), then $\|u^{k+1} - u^k\|_2 \to 0$ as $k \to +\infty$.*

Proof. Define the objective function in (22) by

$$F(u) := \frac{\beta}{2}\|u - u^k\|_2^2 - \frac{R(u^k)}{H(u)}\langle q^k, u\rangle + R(u) + \frac{\lambda}{2}\|Au - f\|_2^2. \tag{23}$$

It is straightforward that

$$F(u^k) = -\frac{R(u^k)}{H(u^k)}\langle q^k, u^k\rangle + R(u^k) + \frac{\lambda}{2}\|Au^k - f\|_2^2. \tag{24}$$

Since H is absolutely one-homogeneous, we use Lemma 3.1 (i) to obtain $\langle q^k, u^k\rangle = H(u^k)$, thus leading to

$$F(u^k) = -R(u^k) + R(u^k) + \frac{\lambda}{2}\|Au^k - f\|_2^2 = \frac{\lambda}{2}\|Au^k - f\|_2^2. \tag{25}$$

It follows from Lemma 3.1 (ii) that $H(u^{k+1}) \geq \langle q^k, u^{k+1}\rangle$, which implies that

$$\begin{aligned}
F(u^{k+1}) &= \frac{\beta}{2}\|u^{k+1} - u^k\|_2^2 - \frac{R(u^k)}{H(u^{k+1})}\langle q^k, u^{k+1}\rangle + R(u^{k+1}) + \frac{\lambda}{2}\|Au^{k+1} - f\|_2^2 \\
&\geq \frac{\beta}{2}\|u^{k+1} - u^k\|_2^2 - \frac{R(u^k)}{H(u^{k+1})}H(u^{k+1}) + R(u^{k+1}) + \frac{\lambda}{2}\|Au^{k+1} - f\|_2^2 \\
&= \frac{\beta}{2}\|u^{k+1} - u^k\|_2^2 - R(u^k) + R(u^{k+1}) + \frac{\lambda}{2}\|Au^{k+1} - f\|_2^2.
\end{aligned}$$

We use the fact that $F(u^{k+1}) \leq F(u^k)$ to deduce:

$$\frac{\beta}{2}\|u^{k+1} - u^k\|_2^2 - R(u^k) + R(u^{k+1}) + \frac{\lambda}{2}\|Au^{k+1} - f\|_2^2 \leq \frac{\lambda}{2}\|Au^k - f\|_2^2. \tag{26}$$

Summing from 1 to N, we get:

$$\begin{aligned}
\frac{\beta}{2}\sum_{k=1}^{N}\|u^{k+1} - u^k\|_2^2 &\leq R(u^1) - R(u^{N+1}) + \frac{\lambda}{2}\Big(\|Au^1 - f\|_2^2 - \|Au^{N+1} - f\|_2^2\Big) \\
&\leq R(u^1) + \frac{\lambda}{2}\|Au^1 - f\|_2^2,
\end{aligned}$$

due to $R(u) \geq 0$ and $\|Au - f\|_2^2 \geq 0$ for any u. Let $N \to \infty$, we obtain that $\sum_{k=1}^{\infty}\|u^{k+1} - u^k\|_2^2$ is bounded, which implies that $\|u^{k+1} - u^k\|_2^2 \to 0$. □

Now that we have proven that $\|u^{k+1} - u^k\|_2 \to 0$ as $k \to +\infty$, we are going to be able to pass the limit up to a subsequence in the optimality condition of (1).

Theorem 3.5. *For absolutely one-homogeneous functionals $J(\cdot), H(\cdot)$, if the sequence $\{u^k\}$ defined by (22) is bounded, and if $q^k - q^{k+1} \to 0$, then there exists a subsequence of $\{(u^k, p^k, q^k)\}$ that converges to (u^*, p^*, q^*). Moreover, we have*

$$p^* \in \partial J(u^*),\ q^* \in \partial H(u^*), \quad \textit{and} \quad 0 = \lambda\, A^T(Au^* - f) + \frac{p^* - R(u^*)q^*}{H(u^*)}. \tag{27}$$

Proof. We assume that u^k is bounded, and we know from Lemma 3.3 that the subgradients p^k and q^k are also bounded. Thus, there exists (u^*, p^*, q^*) such that up to a subsequence, $(u^k, p^k, q^k) \to (u^*, p^*, q^*)$. The optimality condition for (22) can be written as:

$$\begin{aligned}
0 \ = \ & \beta(u^{k+1} - u^k) + \lambda\, A^T(Au^{k+1} - f) \\
& -\frac{R(u^k)q^k}{H(u^{k+1})} + \frac{R(u^k)\langle q^k, u^{k+1}\rangle q^{k+1}}{(H(u^{k+1}))^2} + \frac{p^{k+1} - R(u^{k+1})q^{k+1}}{H(u^{k+1})}.
\end{aligned}$$

Thanks to Theorem 3.4, we can pass to the limit in this last equation to get:

$$0 = \lambda\, A^T(Au^* - f) + \frac{p^* - R(u^*)q^*}{H(u^*)}, \tag{28}$$

where we use Lemma 3.1 for $\langle q^*, u^*\rangle = H(u^*)$. Note that (28) is the original optimality condition $0 \in \partial G(u^*)$, i.e., Equation (9) for the optimization problem (1). ☐

Remark: The assumption $q^k - q^{k+1} \to 0$ in Theorem 3.5 could be removed under some specific choice of the subgradients q^k. Indeed, since from Theorem 3.4, we have $u^k - u^{k+1} \to 0$, and if we assume dom$(H)=\mathbb{R}^n$, which is the case for all the examples studied in the paper, then distance $(\partial H(u^k), \partial H(u^{k+1})) \to 0$ (since the graph of ∂H is closed, see e.g. Proposition 16.26 in [2]). Hence with a proper choice of the subgradients, we could have $q^k - q^{k+1} \to 0$.

3.2. **Uniform boundedness of the sequence $\{u^k\}$.** The goal of this subsection is to explain why the technical assumption on the uniform boundedness of the sequence $\{u^k\}$ is reasonable for Theorem 3.5. Instead of dealing with the discrete sequence $\{u^k\}$, we conduct our analysis in a continuous setting, which enables us to have tractable computations. In particular, we consider a differentiable function u of the continuous flow, that is, $u_t \in -\partial G(u)$ in (8). Notice that u^k defined by (22) can be seen as a discretized version of u.

We show in Theorem 3.6 that a mapping of $t \mapsto \|u\|_2^2$ is a non-increasing function as long as $\|Au\|_2 \geq \|f\|_2$.

Theorem 3.6. *Suppose $u(t)$ is a differentiable function with respect to the time t that satisfies the flow (8), i.e., there exists $p \in \partial J$ and $q \in \partial H$ such that the following equality*

$$u_t = -\lambda\, A^T(Au - f) - \frac{p - R(u)q}{H(u)}, \tag{29}$$

holds. If $\|Au\|_2 \geq \|f\|_2$, then

$$\frac{d}{dt}\left(\|u\|_2^2\right) \leq 0. \tag{30}$$

Proof. Simple calculations lead to

$$\begin{aligned}
\frac{d}{dt}\left(\frac{1}{2}\|u\|_2^2\right) &= \langle u, u_t\rangle \\
&= -\lambda\langle Au - f, Au\rangle - \frac{\langle p, u\rangle - R(u)\langle q, u\rangle}{H(u)} \\
&= -\lambda\langle Au - f, Au\rangle - \frac{J(u) - R(u)H(u)}{H(u)} \\
&= -\lambda\left(\|Au\|_2^2 - \langle f, Au\rangle\right),
\end{aligned}$$

where we use Lemma 3.1 with $p \in \partial J(u)$ and $q \in \partial H(u)$. It further follows from the Cauchy-Schwartz inequality that

$$\frac{d}{dt}\left(\|u\|_2^2\right) \leq \lambda\|Au\|_2\left(\|f\|_2 - \|Au\|_2\right). \tag{31}$$

Consequently, if $\|Au\|_2 \geq \|f\|_2$, then $\|u\|_2^2$ is a non-increasing function. ☐

We give a numerical verification of Theorem 3.6 in Figure 3. A direct consequence of Theorem 3.6 leads to the following two corollaries.

Corollary 3.7. *If A is coercive, i.e., there exists $c > 0$ such that $\|Au\|_2 \geq c\|u\|_2, \forall u$, then any function u satisfying the flow (29) is uniformly bounded.*

Proof. Following (31) with the coercivity assumption on A, we get

$$\frac{d}{dt}\left(\|u\|_2^2\right) \leq \lambda\|Au\|_2\left(\|f\|_2 - c\|u\|_2\right). \qquad (32)$$

Now we have two cases to consider

- If $\|u\|_2 \leq \frac{\|f\|_2}{c}$, then $\|u\|_2$ is bounded.
- If $\|u\|_2 \geq \frac{\|f\|_2}{c}$, then $\frac{d}{dt}\left(\|u\|_2^2\right) \leq 0$, and thus $\|u\|_2$ is also bounded.

□

The coercivity assumption on A can further be weakened. For instance, if we write $u = v+w$ with $v \in \mathrm{Ker}(A)$ and $w \in (\mathrm{Ker}(A))^{\perp}$ (notice that this decomposition exists and is unique), then we only need a uniform boundedness assumption on v, as A is coercive on the orthogonal of its kernel. We summarize these results in Corollary 3.8.

Corollary 3.8. *If $u(t)$ satisfies the flow (29), we can uniquely express $u = v + w$ with $v \in \mathrm{Ker}(A)$ and $w \in (\mathrm{Ker}(A))^{\perp}$. If v is uniformly bounded, then u is uniformly bounded.*

Proof. Let us define $V := \mathrm{Ker}(A), W := (\mathrm{Ker}(A))^{\perp}$, and $c := \min_{\{\|w\|_2=1, w\in W\}} \|Aw\|_2$. since A is a continuous linear operator defined on $\mathbb{R}^N$, we have $c > 0$, which implies that $\|Aw\|_2 \geq c\|w\|_2$ for all $w \in W$, and thus A is coercive on W. Now, if $u \in \mathbb{R}^N$, then there exists a unique decomposition $u = v + w$ with $v \in V$ and $w \in W$. We have $Au = Av + Aw = Aw$. From (31), we get:

$$\begin{aligned}\frac{d}{dt}\left(\|u\|_2^2\right) \leq & \lambda\|Au\|_2\left(\|f\|_2 - \|Au\|_2\right)\\ = & \lambda\|Aw\|_2\left(\|f\|_2 - \|Aw\|_2\right)\\ \leq & \lambda\|Aw\|_2\left(\|f\|_2 - c\|w\|_2\right).\end{aligned}$$

Since we know that v is uniformly bounded from the assumption of the corollary, it remains only to show that w is uniformly bounded. Similarly to the proof of Corollary 3.7, we then discuss two cases:

- If $\|w\|_2 \leq \frac{\|f\|_2}{c}$, then $\|w\|_2$ is bounded. Using the assumption that v is uniformly bounded, we get $u = v + w$ is uniformly bounded.
- If $\|w\|_2 \geq \frac{\|f\|_2}{c}$, then $\frac{d}{dt}\left(\|u\|_2^2\right) \leq 0$, which implies the boundedness of $\|u\|_2$.

□

4. **Numerical results.** In this section, we showcase the effectiveness of the proposed algorithms through a set of numerical experiments. All of these experiments were carried out on a typical laptop featuring a CPU (AMD Ryzen 5 4600U at 2.10GHz) and MATLAB (R2021b).

We start with some numerical insights of the proposed scheme in Section 4.1, followed by case studies of sparse signal recovery in Section 4.2 and MRI reconstruction in Section 4.3. Specifically for signal recovery, we conduct experiments in a noisy setting, aiming to recover an underlying sparse vector $u \in \mathbb{R}^n$ with s non-zero elements from a set of noisy measurements, $f = Au + \nu$, where $A \in \mathbb{R}^{m\times n}$ is a Gaussian random matrix with each column normalized by zero mean and unit Euclidean norm, and ν is Gaussian noise with zero mean and standard deviation σ. We fix the ambient dimension $n = 512$, sparsity $s = 130$, and noise level $\sigma = 0.1$, while varying the number of measurements m to examine the performance of sparse

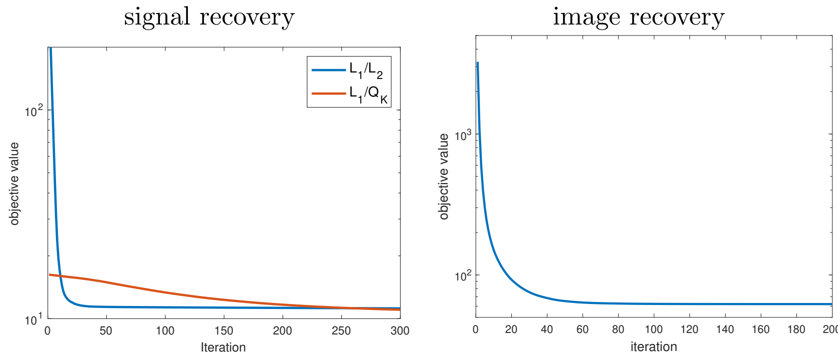

FIGURE 2. The objective function (7) respect to the iteration counts: L_1/L_2 and L_1/Q_K for signal recovery (left) and L_1/L_2 on the gradient for image recovery (right). The decay in each objective function provides empirical evidence of the convergence of the proposed scheme (11).

signal recovery. Notice that fewer measurements result in a more challenging recovery process. We use the mean-square error (MSE) metric to evaluate the recovery performance. we can obtain the ordinary least square (OLS) solution if we know the ground truth of the support set of $\Lambda = \text{supp}(u)$, which refers to the index set of nonzero entries in u. In this case, we can consider the mean squared error (MSE) of OLS as the benchmark for the oracle performance, using $\sigma \text{tr}(A_\Lambda^\top A_\Lambda)^{-1}$, where A_Λ refers to a submatrix of A by taking the columns corresponding to the index set Λ.

4.1. **Algorithm behaviors.** The convergence analysis we conduct in Section 3.1 is based on a modified model (22), as opposed to our numerical scheme (11). Here we empirically demonstrate the convergence of the latter on the three quotient models: L_1/L_2, L_1/Q_K, and L_1/L_2 on the gradient. The first two models are related to signal recovery, and we choose $K = 100$ for the L_1/Q_K model in this experiment, while the last one is stemmed from the image processing literature. The objective function for all these models is expressed in (7). We plot the objective value $R(u^k) + \frac{\lambda}{2}\|Au^k - f\|_2^2$ with respect to k, in which u^k is defined by (11). As illustrated in Figure 2, all the objective curves decrease rapidly, which provides strong evidence that the proposed scheme (11) is convergent. The theoretical analysis of (11) is left for future work.

Theorem 3.6 discusses a monotonically decreasing property of the sequence generated by the fully implicit scheme (22). Here, we verify this property numerically for the sequence generated by the semi-implicit scheme (11), using the L_1/L_2 model as an example. Specifically, we choose an initial guess of u^0 such that $\|Au^0\|_2 - \|f\|_2$ is strictly larger than 0 as Case 1, and $\|Au^0\|_2 - \|f\|_2 < 0$ as Case 2. We plot $\|u^k\|_2$ and $\|Au^k\|_2 - \|f\|_2$ with respect to k in Figure 3, which validates the decrease in $\|u\|_2$ is attributed to $\|Au^k\|_2 \geq \|f\|_2$.

Lastly, we investigate the impact of the parameter K for the L_1/Q_K model. We consider $m = 250$ to 360 with an increment of 10. For each m, we generate a random matrix A, a ground-truth sparse vector u of $s = 130$ nonzero elements, and a noise term ν to obtain the measurement vector f. We conduct 100 random realizations and record in Table 1 the average value of MSEs between the the ground-truth

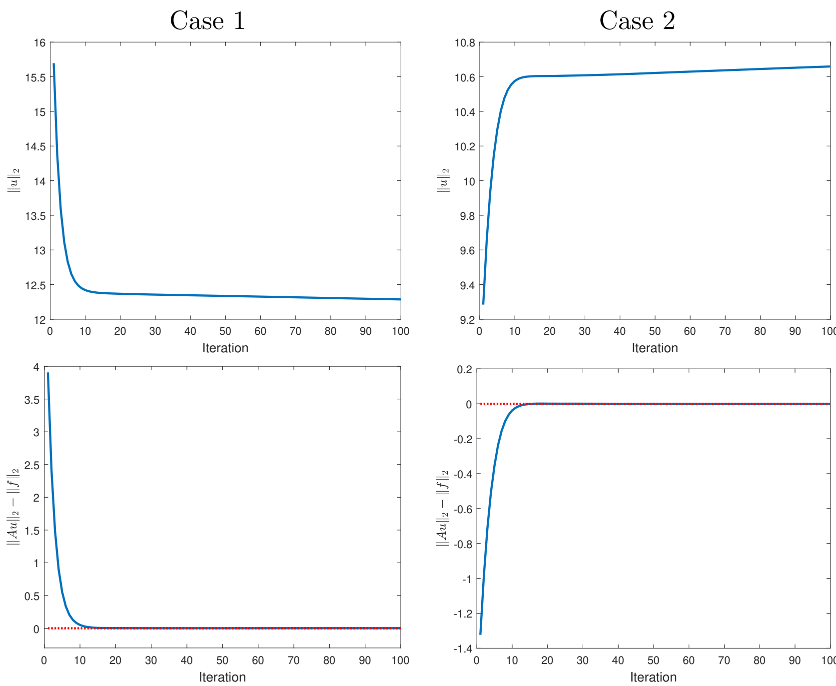

FIGURE 3. Numerical verification for the decreasing property revealed by Theorem 3.6: if $\|Au^0\|_2 \geq \|f\|_2$, then $\|u^k\|_2$ decreases with respect to the iteration (left); otherwise, $\|u^k\|_2$ increases (right). We plot $\|u^k\|_2$ on the top row, while $\|Au^0\|_2 - \|f\|_2$ with a baseline of 0 (red dash line) on the bottom row.

u and reconstructed solutions by L_1/Q_K with $K = 10, 100, 150, n(= 512)$. We use the L_1 solution as the initial condition for L_1/Q_K, which is referred to as the baseline model in Table 1. Notice that for $K = n$, the L_1/Q_K model becomes L_1/L_2. Table 1 shows that the L_1/Q_K model exhibits a close approximation to the oracle performance when $K = 100$ or 150, as the ground-truth sparsity is 130. When the parameter K is close to the ground-truth level, L_1/Q_K achieves top-notch performance at any m. For a smaller value of m, the problem becomes more ill-posed, and hence, all models lead to similar performance. If we choose $K = 10$ (far away from the true sparsity), the performance of L_1/Q_K is worse than the L_1/L_2 model, which implies that K plays an important role in the success of the L_1/Q_K model for sparse recovery.

4.2. **Signal recovery.** This section investigates the signal recovery problem, in which we compare the proposed Algorithm 1 on both L_1/L_2 and L_1/Q_K (choosing $K = 100$) regularizations with a difference of convex algorithm (DCA) scheme [24] implemented by ourselves. Here DCA aims to minimize $D_1(u) - D_2(u)$ with convex functionals D_1, D_2 by iteratively constructing two sequences $\{u^k\}$ and $\{v^k\}$ in the following way,

$$\begin{cases} v^k \in \partial D_2\left(u^k\right) \\ u^{k+1} = \arg\min_u D_1(u) - \langle u, v^k \rangle . \end{cases} \tag{33}$$

TABLE 1. Impact of the parameter K on the sparse recovery via the L_1/Q_K model. The sensing matrix A is of size $m \times n$, where m ranges from 250 to 360 and $n = 512$. The ground-truth sparse vector contains $s = 130$ nonzero elements. Each recorded value is averaged over 100 random realizations. The baseline model refers to the L_1 minimization, whose solution serves as the initial condition for L_1/Q_K. When K is chosen to be close to the true sparsity level (e.g., $K = 100, 150$ versus $s = 130$), L_1/Q_K yields top-notch performance; otherwise (e.g., $K = 10$), $L_1/L_2(K = n)$ is the best.

K \ m	250	260	270	280	290	300
baseline	5.27	4.97	4.59	4.44	4.20	3.91
10	5.20	4.87	4.44	4.19	3.93	3.69
100	4.95	4.57	4.12	**3.80**	3.56	**3.29**
150	**4.92**	**4.55**	**4.10**	**3.80**	**3.53**	**3.29**
n	5.01	4.65	4.19	3.90	3.65	3.43

K \ m	310	320	330	340	350	360
baseline	3.73	3.55	3.49	3.26	3.13	3.02
10	3.42	3.25	3.11	2.97	2.86	2.75
100	**3.01**	**2.86**	**2.70**	**2.57**	**2.46**	**2.34**
150	3.02	**2.86**	2.71	2.58	2.49	2.39
n	3.16	3.04	2.91	2.81	2.73	2.65

We consider splitting the objective function (7) into :

$$\begin{aligned} D_1(u) &= \mu\|u\|_1 + \frac{\lambda}{2}\|Au - f\|_2, \\ D_2(u) &= \mu\|u\|_1 - R(u). \end{aligned} \tag{34}$$

The u-subproblem in DCA (33) amounts to an L_1 regularized problem, which can be solved by ADMM.

In addition, it follows from [20] that the L_1/Q_K model can be formulated by

$$\min_{u\in\mathbb{R}^n} \frac{\|u\|_1 + \frac{\lambda}{2}\|Au - f\|_2^2}{\|u\|_{(K)}}, \tag{35}$$

so that a fractional programming (FP) strategy [34] can be applied. Specifically, a proximal-gradient-subgradient algorithm with backtracked extrapolation (PGSA_BE) was proposed in [20], which investigated a different model of L_1/Q_K with the L_1 norm of the K-largest magnitudes instead of the L_2 norm in the denominator. As L_1/Q_K becomes L_1/L_2 for $K = n$, we implement the ADMM algorithm for the L_1/L_2 model under either QRM (3) or FP (35) setting.

We randomly generate the matrix A of size $m \times 512$ for m varying from 240 to 360 with an increment of 20. Since the quotient models are non-convex, the choice of initial guess u^0 significantly impacts the performance. We adopt the restored solution via the L_1 minimization as the initial guess and terminate the iterations when the relative error $\|u^{k+1} - u^k\|_2/\|u^{k+1}\|_2$ is less than 10^{-8}. This stopping criterion is used for all the algorithms. Table 2 reports the averaged MSE

TABLE 2. MSEs of recovering a sparse vector of length $n = 512$ with $s = 130$ nonzero elements from m noisy measurements ($m = 240 : 20 : 360$ following the MatLab's notation). We compare L_1/L_2 and L_1/Q_K for $K = 100$ under the settings of FP (35) and QRM (3). We observe that QRM is a better framework than FB for sparse recovery. The best results are consistently given by the proposed algorithm for solving the L_1/Q_K model when the value of $K = 100$ is close to the true sparsity level (130). The L_1/L_2 (when $K = n$) model achieves the second best in performance.

	model-algorithm	240	260	280	300	320	340	360
FP	L_1/L_2-ADMM	5.51	4.76	4.00	3.48	3.15	2.86	2.67
	L_1/L_2-PGSA_BE	11.12	8.60	6.28	4.40	3.37	2.83	2.52
	L_1/Q_K-PGSA_BE	5.62	4.75	3.92	3.41	3.08	2.83	2.68
QRM	L_1/L_2-DCA	5.56	4.87	4.14	3.61	3.27	2.95	2.69
	L_1/L_2-ADMM	5.53	4.75	3.96	3.45	3.12	2.86	2.68
	L_1/L_2-proposed	5.50	4.70	3.92	3.40	3.07	2.81	2.64
	L_1/Q_K-DCA	5.52	4.77	4.01	3.48	3.15	2.86	2.67
	L_1/Q_K-proposed	**5.44**	**4.65**	**3.83**	**3.26**	**2.91**	**2.57**	**2.33**

values over 100 random realizations. We observe that the QRM framework always performs better than FP for the same regularization. The L_1/Q_K model solved by our algorithm performs the best in all the cases when $K = 100$ is chosen near the true sparsity level (130), and L_1/L_2 without knowing the sparsity ranks the second best.

Furthermore, we examine the performance of various methods in terms of support identification, which can be viewed as a binary classification problem determining whether an index belongs to the ground-truth support set. Following the statistical analysis, we use the F1 score [18] to evaluate the identification/classification performance. The F1 score is the harmonic mean of the precision (positive predictive value) and recall (hit rate). The closer the F1 score to 1, the better the identification is. Table 3 shows that the best results are consistently obtained by the proposed algorithm for solving the L_1/Q_K model when the value of $K = 100$ is close to the true sparsity level (130).

In summary, the proposed algorithms for solving two QRM models with L_1/L_2 and L_1/Q_K outperform the other relevant approaches in terms of recovery accuracy and support identification.

4.3. **Image recovery.** We consider an MRI reconstruction as a proof-of-concept example in image processing. The MRI measurements are acquired through multiple radial lines in the frequency domain, achieved by performing the Fourier transform. In addition, we add the Gaussian noise, with a mean of zero and standard deviation σ on the MRI measurements. Intuitively, fewer radial lines and a larger σ value bring more ill-posedness and difficulty to the problem. Here we consider two standard phantoms, namely Shepp–Logan (SL) phantom generated using MATLAB's built-in command `phantom` and the FORBILD (FB) phantom [33]. We evaluate the performance in terms of the relative error (RE) and the peak signal-to-noise ratio

Table 3. F1 score on support identification, with $s = 130$ nonzero elements from m noisy measurements ($m = 240 : 20 : 360$ following the MatLab's notation). We compare L_1/L_2 and L_1/Q_K for $K = 100$ under the settings of FP (4) and QRM (3). The best results are consistently given by the proposed algorithm for solving the L_1/Q_K model when the value of $K = 100$ is close to the true sparsity level (130).

	model-algorithm	240	260	280	300	320	340	360
FP	L_1/L_2-ADMM	0.41	0.41	0.41	0.41	0.41	0.41	0.41
	L_1/L_2-PGSA_BE	0.44	0.51	0.58	0.65	0.69	0.70	0.70
	L_1/Q_K-PGSA_BE	0.49	0.50	0.50	0.50	0.50	0.49	0.49
QRM	L_1/L_2-DCA	0.41	0.41	0.41	0.41	0.41	0.41	0.41
	L_1/L_2-ADMM	0.49	0.49	0.50	0.49	0.49	0.48	0.47
	L_1/L_2-proposed	0.49	0.50	0.51	0.50	0.50	0.49	0.48
	L_1/Q_K-DCA	0.41	0.41	0.41	0.41	0.41	0.41	0.41
	L_1/Q_K-proposed	**0.59**	**0.61**	**0.65**	**0.66**	**0.67**	**0.67**	**0.67**

(PSNR), defined by

$$\mathrm{RE}(u^*, \tilde{u}) := \frac{\|u^* - \tilde{u}\|_2}{\|\tilde{u}\|_2} \quad \text{and} \quad \mathrm{PSNR}(u^*, \tilde{u}) := 10 \log_{10} \frac{NP^2}{\|u^* - \tilde{u}\|_2^2},$$

where u^* is the restored image, $\tilde{u}$ is the ground truth, and P is the maximum peak value of $\tilde{u}$.

Similar to the signal-recovering experiments, we regard the performance of the L_1 on the gradient, i.e., the total variation (TV), as the baseline. For L_1/L_2 on the gradient, we compare the proposed algorithm to a previous method based on ADMM [25]. For three sampling schemes (7, 10, and 13 lines) and two noise levels ($\sigma = 0.01$ and 0.05), we record RE and PSNR values of three methods in Table 4, demonstrating significant improvements in the accuracy of the proposed approach over the previous works.

Figures 4 and 5 present visual reconstruction results of the SL phantom and the FB phantom, respectively, both under high additive Gaussian noise ($\sigma = 0.05$). In particular, Figure 4 is to recover the SL phantom using 7 radial lines. The L_1 model has severe streaking artifacts due to this extremely small number of data obtained on the radial lines. The L_1/L_2 minimization on the gradient yields significant improvements over the baseline model (TV). The proposed algorithm outperforms the previous ADMM approach at the outer ring and boundaries of the three middle oval shapes, which are more obvious in the difference map to the ground truth. On the other hand, the FB phantom has finer structures and lower image contrast compared to the SL phantom. As a result, it requires 13 radial lines for a reasonable reconstruction. As we observe in Figure 5, the overall geometric shapes are preserved. At the same time, many speckle artifacts appear in the reconstructed images by L_1/L_2 no matter which algorithm is used.

5. **Conclusions.** In this paper, we proposed a gradient descent flow to minimize a quotient regularization model with a quadratic data fidelity term for signal and image processing applications. We assumed the numerator and the denominator in the quotient model are absolutely one homogeneous. Numerically, we adopted a

TABLE 4. MRI reconstruction from different numbers of radial lines and different noise levels.

Image	σ	Line	L_1		L_1/L_2-ADMM		L_1/L_2-proposed	
			RE	PSNR	RE	PSNR	RE	PSNR
SL	0.01	7	46.06%	19.50	25.36%	24.09	**3.74%**	**40.72**
		10	16.29%	28.66	3.41%	41.53	**2.91%**	**42.90**
		13	6.85%	36.52	1.91%	46.55	**1.71%**	**47.49**
	0.05	7	52.31%	18.33	43.63%	19.38	**31.90%**	**22.10**
		10	33.09%	22.42	14.34%	29.04	**14.08%**	**29.24**
		13	22.67%	26.10	10.50%	31.75	**10.41%**	**31.82**
FB	0.01	7	21.63%	21.49	13.80%	24.89	**1.11%**	**26.94**
		10	18.14%	23.08	14.98%	24.17	**12.90%**	**25.47**
		13	9.51%	28.29	1.41%	44.71	**1.17%**	**46.31**
	0.05	7	26.03%	19.9	22.14%	20.78	**16.50%**	**23.36**
		10	18.14%	23.08	14.98%	24.17	**12.90%**	**25.47**
		13	14.48%	24.79	12.67%	25.64	**12.30%**	**25.89**

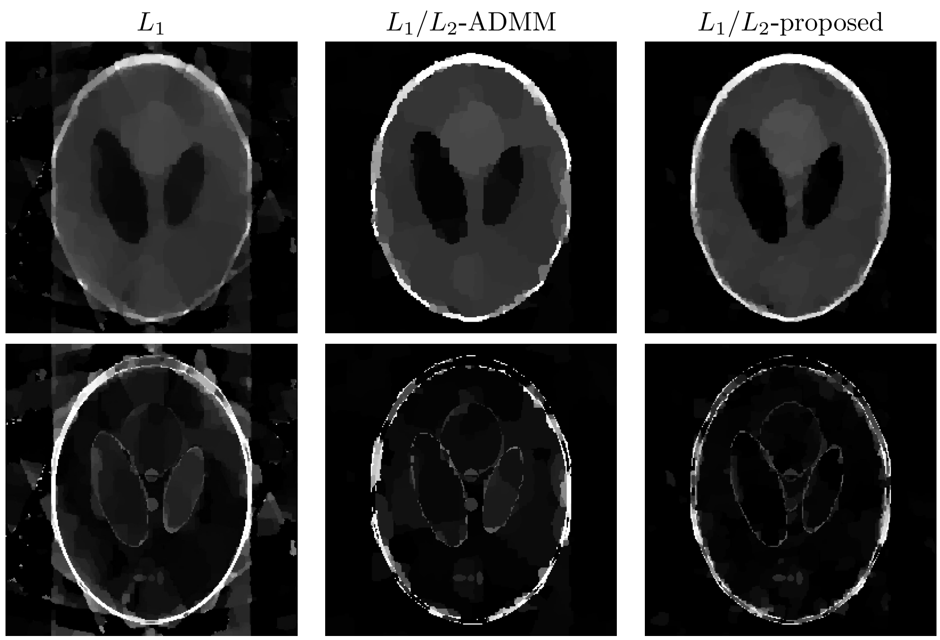

FIGURE 4. MRI reconstruction on the SL phantom with a noise level of $\sigma = 0.05$ with 7 radial lines. Top row – reconstruction results, bottom row – difference from ground truth. The proposed algorithm outperforms the previous ADMM approach at the outer ring and boundaries of the three middle oval shapes, better seen in the difference map.

semi-implicit scheme that involves solving a convex problem iteratively. We theoretically analyzed the convergence of a slightly modified algorithm in a continuous formulation, as opposed to the discretized version. Experimentally, we presented the comparison results of three case studies of L_1/L_2 and L_1/Q_K for signal recovery and L_1/L_2 on the gradient for MRI reconstruction. We demonstrated that the proposed algorithm significantly outperforms the previous methods in each case in terms of accuracy. Future work includes the speed-up of the proposed algorithm,

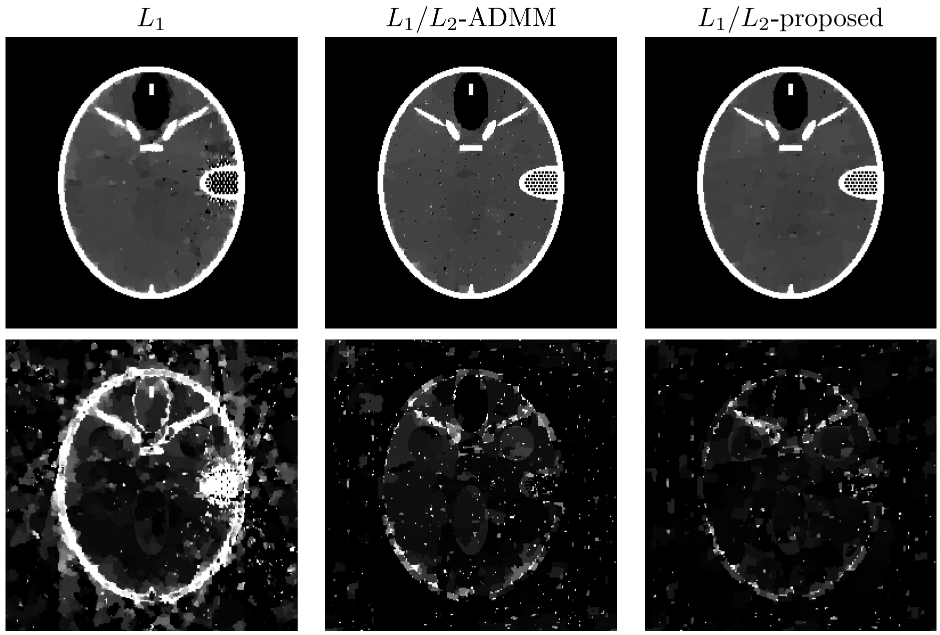

FIGURE 5. MRI reconstruction on the FB phantom with a noise level of $\sigma = 0.05$ with 13 radial lines. Top row – reconstruction results, bottom row – difference from ground truth. The proposed algorithm is able to better preserve the overall geometric shapes, compared to competing methods.

e.g., trying to make a single loop rather than the double loop, and the convergence analysis of the actual scheme.

REFERENCES

[1] J.-F. Aujol, G. Gilboa and N. Papadakis, Theoretical analysis of flows estimating eigenfunctions of one-homogeneous functionals, *SIAM J. Imaging Sci.*, **11** (2018), 1416-1440.

[2] H. Bauschke and P. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2017.

[3] M. Benning, G. Gilboa, J. S. Grah and C.-B. Schönlieb, Learning filter functions in regularisers by minimising quotients, *International Conference on Scale Space and Variational Methods in Computer Vision (SSVM), Kolding, Denmark*, Springer, **6** (2017), 511-523.

[4] M. Benning, G. Gilboa and C.-B. Schönlieb, Learning parametrised regularisation functions via quotient minimisation, *PAMM*, **16** (2016), 933-936.

[5] S. Boyd, N. Parikh and E. Chu, *Distributed Optimization and Statistical Learning Via the Alternating Direction Method of Multipliers*, Now Publishers Inc, 2011.

[6] X. Bresson, T. Laurent, D. Uminsky and J. V. Brecht, Convergence and energy landscape for Cheeger cut clustering, *Adv. Neural Inf. Process. Syst.*, (2012), 1385-1393.

[7] L. Bungert, E. Hait-Fraenkel, N. Papadakis and G. Gilboa, Nonlinear power method for computing eigenvectors of proximal operators and neural networks, *SIAM J. Imaging Sci.*, **14** (2021), 1114-1148.

[8] M. Burger, G. Gilboa, M. Moeller, L. Eckardt and D. Cremers, Spectral decompositions using one-homogeneous functionals, *SIAM Journal on Imaging Sciences*, **9** (2016), 1374-1408.

[9] A. Cherni, E. Chouzenoux, L. Duval and J.-C. Pesquet, SPOQ ℓ_p-over-ℓ_q regularization for sparse signal recovery applied to mass spectrometry, *IEEE Trans. Signal Process.*, **68** (2020), 6070-6084.

[10] L. Demanet and P. Hand, Scaling law for recovering the sparsest element in a subspace, *Information and Inference: A Journal of the IMA*, **3** (2014), 295-309.
[11] T. Feld, J.-F. Aujol, G. Gilboa and N. Papadakis, Rayleigh quotient minimization for absolutely one-homogeneous functionals, *Inverse Probl.*, **35** (2019), 064003.
[12] D. Gabay and B. Mercier, A dual algorithm for the solution of nonlinear variational problems via finite element approximation, *Comput. Math. Appl.*, **2** (1976), 17-40.
[13] M. Hein and T. Bühler, An inverse power method for nonlinear eigenproblems with applications in 1-spectral clustering and sparse pca, *Adv. Neural Inf. Process. Syst.*, **23**.
[14] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 2012.
[15] P. O. Hoyer, Non-negative matrix factorization with sparseness constraints., *J. Mach. Learn. Res.*, **5**.
[16] Y. Hu, D. Zhang, J. Ye, X. Li and X. He, Fast and accurate matrix completion via truncated nuclear norm regularization, *IEEE Trans. Pattern Anal. Mach. Intell.*, **35** (2012), 2117-2130.
[17] N. Hurley and S. Rickard, Comparing measures of sparsity, *IEEE Trans. Inf. Theory*, **55** (2009), 4723-4741.
[18] H. K. Klein and M. D. Myers, A set of principles for conducting and evaluating interpretive field studies in information systems, *MIS Quarterly*, 67-93.
[19] J. Lei, Q. Liu and X. Wang, Physics-informed multi-fidelity learning-driven imaging method for electrical capacitance tomography, *Eng. Appl. Artif. Intell.*, **116** (2022), 105467.
[20] Q. Li, L. Shen, N. Zhang and J. Zhou, A proximal algorithm with backtracked extrapolation for a class of structured fractional programming, *Appl. Comput. Harmon. Anal.*, **56** (2022), 98-122.
[21] B. S. Mordukhovich, *Variational Analysis and Generalized Differentiation II: Applications*, Springer, **331** (2006).
[22] R. Z. Nossek and G. Gilboa, Flows generating nonlinear eigenfunctions, *J. Sci. Comput.*, **75** (2018), 859-888.
[23] T.-H. Oh, Y.-W. Tai, J.-C. Bazin, H. Kim and I. S. Kweon, Partial sum minimization of singular values in robust PCA: Algorithm and applications, *IEEE Trans. Pattern Anal. Mach. Intell.*, **38** (2015), 744-758.
[24] T. Pham-Dinh and H. A. Le-Thi, The DC (difference of convex functions) programming and DCA revisited with DC models of real world nonconvex optimization problems, *Ann. Oper. Res.*, **133** (2005), 23-46.
[25] Y. Rahimi, C. Wang, H. Dong and Y. Lou, A scale-invariant approach for sparse signal recovery, *SIAM J . Sci. Comput.*, **41** (2019), A3649-A3672.
[26] M. Tao, Minimization of L_1 over L_2 for sparse signal recovery with convergence guarantee, *SIAM J. Sci. Comput.*, **44** (2022), A770-A797.
[27] C. Wang, M. Tao, C.-N. Chuah, J. Nagy and Y. Lou, Minimizing L_1 over L_2 norms on the gradient, *Inverse Probl.*, **38** (2022), 065011.
[28] C. Wang, M. Tao, J. G. Nagy and Y. Lou, Limited-angle CT reconstruction via the L_1/L_2 minimization, *SIAM J. Imaging Sci.*, **14** (2021), 749-777.
[29] C. Wang, M. Yan, Y. Rahimi and Y. Lou, Accelerated schemes for the L_1/L_2 minimization, *IEEE Trans. Signal Process.*, **68** (2020), 2660-2669.
[30] J. Wang, A wonderful triangle in compressed sensing, *Information Sciences*, **611** (2022), 95-106.
[31] T. Wu, Z. Mao, Z. Li, Y. Zeng and T. Zeng, Efficient color image segmentation via quaternion-based l_1/l_2 regularization, *J. Sci. Comput.*, **93** (2022), 9.
[32] Y. Xu, A. Narayan, H. Tran and C. G. Webster, Analysis of the ratio of ℓ_1 and ℓ_2 norms in compressed sensing, *Appl. Comput. Harmon. Anal.*, **55** (2021), 486-511.
[33] Z. Yu, F. Noo, F. Dennerlein, A. Wunderlich, G. Lauritsch and J. Hornegger, Simulation tools for two-dimensional experiments in x-ray computed tomography using the FORBILD head phantom, *Phys. Med. Biol.*, **57** (2012), N237.
[34] N. Zhang and Q. Li, First-order algorithms for a class of fractional optimization problems, *SIAM J. Optim.*, **32** (2022), 100-129.

Received November 2023; revised July 2024; early access October 2024.