Partial gradient optimal thresholding algorithms for a class of sparse optimization problems

Nan Meng[1] · Yun-Bin Zhao[2] · Michal Kočvara[1] · Zhongfeng Sun[3]

Abstract

The optimization problems with a sparsity constraint is a class of important global optimization problems. A typical type of thresholding algorithms for solving such a problem adopts the traditional full steepest descent direction or Newton-like direction as a search direction to generate an iterate on which a certain thresholding is performed. Traditional hard thresholding discards a large part of a vector, and thus some important information contained in a dense vector has been lost in such a thresholding process. Recent study (Zhao in SIAM J Optim 30(1): 31–55, 2020) shows that the hard thresholding should be applied to a compressible vector instead of a dense vector to avoid a big loss of information. On the other hand, the optimal k-thresholding as a novel thresholding technique may overcome the intrinsic drawback of hard thresholding, and performs thresholding and objective function minimization simultaneously. This motivates us to propose the so-called partial gradient optimal thresholding (PGOT) method and its relaxed versions in this paper. The PGOT is an integration of the partial gradient and the optimal k-thresholding technique. The solution error bound and convergence for the proposed algorithms have been established in this paper under suitable conditions. Application of our results to the sparse optimization problems arising from signal recovery is also discussed. Experiment results from synthetic data indicate that the proposed algorithm is efficient and comparable to several existing algorithms.

✉ Yun-Bin Zhao
  yunbinzhao@cuhk.edu.cn

  Nan Meng
  nxm563@bham.ac.uk

  Michal Kočvara
  m.kocvara@bham.ac.uk

  Zhongfeng Sun
  zfsun@sdut.edu.cn

[1] School of Mathematics, University of Birmingham, Edgbaston, Birmingham B15 2TT, UK

[2] Shenzhen Research Institute of Big Data, Chinese University of Hong Kong, Shenzhen, China

[3] School of Mathematics and Statistics, Shandong University of Technology, Zibo, China

1 Introduction

The optimization problem with a sparsity constraint

$$\min_{x \in \mathbb{R}^n} \{ f(x) : \ \|x\|_0 \leq k \} \tag{1}$$

arises in many practical fields such as compressive sensing [20, 22, 37], signal processing [8], wireless communication [15], pattern recognition [33], to name a few. In the model (1), $\| \cdot \|_0$ is called the 'ℓ_0-norm' which counts the number of nonzero entries of a vector. Depending on application, the function $f(x)$ may take different specific forms. For instance, in compressed sensing scenarios, $f(x)$ is usually taken as $\|y - Ax\|_2^2$ which is an error metric for signal measurements. The problem (1) is known to be NP-hard, and the main difficulty for solving this problem lies in locating the position of nonzero entries of a feasible sparse vector at which $f(x)$ is minimized.

The algorithms for solving (1) can be sorted into several categories including convex optimization methods, heuristic algorithms, and thresholding algorithms. The convex optimization methods include ℓ_1-minimization [11, 14], reweighted ℓ_1-minimization [13, 40], and dual-density-based reweighted ℓ_1-minimization [37, 39, 41]. The heuristic-type methods include orthogonal matching pursuit (OMP) [10, 32, 34], compressive sampling matching pursuit (CoSaMP) [31], subspace pursuit (SP) [16, 17], and their variants. Thresholding-type algorithms attract much attention due to their easy implementation and low computational complexity [4, 6, 7, 9, 23, 27, 38].

The key step for thresholding-type iterative algorithms can be stated as

$$x^{p+1} = \mathcal{T}_k \left(x^p + \lambda d \right), \tag{2}$$

where $\mathcal{T}_k(\cdot)$ represents a thresholding operator that is used to produce a k-sparse vector, λ denotes the stepsize and d is a search direction at the current iterate x^p. Throughout the paper, a vector x is said to be k-sparse if $\|x\|_0 \leq k$. Several thresholding operators are widely used in the literature, such as the hard thresholding [4, 6, 7, 9, 23, 27], soft thresholding [18, 21, 26, 28], and optimal k-thresholding [38, 42]. The steepest descent direction [5–7, 9, 24, 30] and Newton-type direction [29, 35, 43, 44] are two search directions that are used by many researchers.

Let $\mathcal{H}_k$ denote the hard thresholding operator which retains the largest k magnitudes and zeroes out other entries of a vector. By setting $\mathcal{T}_k = \mathcal{H}_k$ and $d = -\nabla f(x)$, the iterative formula (2) is reduced to

$$x^{p+1} = \mathcal{H}_k \left(x^p - \lambda \nabla f(x^p) \right), \tag{3}$$

where $\nabla f(x^p)$ is the gradient of f at x^p. The formula (3) is the well-known iterative hard thresholding (IHT) initially studied in [5, 6]. The IHT can be enhanced by either attaching an orthogonal projection (a pursuit step) to obtain the so-called hard thresholding pursuit (HTP) method [9, 23] or by using an adaptive stepsize strategy to yield the so-called normalized iterative hard thresholding (NIHT) [7]. While the algorithm (3) can reconstruct the vector under suitable conditions (see [6, 22, 36]), but as pointed in [38, 42], the operator $\mathcal{H}_k$ may cause certain numerical problems as well.

To improve the performance of IHT, Zhao [38, 42] recently proposed the optimal k-thresholding technique which stresses that thresholding of a vector should be performed simultaneously with objective function reduction in the course of iterations. Replacing $\mathcal{H}_k$ by the optimal k-thresholding operator $\mathcal{Z}_k^{\#}$ in (3) leads to the following iterative optimal k-thresholding scheme:

$$x^{p+1} = \mathcal{Z}_k^{\#}\left(x^p - \lambda \nabla f(x^p)\right).$$

The optimal k-thresholding of a vector $u \in \mathbb{R}^n$ with respect to the objective function $f(x)$ is defined as $\mathcal{Z}_k^{\#}(u) := u \otimes w^*$ with

$$w^* = \arg\min_{w} \left\{ f(u \otimes w) : \mathbf{e}^T w = k, \ w \in \{0, 1\}^n \right\}, \tag{4}$$

where $\mathbf{e} \in \mathbb{R}^n$ is the vector of ones and $\otimes$ denotes the Hadamard product of two vectors. To avoid solving the above binary optimization problem, Zhao [38] suggests solving the tightest convex relaxation of (4) instead. That is, replacing the binary constraint by its convex hull, we obtain the following convex relaxation of the problem (4):

$$\overline{w} = \arg\min_{w} \left\{ f(u \otimes w) : \mathbf{e}^T w = k, \ 0 \le w \le \mathbf{e} \right\}.$$

The vector $u \otimes \overline{w}$ is called the relaxed optimal k-thresholding of u.

The hard thresholding operator in (3) discards a large important part of a vector when the vector is dense. This means some important information of the vector has been lost in the process of hard thresholding. As pointed out in [38, 42], the hard thresholding should be applied to a compressible vector instead of a dense vector in order to avoid a big loss of information. Note that the vector $x^p - \lambda \nabla f(x^p)$ in (3) is usually dense since the search direction $-\nabla f(x^p)$ is not necessarily sparse. This motivates us to adopt the partial gradient instead of the full gradient as a search direction in order to generate the following sparse or compressible vector:

$$u^p := x^p - \lambda \mathcal{H}_q(\nabla f(x^p))$$

on which some thresholding is then performed to generate the next iterate x^{p+1}. In the above formula, the integer number $q > 0$ controls the number of elements extracted from the full gradient. In other words, we only use q significant entries of the gradient as our search direction. We may use the hard thresholding of u^p to produce an iterate satisfying the constraint of the problem (1). However, as we pointed out before, the optimal k-thresholding is more powerful and more efficient than the hard thresholding. This stimulates the following iterative scheme:

$$x^{p+1} = \mathcal{Z}_k^{\#}\left(x^p - \lambda \mathcal{H}_q(\nabla f(x^p))\right). \tag{5}$$

This is refer to as the partial gradient optimal thresholding (PGOT) algorithm in this paper, which is described in detail in Sect. 2. The enhanced version of PGOT is called the partial gradient relaxed optimal thresholding (PGROT). In order to maintain the k-sparsity of the iterate, a further enhancement of PGOT and PGROT can be made by adding a pursuit step to PGROT to eventually obtain the more efficient algorithm called the partial gradient relaxed optimal thresholding pursuit (PGROTP), which is treated as the final version of the proposed algorithm in the paper, actually being used to solve the problems. The solution error bound and convergence analysis for our algorithms with q in the range $q \in [2k, n]$ are shown under the assumption of restricted isometry property (RIP). Simulations from synthetic data indicate that PGROTP algorithm is robust and comparable to several existing methods.

The paper is organized as follows. The algorithms are described in Sect. 2. The error bounds and global convergence of the proposed algorithms are established in Sect. 3. Numerical results are given in Sect. 4 and conclusions are given in Sect. 5.

1.1 Notations

We first introduce some notations used throughout the paper. $\mathbb{R}^n$ is the n-dimensional Euclidean space, and $\mathbb{R}^{m \times n}$ is the set of $m \times n$ matrices. Vector $\mathbf{e}$ is the vector of ones. Denote $[N]$ as the set $\{1, \ldots, n\}$. Given a set $\Omega \subseteq [N]$, $\overline{\Omega} := [N] \backslash \Omega$ denotes the complement set of Ω, and $|\Omega|$ is the cardinality of set Ω. For a vector $x \in \mathbb{R}^n$, $x_\Omega \in \mathbb{R}^n$ denotes the vector obtained from x by retaining elements indexed by Ω and zeroing out the remaining ones. The set $\text{supp}(x) = \{i : x_i \neq 0\}$ is called the support of x, $\mathcal{L}_k(x)$ denotes the support of $\mathcal{H}_k(x)$, and $\mathcal{Z}_k^\#(\cdot)$ denotes the optimal k-thresholding operator. For a matrix A, A^T denotes its transpose. The notation $\otimes$ represents the Hadamard product of two vectors, i.e., $u \otimes v = [u_1 v_1, \ldots, u_n v_n]^T$. Given a number α, $\lceil \alpha \rceil$ is the smallest integer number that is larger than or equal to α.

2 Algorithms

In this paper, we focus on the following specific objective function:

$$f(x) = \frac{1}{2} \|y - Ax\|_2^2, \tag{6}$$

where A is a given $m \times n$ matrix with $m \ll n$, and y is a given vector. Using this quadratic function, the model (1) becomes

$$\min_{x \in \mathbb{R}^n} \left\{ \frac{1}{2} \|y - Ax\|_2^2 : \|x\|_0 \leq k \right\}. \tag{7}$$

This problem has been widely used in data sparse representation, statistical regression, and signal reconstruction via compressive sensing and in many other application settings. The gradient of the function (6) is given as

$$\nabla f(x) = -A^T (y - Ax).$$

By using the major part of this specific gradient, we define the u^p as follows:

$$u^p = x^p + \lambda \mathcal{H}_q \left(A^T (y - Ax^p) \right),$$

where $0 < q \leq n$ is an integer number. The optimal thresholding of u^p with respect to the function (6) is given by $\mathcal{Z}_k^\#(u^p) = u^p \otimes w^*$, where

$$w^* = \arg \min_w \left\{ \|y - A(u \otimes w)\|_2^2 : \mathbf{e}^T w = k, \ w \in \{0, 1\}^n \right\}.$$

Thus the partial gradient optimal k-thresholding (PGOT) algorithm (5) for solving problem (7) can be stated as

$$x^{p+1} = \mathcal{Z}_k^\# \left(x^p + \lambda \mathcal{H}_q \left(A^T (y - Ax^p) \right) \right).$$

For simplicity of algorithmic description and analysis, we set $\lambda = 1$ throughout the rest of the paper. It should be stressed that in practical applications, suitable stepsize should be used

in order to speed up the convergence of the algorithms. By the definition of $\mathcal{Z}_k^\#$, the PGOT algorithm can be described explicitly as follows.

Algorithm 1 Partial Gradient Optimal k-Thresholding (PGOT)

- Input: matrix A, vector y, sparsity level k, integer number $q \geq k$, and initial point $x^0 = 0$.
- Iteration:

$$u^p = x^p + \mathcal{H}_q\left(A^T\left(y - Ax^p\right)\right), \tag{OT}$$

$$w^p = \arg\min_w\{\|y - A(w \otimes u^p)\|_2^2 : \mathbf{e}^T w = k, \ w \in \{0, 1\}^n\},$$

$$x^{p+1} = w^p \otimes u^p \ (= \mathcal{Z}_k^\#(u^p)).$$

- Output: k-sparse vector $\hat{x}$.

To avoid solving the integer programming problem (OT), as suggested in [38], the binary constraint $w \in \{0, 1\}^n$ in (OT) can be relaxed to $0 \leq w \leq \mathbf{e}$ so that we obtain the partial gradient relaxed optimal thresholding (PGROT) algorithm.

Algorithm 2 Partial Gradient Relaxed Optimal k-Thresholding (PGROT)

- Input: matrix A, vector y, sparsity level k, integer number $q \geq k$, and initial point $x^0 = 0$.
- Iteration:

$$u^p = x^p + \mathcal{H}_q\left(A^T\left(y - Ax^p\right)\right), \tag{ROT}$$

$$\overline{w}^p = \arg\min_w\{\|y - A(w \otimes u^p)\|_2^2 : \mathbf{e}^T w = k, \ 0 \leq w \leq \mathbf{e}\},$$

$$x^{p+1} = \mathcal{H}_k(\overline{w}^p \otimes u^p).$$

- Output: k-sparse vector $\hat{x}$.

The solution $\overline{w}^p$ to (ROT) is not k-sparse in general. So the purpose of the final thresholding step in PGROT to restore the k-sparsity of iterate. It is worth emphasizing that the use of $\mathcal{H}_k$ here is different from the settings in traditional IHT, since the vector u^p generated by the partial gradient is $(k + q)$-sparse instead of being a usually dense vector in IHT.

The PGROT can be further enhanced by including a pursuit step (i.e., an orthogonal projection step) to find a possibly better iterate than the point generated by PGROT. This consideration leads to so-called PGROTP algorithm which is described in Algorithm 3. In the next section, we perform a theoretical analysis for the proposed algorithms.

3 Error bound and convergence analysis

In this section, we establish the error bounds for the solution of the problem via the proposed algorithms. The purpose is to estimate the distance between the iterate x^p, generated by the proposed algorithms, and the global solution of the problem (7). As an implication of the error bounds, the global convergence of our algorithms can be instantly obtained for the

---

Algorithm 3 Partial Gradient Relaxed Optimal k-thresholding Pursuit (PGROTP)

- Input: matrix A, vector y, sparsity level k, integer number $q \geq k$, and initial point $x^0 = 0$.
- Iteration:

$$u^p = x^p + \mathcal{H}_q \left(A^T (y - Ax^p) \right), \qquad \text{(Pursuit step)}$$

$$\overline{w}^p = \arg\min_w \{ \left\| y - A(w \otimes u^p) \right\|_2^2 : \mathbf{e}^T w = k, \ 0 \leq w \leq \mathbf{e} \},$$

$$S^{p+1} = \mathcal{L}_k(\overline{w}^p \otimes u^p),$$

$$x^{p+1} = \arg\min_z \{ \|y - Az\|_2 : \ \text{supp}(z) \subseteq S^{p+1} \}.$$

- Output: k-sparse vector $\hat{x}$.

---

problem (7) in the scenarios of sparse signal recovery. Before going ahead, we first introduce the restricted isometry constant (RIC) of the matrix A.

Definition 1 [11, 22] The s-th order restricted isometry constant (RIC) δ_s of a matrix $A \in \mathbb{R}^{m \times n}$ is the smallest number $\delta_s \geq 0$ such that

$$(1 - \delta_s) \|x\|_2^2 \leq \|Ax\|_2^2 \leq (1 + \delta_s) \|x\|_2^2$$

for all s-sparse vector x, where $s > 0$ is an integer number.

We usually say that matrix A satisfies the RIP of order s if $\delta_s < 1$. It is well known that the random matrices such as Bernoulli and Gaussian matrices satisfy the RIP with an overwhelming probability [11, 12]. The following property of RIC is frequently used in the paper.

Lemma 1 [11, 23, 31] *Suppose matrix A satisfy the RIP of order k. Given a vector $u \in \mathbb{R}^n$ and a set $\Omega \subseteq [N]$, one has*

(i) $\left\| \left(\left(I - A^T A \right) v \right)_\Omega \right\|_2 \leq \delta_t \|u\|_2$ *if* $|\Omega \cup \text{supp}(v)| \leq t$.
(ii) $\left\| (A^T u)_\Omega \right\|_2 \leq \sqrt{1 + \delta_t} \|u\|_2$ *if* $|\Omega| \leq t$.

3.1 Main results for PGOT

The following two technical results concerning the properties of optimal k-thresholding and hard thresholding operators are useful.

Lemma 2 [38, 42] *Let $x^* \in \mathbb{R}^n$ be the solution to the problem (7) and denote by $\eta = y - Ax^*$. Given an arbitrary vector $u \in \mathbb{R}^n$, let $\mathcal{Z}_k^\#(u)$ be the optimal k-thresholding vector of u. Then for any k-sparse binary vector $w^* \in \{0, 1\}^n$ satisfying $\text{supp}(x^*) \subseteq \text{supp}(w^*)$, one has*

$$\left\| \mathcal{Z}_k^\#(u) - x^* \right\|_2 \leq \sqrt{\frac{1 + \delta_k}{1 - \delta_{2k}}} \left\| (x^* - u) \otimes w^* \right\|_2 + \frac{2}{\sqrt{1 - \delta_{2k}}} \|\eta\|_2.$$

This result can be found from the proof of Theorem 4.3 in [38].

Lemma 3 [36] *Let $q \geq s$ be an integer number. For any vector $z \in \mathbb{R}^n$ and any s-sparse vector $u \in \mathbb{R}^n$, one has*

$$\left\| u - \mathcal{H}_q(z) \right\|_2 \leq \frac{\sqrt{5} + 1}{2} \left\| (u - z)_{\Lambda \cup \Omega} \right\|_2,$$

where $\Lambda = \mathrm{supp}(u)$ and $\Omega = \mathrm{supp}\left(\mathcal{H}_q(z)\right)$.

When $s \leq q$, a s-sparse vector is also q-sparse. Thus Lemma 3 above follows exactly from Lemma 2.2 in [36]. We are ready to prove the error bound and global convergence of the algorithm PGOT.

Theorem 1 *Let $x^* \in \mathbb{R}^n$ be the solution of the problem (7) and $\eta := y - Ax^*$. Let $q \geq 2k$ be a positive integer number. Suppose that the restricted isometry constant of A satisfies*

$$\delta_{3k} < \alpha^*, \tag{8}$$

where $\alpha^* \in (0, 1)$ is the unique real root of the univariate equation

$$\left(\left\lceil \frac{q}{k} \right\rceil + 1\right)^2 \alpha^3 + \left(\left\lceil \frac{q}{k} \right\rceil + 1\right)^2 \alpha^2 + \frac{2}{3 + \sqrt{5}}\alpha - \frac{2}{3 + \sqrt{5}} = 0. \tag{9}$$

Then the sequence $\{x^p\}$ generated by PGOT satisfies that

$$\left\| x^{p+1} - x^* \right\|_2 \leq \rho^p \left\| x^0 - x^* \right\|_2 + \frac{\tau}{1 - \rho} \|\eta\|_2,$$

where

$$\rho := \frac{\sqrt{5} + 1}{2} \left(\left\lceil \frac{q}{k} \right\rceil + 1\right) \delta_{3k} \sqrt{\frac{1 + \delta_k}{1 - \delta_{2k}}} < 1 \tag{10}$$

is guaranteed under the condition (8), and the constant τ is given as

$$\tau = \frac{(\sqrt{5} + 1)\left(\left\lceil \frac{q}{k} \right\rceil + 1\right)(1 + \delta_{2k}) + 4}{2\sqrt{1 - \delta_{2k}}}. \tag{11}$$

In particular, if $\eta = 0$, the sequence $\{x^p\}$ generated by PGOT converges to x^*.

Proof Let $\eta = y - Ax^*$ and u^p, x^{p+1} be the vectors generated at p-th iteration of PGOT, i.e., $u^p = x^p + \mathcal{H}_q\left(A^T(y - Ax^p)\right)$ and $x^{p+1} = \mathcal{Z}_k^{\#}(u^p)$. Let $w^* \in \{0, 1\}^n$ be a k-sparse vector such that $\mathrm{supp}(x^*) \subseteq \mathrm{supp}(w^*)$. Applying Lemma 2 leads to

$$
\begin{aligned}
\left\| x^* - x^{p+1} \right\|_2 &= \left\| x^* - \mathcal{Z}_k^{\#}(u^p) \right\|_2 \\
&\leq \sqrt{\frac{1 + \delta_k}{1 - \delta_{2k}}} \left\| (x^* - u^p) \otimes w^* \right\|_2 + \frac{2}{\sqrt{1 - \delta_{2k}}} \|\eta\|_2 \\
&\leq \sqrt{\frac{1 + \delta_k}{1 - \delta_{2k}}} \left\| x^* - u^p \right\|_2 + \frac{2}{\sqrt{1 - \delta_{2k}}} \|\eta\|_2 \quad (\text{since } 0 \leq w^* \leq e). \tag{12}
\end{aligned}
$$

Denote by $\Omega := \mathcal{L}_q(A^T(y - Ax^p))$. It is easy to see that $|\mathrm{supp}(x^* - x^p)| \leq 2k$. Thus, if $q \geq 2k$, by Lemma 3, we have

$$
\begin{aligned}
\left\| x^* - u^p \right\|_2 &= \left\| x^* - x^p - \mathcal{H}_q\left(A^T(y - Ax^p)\right) \right\|_2 \\
&\leq \frac{\sqrt{5} + 1}{2} \left\| \left(x^* - x^p - A^T(y - Ax^p)\right)_{\Omega \cup (S \cup S^p)} \right\|_2, \tag{13}
\end{aligned}
$$

where $S = \text{supp}(x^*)$ and $S^p = \text{supp}(x^p)$. Given a vector $v \in \mathbb{R}^n$ and two support sets $\Lambda_1, \Lambda_2 \in [N]$. It is easy to verify $\left\| v_{\Lambda_1 \cup \Lambda_2} \right\|_2 \leq \left\| v_{\Lambda_1} \right\|_2 + \left\| v_{\Lambda_2} \right\|_2$. Therefore,

$$\| (x^* - x^p - A^T (y - Ax^p))_{\Omega \cup (S \cup S^p)} \|_2 \leq \left\| \left(x^* - x^p - A^T \left(y - Ax^p \right) \right)_{\Omega} \right\|_2$$
$$+ \left\| \left(x^* - x^p - A^T \left(y - Ax^p \right) \right)_{S \cup S^p} \right\|_2. \tag{14}$$

Noting that $|\,\text{supp}(x_S - x^p)| \leq |S \cup S^p| \leq 2k$. The second term on the right-hand side of (14) can be bounded. In fact, by Lemma 1, we have

$$\| (x^* - x^p - A^T (y - Ax^p))_{S \cup S^p} \|_2$$
$$= \left\| \left((I - A^T A)(x^* - x^p) + A^T \eta \right)_{S \cup S^p} \right\|_2$$
$$\leq \left\| \left((I - A^T A)(x^* - x^p) \right)_{S \cup S^p} \right\|_2 + \left\| (A^T \eta)_{S \cup S^p} \right\|_2$$
$$\leq \delta_{2k} \left\| x^* - x^p \right\|_2 + \sqrt{1 + \delta_{2k}} \left\| \eta \right\|_2. \tag{15}$$

Setting $t = \lceil \frac{q}{k} \rceil$, the set Ω can be separated into t disjoint sets such that $\Omega = T_1 \cup T_2 \ldots, T_t$, where $|T_i| \leq k$ for $i = 1, \ldots, t$, and $T_i \cap T_j = \emptyset$ if $i \neq j$. Thus we have

$$\| (x^* - x^p - A^T (y - Ax^p))_{\Omega} \|_2$$
$$\leq \sum_{i=1}^{t} \left\| \left(x^* - x^p - A^T \left(y - Ax^p \right) \right)_{T_i} \right\|_2$$
$$\leq \sum_{i=1}^{t} \left\| \left[(I - A^T A)(x^* - x^p) \right]_{T_i} \right\|_2 + \sum_{i=1}^{t} \left\| \left[A^T \eta \right]_{T_i} \right\|_2$$
$$\leq t\delta_{3k} \left\| x^* - x^p \right\|_2 + t\sqrt{1 + \delta_k} \left\| \eta \right\|_2, \tag{16}$$

where the last inequality follows from Lemma 1 because $|T_i \cup \text{supp}(x^* - x^p)| \leq 3k$. Since $\delta_{2k} \leq \delta_{3k}$, combining (13)–(16) leads to

$$\left\| x^* - u^p \right\|_2 \leq \frac{\sqrt{5}+1}{2}(t+1)\delta_{3k} \left\| x^* - x^p \right\|_2 + \frac{\sqrt{5}+1}{2}(t+1)\sqrt{1+\delta_{2k}} \left\| \eta \right\|_2. \tag{17}$$

Substituting (17) into (12) yields

$$\left\| x^{p+1} - x^* \right\|_2 \leq \rho \left\| x^p - x^* \right\|_2 + \tau \left\| \eta \right\|_2, \tag{18}$$

where ρ and τ are given as (10) and (11), respectively. Since $\delta_k \leq \delta_{2k} \leq \delta_{3k}$, the constant $\rho < 1$ is ensured if

$$\frac{\sqrt{5}+1}{2}(t+1)\delta_{3k}\sqrt{\frac{1+\delta_{3k}}{1-\delta_{3k}}} < 1. \tag{19}$$

Squaring both sides of (19) and rearranging terms yield

$$g(\delta_{3k}) := (t+1)^2\delta_{3k}^3 + (t+1)^2\delta_{3k}^2 + \frac{2}{3+\sqrt{5}}\delta_{3k} - \frac{2}{3+\sqrt{5}} < 0.$$

The gradient of g with respect to δ_{3k} is given as

$$\nabla g(\delta_{3k}) = 3(t+1)^2\delta_{3k}^2 + 2(t+1)^2\delta_{3k} + \frac{2}{3+\sqrt{5}} > 0.$$

Thus the function g is strictly and monotonically increasing over the interval $\delta_{3k} \in (0, 1]$. Note that

$$g(0) = -\frac{2}{3 + \sqrt{5}} < 0 \text{ and } g(1) = 2(t + 1)^2 > 0.$$

Thus there exists a unique real root α^* for the equation $g(\alpha^*) = 0$ in $[0, 1]$. Therefore, $\delta_{3k} < \alpha^*$ ensures that the constant $\rho < 1$ in (18), and hence it follows from (18) that

$$\|x^{p+1} - x^*\|_2 \le \rho^p \|x^0 - x^*\|_2 + \frac{\tau}{1 - \rho} \|\eta\|_2,$$

which is exactly the desired error bound. In particular, when $\eta = 0$, it follows immediately from the above error bound that the sequence $\{x^p\}$ generated by PGOT converges to x^* as $p \to \infty$. □

A more explicitly given RIC bound than (8) for PGOT can be derived as follows. Since $\sqrt{\frac{1+\delta_{3k}}{1-\delta_{3k}}} < \frac{1+\delta_{3k}}{1-\delta_{3k}}$, the inequality (19) is guaranteed provided the following inequality is satisfied:

$$\frac{\sqrt{5} + 1}{2} \left(\left\lceil \frac{q}{k} \right\rceil + 1 \right) \delta_{3k} \frac{1 + \delta_{3k}}{1 - \delta_{3k}} < 1,$$

which can be written as

$$\phi \delta_{3k}^2 + (\phi + 1) \delta_{3k} - 1 < 0, \tag{20}$$

where

$$\phi = \frac{\sqrt{5} + 1}{2} \left(\left\lceil \frac{q}{k} \right\rceil + 1 \right).$$

To guarantee (20), it is sufficient to require that

$$\delta_{3k} < \frac{-(\phi + 1) + \sqrt{\phi^2 + 6\phi + 1}}{2\phi} = \frac{2}{\sqrt{\phi^2 + 6\phi + 1} + \phi + 1}.$$

The right-hand side above is the positive root in $[0, 1]$ of the quadratic equation $\phi \delta_{3k}^2 + (\phi + 1) \delta_{3k} - 1 = 0$. From the above analysis, we immediately obtain the following result.

Corollary 1 *Let $x^* \in \mathbb{R}^n$ be the solution of the problem (7) and $\eta := y - Ax^*$. Let $q \ge 2k$ be a positive integer number. Suppose the restricted isometry constant of A satisfy*

$$\delta_{3k} < \frac{2}{\sqrt{\phi^2 + 6\phi + 1} + \phi + 1}, \tag{21}$$

where

$$\phi = \frac{\sqrt{5} + 1}{2} \left(\left\lceil \frac{q}{k} \right\rceil + 1 \right).$$

Then the sequence $\{x^p\}$ generated by PGOT satisfies that

$$\|x^{p+1} - x^*\|_2 \le \rho^p \|x^0 - x^*\|_2 + \frac{\tau}{1 - \rho} \|\eta\|_2,$$

where $\rho < 1$ and τ are given by (10) and (11), respectively.

The bound (21) depends only on the given integer number q. It is easy to verify that, for instance, $\delta_{3k} < 0.1517$ when $q = 2k$, $\delta_{3k} < 0.1211$ when $2k < q \leq 3k$, and $\delta_{3k} < 0.1009$ when $3k < q \leq 4k$.

Theorem 1 demonstrates how far the iterate point x^{p+1} generated by PGOT is from the solution x^* of the problem (7). It shows that the bound of $\|x^{p+1} - x^*\|_2$ depends on the value $\|\eta\|_2 = \|y - Ax^*\|_2$. It is worth mentioning that the sparse optimization problems are a large class of problems arising from many different applications, such as statistical regression, data sparse representation, data compression, channel estimation in wireless communication. The main results established in this section are general enough to apply to these broad situations. For instance, in sparse signal recovery, $y := Ax^* + \eta$ are the linear measurements of the signal x^*. In this case, $\|\eta\|_2 = \|y - Ax^*\|_2$ is the size of measurement error which would be small. In particular, $\eta = 0$ when measurements are accurate. For such a specific application, our main results can be further enhanced. See the discussion below in more detail.

3.1.1 Application to sparse signal recovery

Let x^* be a k-sparse signal to recover. To recover x^*, we take the signal measurements $y := Ax^* + \eta$ with a measurement matrix A, where $\eta = y - Ax^*$ denotes the measurement error which is small. Recovering x^* from the measurements y can be exactly modeled as the optimization problem (7). From the results in Sect. 3.1, we immediately obtain the next result concerning sparse signal recovery.

Theorem 2 *Let $y := Ax^* + \eta$ be the measurements of the k-sparse signal $x^* \in \mathbb{R}^n$ with measurement error η. Let $q \geq 2k$ be a positive integer number. Suppose the restricted isometry constant of measurement matrix A satisfies one of the following conditions:*

(i) $\delta_{3k} < \alpha^*$, where $\alpha^* \in (0, 1)$ is the unique real root of (9),
(ii) δ_{3k} satisfies (21).

Then the sequence generated by PGOT satisfies

$$\left\| x^{p+1} - x^* \right\|_2 \leq \rho^p \left\| x^0 - x^* \right\|_2 + \frac{\tau}{1 - \rho} \left\| \eta \right\|_2, \tag{22}$$

where ρ and τ are the same as (10) and (11), respectively. In particular, if the measurements are accurate, i.e., $y = Ax^*$, then the sequence $\{x^p\}$ generated by PGOT converges to x^*.

From (22), we see that when the measurements are accurate enough, i.e., $\|\eta\|_2$ is sufficient small, then $x^{p+1} \approx x^*$. This means the x^{p+1} is a high-quality reconstruction of x^*.

3.2 Main results for PGROT

Before analyzing the PGROT, we introduce the following lemma.

Lemma 4 *[38] Let $x^* \in \mathbb{R}^n$ be the solution to the problem (7) and $\eta = y - Ax^*$ be the error. Denote $S = \mathrm{supp}(x^*)$ and $S^{p+1} = \mathrm{supp}(x^{p+1})$. Let u^p and w^p be the vector defined as in PGROT, and $w^* \in \{0, 1\}^n$ be a binary k-sparse vector such that $S \subseteq \mathrm{supp}(w^*)$. Then*

$$\left\| \left(x^* - u^p \otimes w^p \right)_{S \cup S^{p+1}} \right\|_2 \leq \sqrt{\frac{1 + \delta_k}{1 - \delta_{2k}}} \left\| \left(x^* - u^p \right) \otimes w^* \right\|_2 + \frac{2}{\sqrt{1 - \delta_{2k}}} \left\| \eta \right\|_2$$

$$+ 2 \sqrt{\frac{1 + \delta_k}{1 - \delta_{2k}}} \left\| \mathcal{H}_k \left(u^p - x^* \right) \right\|_2.$$

Theorem 3 *Let $x^* \in \mathbb{R}^n$ be the solution to the problem (7) and $\eta = y - Ax^*$. Let $q \geq 2k$ be a positive integer number. Suppose the restricted isometry constant of matrix A satisfies*

$$\delta_{3k} < \beta^*, \tag{23}$$

where β^* is the unique real root of the equation

$$9\left(\left\lceil \frac{q}{k} \right\rceil + 1\right)^2 \beta^3 + 9\left(\left\lceil \frac{q}{k} \right\rceil + 1\right)^2 \beta^2 + \frac{2}{7 + 3\sqrt{5}}\beta - \frac{2}{7 + 3\sqrt{5}} = 0 \tag{24}$$

in $(0, 1)$. Then the sequence $\{x^p\}$ generated by PGROT satisfies

$$\left\| x^{p+1} - x^* \right\|_2 \leq \overline{\rho}^p \left\| x^0 - x^* \right\|_2 + \frac{\overline{\tau}}{1 - \overline{\rho}} \|\eta\|_2,$$

where

$$\overline{\rho} = 3\left(\frac{\sqrt{5}+1}{2}\right)^2 \left(\left\lceil \frac{q}{k} \right\rceil + 1\right) \delta_{3k} \sqrt{\frac{1 + \delta_k}{1 - \delta_{2k}}} < 1 \tag{25}$$

is ensured under (23), and the constant $\overline{\tau}$ is given as

$$\overline{\tau} = \left(\frac{\sqrt{5}+1}{2}\right)^2 \frac{3\left(\left\lceil \frac{q}{k} \right\rceil + 1\right)(1 + \delta_k)}{\sqrt{1 - \delta_{2k}}} + \frac{\sqrt{5}+1}{\sqrt{1 - \delta_{2k}}}. \tag{26}$$

In particular, if $\eta = 0$, then the sequence $\{x^p\}$ generated by PGROT converges to x^*.

Proof Let x^{p+1}, u^p and $\overline{w}^p$ be defined in PGROT. Denote by $S = \text{supp}(x^*)$. Note that $S^{p+1} = \text{supp}(x^{p+1}) = \text{supp}(\mathcal{H}_k(\overline{w}^p \otimes u^p))$. By Lemma 3, we have

$$\left\| x^* - x^{p+1} \right\|_2 = \left\| x^* - \mathcal{H}_k\left(\overline{w}^p \otimes u^p\right) \right\|_2 \leq \frac{\sqrt{5}+1}{2} \left\| (\mathbf{x}^* - \overline{\mathbf{w}}^{\mathbf{p}} \otimes \mathbf{u}^{\mathbf{p}})_{\mathbf{S} \cup \mathbf{S}^{\mathbf{p+1}}} \right\|_2. \tag{27}$$

Note that w^* is a k-sparse binary vector satisfying $\text{supp}(x^*) \subseteq \text{supp}(w^*)$. By Lemma 4, we obtain

$$\left\| \left(x^* - u^p \otimes \overline{w}^p\right)_{S \cup S^{p+1}} \right\|_2$$

$$\leq \sqrt{\frac{1 + \delta_k}{1 - \delta_{2k}}} \left(\left\| \left(x^* - u^p\right) \otimes w^* \right\|_2 + 2\left\| \mathcal{H}_k\left(u^p - x^*\right) \right\|_2\right) + \frac{2}{\sqrt{1 - \delta_{2k}}} \|\eta\|_2$$

$$\leq 3\sqrt{\frac{1 + \delta_k}{1 - \delta_{2k}}} \|\| x^* - u^p{}_2 + \frac{2}{\sqrt{1 - \delta_{2k}}} \|\eta\|_2, \tag{28}$$

where the last inequality follows from

$$\left\| \left(x^* - u^p\right) \otimes w^* \right\|_2 \leq \left\| \mathcal{H}_k\left(u^p - x^*\right) \right\|_2 \leq \left\| x^* - u^p \right\|_2.$$

Based on (17), we have

$$\left\| x^* - u^p \right\|_2 \leq \frac{\sqrt{5}+1}{2}(t + 1)\delta_{3k} \left\| x^* - x^p \right\|_2 + \frac{\sqrt{5}+1}{2}(t + 1)\sqrt{1 + \delta_k} \|\eta\|_2, \tag{29}$$

where $t = \left\lceil \frac{q}{k} \right\rceil$. Combining (27)–(29) yields

$$\left\| x^* - x^{p+1} \right\|_2 \leq \overline{\rho} \left\| x^* - x^p \right\|_2 + \overline{\tau} \|\eta\|_2 \tag{30}$$

where $\overline{\rho}$ and $\overline{\tau}$ are given by (25) and (26), respectively. We now prove that (23) implies $\overline{\rho} < 1$. Due to the fact $\delta_k \leq \delta_{2k} \leq \delta_{3k}$, to guarantee that $\overline{\rho} < 1$, it is sufficient to require

$$3 \left(\frac{\sqrt{5}+1}{2} \right)^2 (t+1)\delta_{3k} \sqrt{\frac{1+\delta_{3k}}{1-\delta_{3k}}} < 1, \tag{31}$$

which, by squaring both sides and rearranging terms, is equivalent to $g(\delta_{3k}) < 0$ where

$$g(\delta_{3k}) = 9(t+1)^2 \delta_{3k}^3 + 9(t+1)^2 \delta_{3k}^2 + \frac{2}{7+3\sqrt{5}} \delta_{3k} - \frac{2}{7+3\sqrt{5}}.$$

The gradient of $g(\delta_{3k})$ is given as

$$\nabla g(\delta_{3k}) = 27(t+1)^2 \delta_{3k}^2 + 18(t+1)^2 \delta_{3k} + \frac{2}{7+3\sqrt{5}},$$

which is positive over the interval $[0, 1]$. This together with

$$g(0) = -\frac{2}{7+3\sqrt{5}} < 0, \; g(1) = 18(t+1)^2 > 0,$$

implies that there exists a unique real positive root $\beta^* \in (0, 1)$ satisfying $g(\beta^*) = 0$. Therefore, the condition $\delta_{3k} < \beta^*$ guarantees the inequality (31), and thus ensures that $\overline{\rho} < 1$. Thus it follows from (30) that

$$\left\| x^{p+1} - x^* \right\|_2 \leq \overline{\rho}^p \left\| x^0 - x^* \right\|_2 + \frac{\overline{\tau}}{1-\overline{\rho}} \|\eta\|_2.$$

When $\eta = 0$, the relation above implies that $\left\| x^* - x^{p+1} \right\|_2 \leq \overline{\rho}^p \left\| x^* - x^0 \right\|_2 \to 0$ as $p \to \infty$. Therefore, the sequence $\{x^p\}$ generated by RPGOT in this case converges to the solution x^* of (7). □

Similar to the discussion in the end of Sect. 3.1, an explicit bound of δ_{3k} for PGROT can be given. Since $\sqrt{\frac{1+\delta_{3k}}{1-\delta_{3k}}} < \frac{1+\delta_{3k}}{1-\delta_{3k}}$, a sufficient condition for (31) is

$$3 \left(\frac{\sqrt{5}+1}{2} \right)^2 \left(\left\lceil \frac{q}{k} \right\rceil + 1 \right) \delta_{3k} \frac{1+\delta_{3k}}{1-\delta_{3k}} < 1,$$

which is equivalent to

$$\psi \delta_{3k}^2 + (\psi + 1)\delta_{3k} - 1 < 0,$$

where

$$\psi = 3 \left(\frac{\sqrt{5}+1}{2} \right)^2 \left(\left\lceil \frac{q}{k} \right\rceil + 1 \right).$$

By the same analysis in Sect. 3.1, we immediately have the next corollary.

Corollary 2 *Let $x^* \in \mathbb{R}^n$ be the solution to the problem (7) and $\eta := y - Ax^*$. Let $q \geq 2k$ be a positive integer number. Suppose the restricted isometry constant of matrix A satisfies*

$$\delta_{3k} < \frac{2}{\sqrt{\psi^2 + 6\psi + 1} + \psi + 1},$$

where

$$\psi = 3 \left(\frac{\sqrt{5}+1}{2} \right)^2 \left(\left\lceil \frac{q}{k} \right\rceil + 1 \right).$$

Then the sequence $\{x^p\}$ generated by PGROT satisfies

$$\left\| x^{p+1} - x^* \right\|_2 \le (\overline{\rho})^p \left\| x^0 - x^* \right\|_2 + \frac{\overline{\tau}}{1 - \overline{\rho}} \| \eta \|_2,$$

where $\overline{\rho}$ and $\overline{\tau}$ are given as (25) and (26), respectively.

Similar to Corollary 1, we may apply the above result (Theorem 3 and Corollary 2) to the scenario of sparse signal recovery via compressed sensing for which $\eta = y - Ax^*$ is very small, and thus $x^p \approx x^*$ when p is large enough. That is, the iterate x^p generated by PGROT is a quality approximation to the signal.

3.3 Main result for PGROTP

The error bound for the solution of (7) via PGROTP algorithm can be also established. The next lemma concerning a property of pursuit step is useful in this analysis.

Lemma 5 *[38] Let $x^* \in \mathbb{R}^n$ be the solution to the problem (7) and $\eta = y - Ax^*$. The vector $u \in \mathbb{R}^n$ is an arbitrary k-sparse vector. Then the optimal solution of the pursuit step*

$$z^* = \arg \min_z \left\{ \| y - Az \|_2^2 : \ \operatorname{supp}(z) \subseteq \operatorname{supp}(u) \right\}$$

satisfies that

$$\left\| z^* - x^* \right\|_2 \le \frac{1}{\sqrt{1 - (\delta_{2k})^2}} \| x^* - u \|_2 + \frac{\sqrt{1 + \delta_k}}{1 - \delta_{2k}} \| \eta \|_2.$$

The main result for PGROTP algorithm is given as follows.

Theorem 4 *Let $x^* \in \mathbb{R}^n$ be the solution to the problem (7) and $\eta = y - Ax^*$. Let $q \ge 2k$ be a positive integer number. Suppose the restricted isometry constant of matrix A satisfies*

$$\delta_{3k} < \frac{1}{3 \left(\frac{\sqrt{5}+1}{2} \right)^2 \left(\left\lceil \frac{q}{k} \right\rceil + 1 \right) + 1}. \tag{32}$$

Then the sequence $\{x^p\}$ generated by PGROTP satisfies

$$\left\| x^{p+1} - x^* \right\|_2 \le (\rho^*)^p \left\| x^0 - x^* \right\|_2 + \frac{\tau^*}{1 - \rho^*} \| \eta \|_2, \tag{33}$$

where

$$\rho^* = \left(\frac{\sqrt{5}+1}{2} \right)^2 \frac{3 \left(\left\lceil \frac{q}{k} \right\rceil + 1 \right) \delta_{3k}}{1 - \delta_{3k}} < 1 \tag{34}$$

is guaranteed by (32), and the constant τ^* is given as

$$\tau^* = \frac{3 \left(\frac{\sqrt{5}+1}{2} \right)^2 \left(\left\lceil \frac{q}{k} \right\rceil + 1 \right)(1 + \delta_k) + \sqrt{5} + 1}{(1 - \delta_{2k})\sqrt{1 + \delta_{2k}}} + \frac{\sqrt{1 + \delta_k}}{1 - \delta_{2k}}. \tag{35}$$

In particular, when $\eta = 0$, the sequence $\{x^p\}$ generated by PGROTP converges to x^*.

Table 1 The upper bounds of δ_{3k} for several different q

The value of q	PGOT	RPGOT	RPGOTP
$q = 2k$	$\delta_{3k} < 0.1729$	$\delta_{3k} < 0.0407$	$\delta_{3k} < 0.0407$
$2k < q \leq 3k$	$\delta_{3k} < 0.1348$	$\delta_{3k} < 0.0308$	$\delta_{3k} < 0.0308$
$3k < q \leq 4k$	$\delta_{3k} < 0.1106$	$\delta_{3k} < 0.0248$	$\delta_{3k} < 0.0248$

Proof The PGROTP can be regarded as a combination of PGROT with a pursuit step. Denote $\overline{x}^{p+1}$ as the intermediate vector generated by PGROT. Based on the analysis of PGROT algorithm, we have

$$\|\overline{x}^{p+1} - x^*\|_2 \leq \overline{\rho} \|x^* - x^p\|_2 + \overline{\tau} \|\eta\|_2 ,$$

where $\overline{\rho}$ and $\overline{\tau}$ are the same as (25) and (26), respectively. By Lemma 5, we immediately have that

$$\|x^* - x^{p+1}\|_2 \leq \frac{1}{\sqrt{1 - (\delta_{2k})^2}} \|\overline{x}^{p+1} - x^*\|_2 + \frac{\sqrt{1 + \delta_k}}{1 - \delta_{2k}} \|\eta\|_2 .$$

As $\delta_k \leq \delta_{2k} \leq \delta_{3k}$, combining two inequalities above yields

$$\|x^* - x^{p+1}\|_2 \leq \rho^* \|x^* - x^p\|_2 + \tau^* \|\eta\|_2 ,$$

where ρ^* and τ^* are given in (34) and (35), respectively. To guarantee $\rho^* < 1$, i.e.,

$$\left(\frac{\sqrt{5}+1}{2}\right)^2 \frac{3 \left(\lceil \frac{q}{k} \rceil + 1\right) \delta_{3k}}{1 - \delta_{3k}} < 1,$$

it is sufficient to require that

$$\delta_{3k} < \frac{1}{3 \left(\frac{\sqrt{5}+1}{2}\right)^2 \left(\lceil \frac{q}{k} \rceil + 1\right) + 1},$$

which is exactly the assumption (33) of the theorem.

If $\eta = 0$, the sequence $\{x^p\}$ generated by PGROTP converges to x^* as $p \to \infty$, since in this case, (33) is reduced to $\|x^{p+1} - x^*\|_2 \leq (\rho^*)^p \|x^0 - x^*\|_2$. $\qquad\square$

For sparse signal recovery, similar comments to that of Sect. 3.1.1 can be made to PGROTP. The discussion is omitted here. Before we close this section, we list a few RIC conditions in terms of δ_{3k} for the proposed algorithms with different q, i.e., $q = 2k, 3k, 4k$. The results shown in Table 1 are derived based on (9), (24) and (32) for the q given as above, respectively. It is worth mentioning that, when $q = n$, the partial gradient $\mathcal{H}_q (\nabla f(x))$ becomes the full gradient $\nabla f(x)$, and the algorithms in this paper are reduced to the optimal k-thresholding (OT), relaxed optimal k-thresholding (ROT) and relaxed optimal k-thresholding pursuit (ROTP) algorithm, respectively. The sufficient conditions for the convergence of these algorithms were studied in [38, 42].

Remark 1 The iterative complexity of the interior point method for ROTP is about $O\left(pm^3 + pmn + pn^{3.5}L\right)$, as pointed out in [42], where p is the number of iterations and L is the size of the problem data encoding in binary. Solving quadratic optimization is the main

(a) Sparsity level $k = 162$

(b) Sparsity level $k = 197$
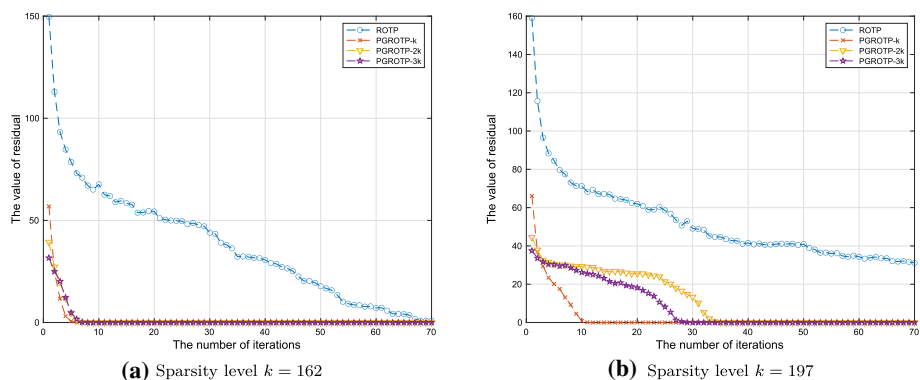
Fig. 1 Objective change in the course of iterations for PGROTP with different q, i.e., $q = k, 2k, 3k, n$

cost for the ROTP algorithm. The algorithms proposed in this paper possess similar a computational complexity. To improve the chance for the solution of (ROT) to be binary or nearly binary, two or more relaxed optimal k-thresholding steps are employed by the algorithms in [38]. The partial-gradient-based algorithm in this paper is quite different from the strategy in [38]. The iterate generated by partial gradient methods is at most $(k + q)$-sparse, based on which the vector derived from the relaxed k-optimal thresholding is more compressible than the one generated by those algorithms in [38].

4 Numerical experiments

Simulations via synthetic data are carried out to demonstrate the numerical performance of the PGROTP which is the main implementable algorithm proposed in this paper. We test the algorithm from four aspects: objective reduction, average number of iterations required for solving the problem (7), success frequency in vector reconstruction and phase transition in terms of success rate. The PGROTP with $q = k, 2k, 3k$ and n are tested and compared. Unless otherwise stated, the measurement matrices used in experiments are Gaussian random matrices whose entries follow the standard normal distribution $\mathcal{N}(0, 1)$. For sparse vectors, their entries also follow the $\mathcal{N}(0, 1)$ and the position of nonzero entries of the vector follows the uniform distribution. All involved convex optimization problems were solved by CVX developed by Grant and Boyd [25] with solver 'Mosek' [1].

4.1 Objective reduction

This experiment is used to investigate the objective-reduction performance of the PGROTP with different q, including $q = k, 2k, 3k$ and n. We set $A \in \mathbb{R}^{500 \times 1000}$ and $y = Ax^*$, where x^* is a generated k-sparse vector. Thus x^* is a global solution of the problem (7). Figure 1 records the changes of the objective value $\|y - Ax\|_2$ in the course of algorithm up to 70 iterations. Figure 1a, b include the results for the sparsity level $\|x^*\|_0 = 162$ and 197, respectively. It can be seen that PGROTP is able to reduce the objective value during iterations. Moreover, this experiment also indicates that the optimal k-thresholding methods with partial gradients may perform better in objective reduction than the ones using full gradients in many situations.

(a) $m = 300$

(b) $m = 400$

(c) $m = 500$

(d) $m = 600$
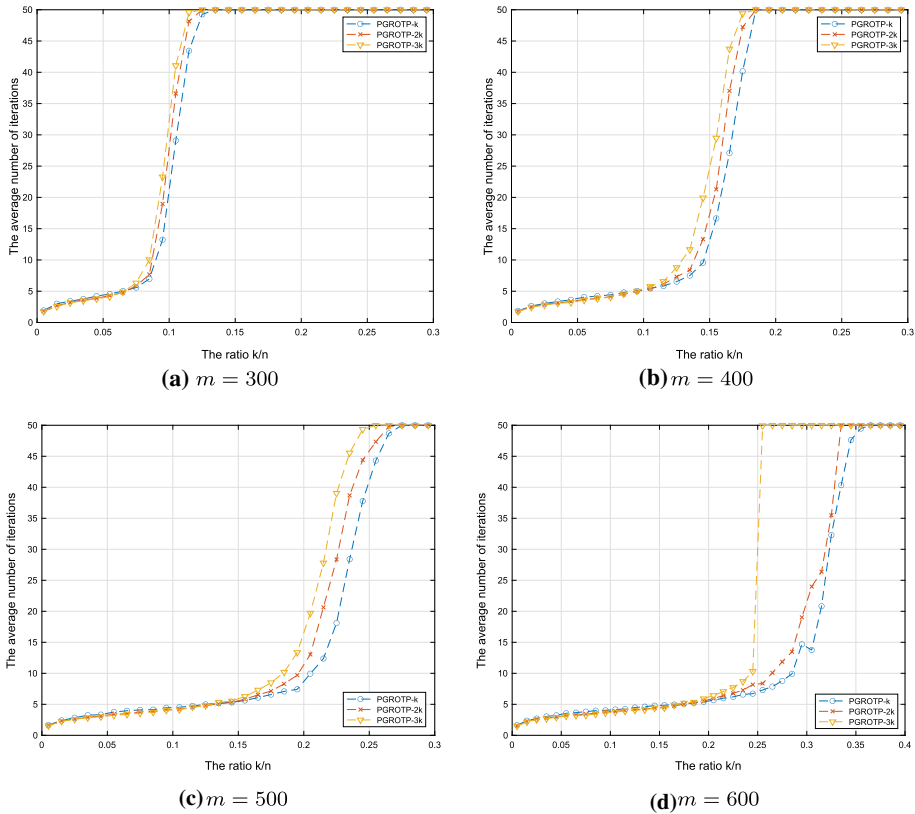
Fig. 2 Comparison of the average number of iterations required by PGROTP with different q

4.2 Number of iterations

Experiments were also performed to demonstrate the average number of iterations needed for PGROTP to solve the sparse optimization problems from the sparse vector reconstruction. The vector dimension is fixed to be 1000, and the size of the measurement matrix is $m \times 1000$, where m takes the following a few different values: $m = 300, 400, 500, 600$. The measurements $y = Ax^*$ are accurate, where x^* is the sparse vector to recover. In this experiment, if the iterate x^p satisfies the criterion

$$\|x - x^*\|_2 / \|x^*\|_2 \leq 10^{-3}, \tag{36}$$

then the algorithm terminates and the number of iterations p is recorded. If the algorithm within 50 iterations cannot meet the criterion (36), then the algorithm stops, and the number of iterations performed is recorded as 50. For each given sparsity level, the average number of iterations is obtained by attempting 50 trials. The outcomes are shown in Fig. 2 which indicate that the required iterations of PGROTP for vector reconstructions are usually low when the sparsity level of x^* is low, and that the number of iterations required for solving the problem increases as the sparsity level increases. This figure also shows that for a given sparsity level, the more measurements are required, the lower the average number of iterations are needed by the PGROTP to meet the reconstruction criterion (36).
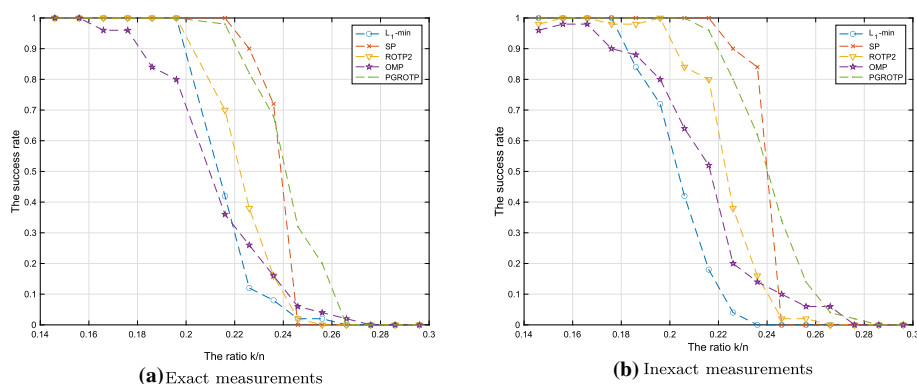
Fig. 3 Comparisons of success rates of sparse signal reconstruction between several algorithms via Gaussian random matrices

4.3 Sparse signal recovery

Simulations were also carried out to compare the success rates of the PGROTP algorithm in sparse vector reconstruction with several existing algorithms, such as ℓ_1-minimization, subspace pursuit (SP), orthogonal matching pursuit (OMP) and ROTP2 (in [38, 42]). The size of the measurement matrix is still 500×1000. For every given ratio of the sparsity level k and n, the success rate of the algorithm is obtained by 50 random attempts. In this experiment, SP, ROTP2 and PGROTP($q = k$) perform a total of 50 iterations, whereas OMP is performed k iterations. After performing the required number of iterations, the algorithm is counted as success if the condition (36) is satisfied. The success rates for accurate and inaccurate measurements are summarized in Fig. 3a, b, respectively. The inaccurate measurements are given by $y = Ax^* + 0.001\eta$, where η is a standard Gaussian vector. Compared with several existing algorithms, it can be seen that the PGROTP is robust and efficient for the sparse vector reconstruction in both noise and noiseless environments.
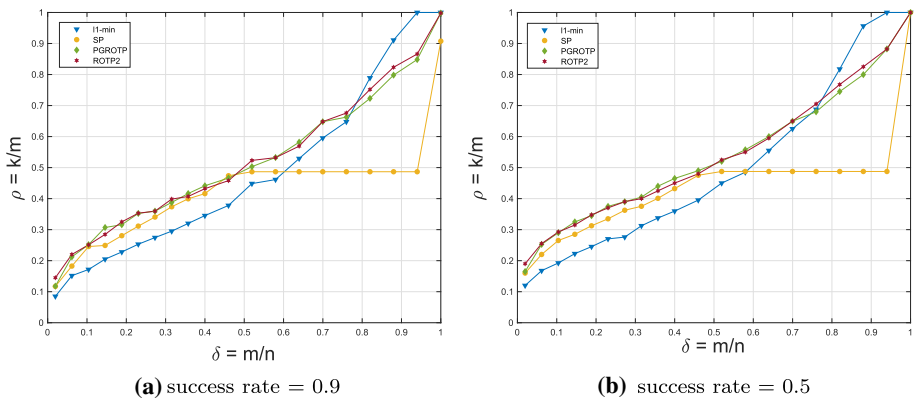
4.4 Phase transition

Phase transition analysis was introduced by Donoho and Tanner [19]. Blanchard and Tanner [2] and Blanchard, Tanner and Wei [3] provided substantial analyses and experiments to compare the phase transition features of several compressed sensing algorithms. We also conducted such experiments to demonstrate the phase transition of PGROTP($q = k$) and compare with ℓ_1-minimization, SP and ROTP2 algorithms.

In this experiment, the length of the target vector x^* is fixed as $n = 2048$. We take 40 different sparsity levels uniformly from 1 to m, i.e., $k = \lfloor \frac{1}{40}m \rfloor, \ldots, \lfloor \frac{39}{40}m \rfloor, m$. For the number of measurements m, we require $m \geq 40$ to ensure that there are 40 different sparsity levels. We take a total of 20 different values for the measurements, 10 uniformly from $[40, 0.4n]$ and 10 from $(0.4n, n]$. The matrix A is a Gaussian random matrix whose entries follow the normal distribution $\mathcal{N}(0, 1/\sqrt{m})$. The measurements of x^* are inexact, i.e., $y = Ax^* + 10^{-5}\eta$, where η is a standard Gaussian vector. The iterative algorithms terminate either when the criterion (36) is met, or the maximum of 30 iterations are reached. The algorithm is counted as success when (36) is satisfied. For each pair of (m, k), we use 10 trials to calculate the success rate and the average time consumption for success recoveries.
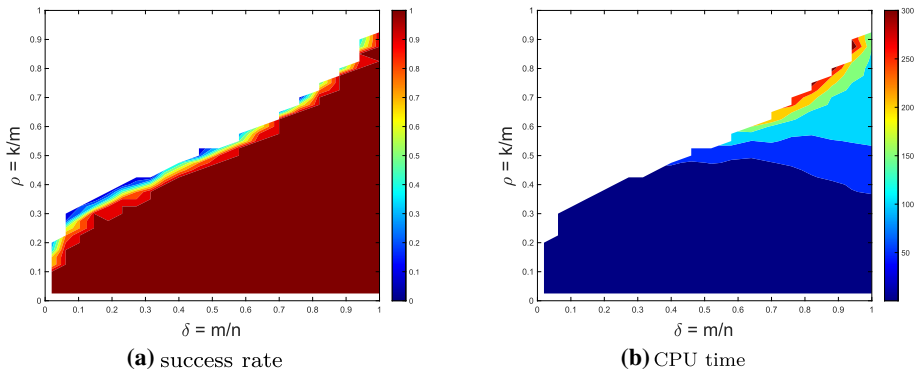
Fig. 4 Phase transition of PGROTP with respect to success rate and CPU time under inexact measurements


Fig. 5 Comparison of phase transition curves of ℓ_1-minimization, SP, PGROTP and ROTP2 algorithms under inexact measurements in terms of success rates

In Figs. 4, 5, 6, the vertical axis represents over-sampling rate $\rho = k/m$, while the horizontal axis represents under-sampling rate $\delta = m/n$. The range of both ρ and δ is between 0 and 1.

Figure 4 demonstrates the phase transition regions of PGROTP in terms of success rates and CPU time for recovery success, respectively. The shaded area in Fig. 4a indicates the region for the recovery success of the algorithm. We can see from Fig. 4a that PGROTP has ability to recover the signal with relatively high δ and ρ. Figure 4b shows that PGROTP takes more time to achieve recovery success when more measurements are used and when the signal is relatively dense.

We now compare the phase transition curves between ℓ_1-minimization, SP, PGROTP and ROTP2 algorithms. We carry out the experiment in the same environment as that of Fig. 4. For every given δ, the logistic regression model is used to determine the specific success rate corresponding to ρ. The phase transition curves can be obtained in terms of different level of success rates. Figure 5a, b demonstrate the curves for the success rates 0.9 and 0.5, respectively. Such curves delineate the regions where the algorithms can recover the target signal with a specific success rate and the region in which they cannot. In general, the higher the curve, the better the performance of the algorithm. Figure 5 indicates that PGROTP performs comparably to the other three algorithms. We also see that although PGROTP
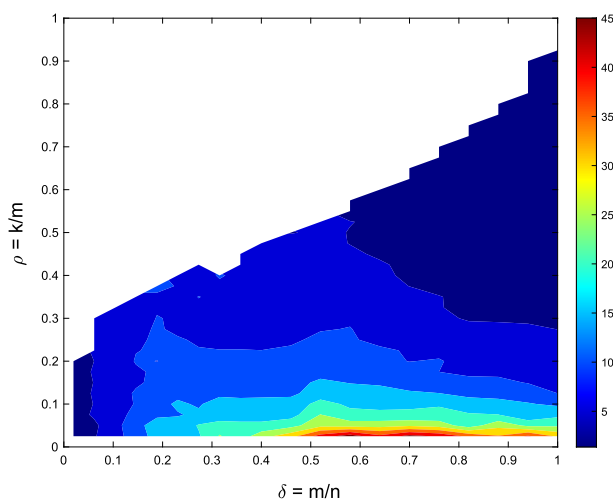
Fig. 6 Time consumption ratio of PGROTP and ROTP2

adopts a partial gradient direction, it maintains a similar performance to ROTP2 in sparse signal recovery.

Finally, we compare the performance of PGROTP and ROTP2 via time consumption ratio. Figure 6 shows the time ratios $r = \frac{\text{time(ROTP2}(\delta, \rho))}{\text{time(PGROTP}(\delta, \rho))}$ over the phase transition region. The difference in time consumption between the two algorithms is more pronounced when more measurements are used to recover a sparse signal. Figure 6 clearly shows that when dealing with signals with low sparsity levels, the PGROTP algorithm can recover wide range of signals with less time than ROTP2. This does show that using the partial gradient information may help reduce the runtime of the optimal k-thresholding method.

5 Conclusions and final remarks

Motivated by the recent optimal k-thresholding technique, we proposed the partial-gradient-based optimal k-thresholding methods for solving a class of sparse optimization problems. Under the restricted isometry property, we established a global error bound for the iterates produced by our algorithms. For sparse signal recovery, our results claim that the proposed algorithms with q satisfying $2k \leq q < n$ are guaranteed to recover the sparse vector. The optimal-thresholding-type algorithms usually require more time to solve the sparse optimization problem than traditional hard thresholding methods and existing heuristic algorithms. The primary computation time is solving quadratic optimization problems. However, the advantage of optimal-thresholding-type algorithms is that they are more stable, robust and have stronger reconstruction capability than existing heuristic methods. Although we focus on solving the specific model (7) in this paper, it is not difficult to extend the framework of the proposed algorithms to the general model (1). We leave this as future work.

References

1. Andersen, E. D., Andersen, K. D.: The MOSEK interior point optimizer for linear programming: an implementation of the homogeneous algorithm. High performance optimization. Springer, Boston, MA, 33, 197-232 (2000)
2. Blanchard, J.D., Tanner, J.: Performance comparisons of greedy algorithms in compressed sensing. Numer. Linear Algebra Appl. **22**(2), 254–282 (2015)
3. Blanchard, J.D., Tanner, J., Wei, K.: CGIHT: conjugate gradient iterative hard thresholding for compressed sensing and matrix completion. Inform. Inference J. IMA **4**(4), 289–327 (2015)
4. Blumensath, T.: Accelerated iterative hard thresholding. Signal Process. **92**, 752–756 (2012)
5. Blumensath, T., Davies, M.: Iterative hard thresholding for sparse approximation. J. Fourier Anal. Appl. **14**, 629–654 (2008)
6. Blumensath, T., Davies, M.: Iterative hard thresholding for compressed sensing. Appl. Comput. Harmon. Anal. **27**, 265–274 (2009)
7. Blumensath, T., Davies, M.: Normalized iterative hard thresholding: guaranteed stability and performance. IEEE J. Sel. Top. Signal Process. **4**, 298–309 (2010)
8. Boche, H., Calderbank, R., Kutyniok, G., Vybiral, J.: Compressed Sensing and its Applications. Springer, New York (2019)
9. Bouchot, J., Foucart, S., Hitczenki, P.: Hard thresholding pursuit algorithms: number of iterations. Appl. Comput. Harmon. Anal. **41**, 412–435 (2016)
10. Cai, T., Wang, L.: Orthogonal matching pursuit for sparse signal recovery with noise. IEEE Trans. Inform. Theory **57**(7), 4680–4688 (2011)
11. Candès, E., Tao, T.: Decoding by linear programming. IEEE Trans. Inform. Theory **51**(12), 4203–4215 (2005)
12. Candès, E., Romberg, J., Tao, T.: Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. IEEE Trans. Inform. Theory **52**(2), 489–509 (2006)
13. Candès, E., Wakin, M., Boyd, S.: Enhancing sparsity by reweighted ℓ_1-minimization. J. Fourier Anal. Appl. **14**, 877–905 (2008)
14. Chen, S., Donoho, D., Saunders, M.: Atomic decomposition by basis pursuit. SIAM Rev. **43**(1), 129–159 (2001)
15. Choi, J., Shim, B., Ding, Y., Rao, B., Kim, D.: Compressed sensing for wireless communications: useful tips and tricks. IEEE Commun. Surv. Tutor. **19**(3), 1527–1549 (2017)
16. Dai, W., Milenkovic, O.: Subspace pursuit for compressive sensing: Closing the gap between performance and complexity. ILLINOIS UNIV AT URBANA-CHAMAPAIGN (2008)
17. Dai, W., Milenkovic, O.: Subspace pursuit for compressive sensing signal reconstruction. IEEE Trans. Inform. Theory **55**, 2230–2249 (2009)
18. Donoho, D.: De-noising by soft-thresholding. IEEE Trans. Inform. Theory **41**, 613–627 (1995)
19. Donoho, D., Tanner, J.: Observed universality of phase transitions in high-dimensional geometry, with implications for modern data analysis and signal processing. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. **367**(1906), 4273–4293 (2009)
20. Eldar, Y., Kutyniok, G.: Compressed Sensing: Theory and Applications. Cambridge University Press, Cambridge (2012)
21. Fornasier, M., Rauhut, H.: Iterative thresholding algorithms. Appl. Comput. Harmon. Anal. **25**(2), 187–208 (2008)
22. Foucart, S., Rauhut, H.: A Mathematical Introduction to Compressive Sensing. Birkhäuser Basel (2013)
23. Foucart, S.: Hard thresholding pursuit: an algorithm for compressive sensing. SIAM J. Numer. Anal. **49**(6), 2543–2563 (2011)

24. Garg, R., Khandekar, R.: Gradient descent with sparsification: an iterative algorithm for sparse recovery with restricted isometry property. In: Proceedings of the 26th annual international conference on machine learning (2009)
25. Grant, M., Boyd, S.: *CVX: matlab software for disciplined convex programming*. Version 1.21, (2017)
26. Huang, G., Wang, L.: Soft-thresholding orthogonal matching pursuit for efficient signal reconstruction. In: Proceedings of the 2013 IEEE international conference on acoustics, speech and signal processing. 2543–2547 (2013)
27. Khanna, R., Kyrillidis, A.: IHT dies hard: Provable accelerated iterative hard thresholding. Int. Conf. Artif. Intell. Stat. 188–198 (2018)
28. Liu, Y., Zhan, Z., Cai, J.F., et al.: Projected iterative soft-thresholding algorithm for tight frames in compressed sensing magnetic resonance imaging. IEEE Trans. Med. Imag. **35**(9), 2130–2140 (2016)
29. Meng, N., Zhao, Y.-B.: Newton-type optimal thresholding algorithms for sparse optimization problems. arXiv preprint arXiv:2104.02371 (2021)
30. Meng, N., Zhao, Y.-B.: Newton-step-based hard thresholding algorithms for sparse signal recovery. IEEE Trans. Signal Process. **68**, 6594–6606 (2020)
31. Needell, D., Tropp, J.: CoSaMP: iterative signal recovery from incomplete and inaccurate samples. Appl. Comput. Harmon. Anal. **26**, 301–321 (2009)
32. Needell, D., Vershynin, R.: Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. IEEE J. Sel. Top. Signal Process. **4**(2), 310–316 (2010)
33. Patel, V., Chellappa, R.: Sparse representations, compressive sensing and dictionaries for pattern recognition. The First Asian Conference on Pattern Recognition, IEEE. 325–329 (2011)
34. Tropp, J., Gilbert, A.: Signal recovery from random measurements via orthogonal mathcing pursuit. IEEE Trans. Inform. Theory **53**(12), 4655–4666 (2007)
35. Yuan, X.-T., Liu, Q.: Newton greedy pursuit: A quadratic approximation method for sparsity-constrained optimization. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 4122–4129 (2014)
36. Zhao, Y.-B., Luo, Z.-Q.: Improved Rip-based bounds for guaranteed performance of two compressed sensing algorithms. arXiv:2007.01451v3 (2020)
37. Zhao, Y.-B.: Sparse Optimization Theory and Methods. CRC Press, Boca Raton, FL (2018)
38. Zhao, Y.-B.: Optimal k-thresholding algorithms for sparse optimization problems. SIAM J. Optim. **30**(1), 31–55 (2020)
39. Zhao, Y.-B., Kočvara, M.: A new computational method for the sparsest solutions to systems of linear equations. SIAM J. Optim. **25**(2), 1110–1134 (2015)
40. Zhao, Y.-B., Li, D.: Reweighted ℓ_1-minimization for sparse solutions to underdetermined linear systems. SIAM J. Optim. **22**(3), 1065–1088 (2012)
41. Zhao, Y.-B., Luo, Z.-Q.: Constructing new reweighted ℓ_1-algorithms for sparsest points of polyhedral sets. Math. Oper. Res. **42**, 57–76 (2017)
42. Zhao, Y.-B., Luo, Z.-Q.: Analysis of optimal thresholding algorithms for compressed sensing. Signal Process. **187**, 108148 (2021)
43. Zhou, S., Pan, L., Xiu, N.: Subspace Newton method for the ℓ_0-regularized optimization. arXiv:2004.05132 (2020)
44. Zhou, S., Xiu, N., Qi, H.: Global and quadratic convergence of Newton hard-thresholding pursuit. J. Mach. Learn. Res. **22**(12), 1–45 (2021)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.