



Quadratic Growth Conditions and Uniqueness of Optimal Solution to Lasso

Yunier Bello-Cruz¹ · Guoyin Li² · Tran Thai An Nghia³

Received: 10 January 2020 / Accepted: 18 January 2022 / Published online: 25 March 2022
© The Author(s) 2022

Abstract

In the previous paper Bello-Cruz et al. (J Optim Theory Appl 188:378–401, 2021), we showed that the quadratic growth condition plays a key role in obtaining Q-linear convergence of the widely used forward–backward splitting method with Beck–Teboulle’s line search. In this paper, we analyze the property of quadratic growth condition via second-order variational analysis for various structured optimization problems that arise in machine learning and signal processing. This includes, for example, the Poisson linear inverse problem as well as the ℓ_1 -regularized optimization problems. As a by-product of this approach, we also obtain several full characterizations for the uniqueness of optimal solution to Lasso problem, which complements and extends recent important results in this direction.

Keywords Nonsmooth and convex optimization problems · Forward–backward splitting method · Linear convergence · Uniqueness · Lasso · Quadratic growth condition · Variational analysis

Mathematics Subject Classification 65K05 · 90C25 · 90C30

Communicated by Liqun Qi.

This work was partially supported by the National Science Foundation (NSF) Grant DMS - 1816386 and 1816449, and a Discovery Project from Australian Research Council (ARC), DP190100555.

✉ Guoyin Li
g.li@unsw.edu.au
Yunier Bello-Cruz
yunierbello@niu.edu
Tran Thai An Nghia
nttran@oakland.edu

¹ Department of Mathematical Sciences, Northern Illinois University, DeKalb, IL 60115, USA

² Department of Applied Mathematics, University of New South Wales, Sydney 2052, Australia

³ Department of Mathematics and Statistics, Oakland University, Rochester, MI 48309, USA

1 Introduction

This paper is a continuation of our previous work [8] at which we studied convergence properties of the forward–backward splitting method (FBS in brief, also known as the proximal gradient method). The FBS [5, 7, 12, 14, 15, 28, 37] is a simple and efficient method for solving an optimization problem whose objective function is the sum of two convex functions: one of which is differentiable in its domain, and the other one is proximal-friendly (that is, its proximal mapping can be easily computed) and can be non-differentiable. It is well known that FBS is globally convergent to an optimal solution with the complexity $o(k^{-1})$ [7, 9, 19, 40] in general settings. Linear convergence for FBS has been studied in many papers via Kurdya–Łojasiewicz inequality [10, 25, 26, 31, 49] or error bound conditions [21, 38, 41, 53] with the base from [34]. Without assuming the usual condition that the gradient of the differentiable function involved is globally Lipschitz continuous, our previous paper [8] studied convergence properties and the complexity of FBS method with Beck–Teboulle’s line search. In particular, under the so-called *quadratic growth condition* also known as *2-conditioned property*, which is close to the idea in [21, 25, 26], we showed that the sequence generated by the FBS with Beck–Teboulle’s line search is Q-linearly convergent. Our derived linear rates complement and sometimes improve those in [21, 25, 26].

One of the main aims of this paper is to analyze the quadratic growth condition for several structured optimization problems. This allows us to understand the performance of FBS methods for solving specific optimization problems by considering the specific structure of the problems. In particular, we show that the quadratic growth condition is automatically satisfied for the standard Poisson inverse regularized problems with Kullback–Leibler divergence [16, 47], which does not satisfy the usual global Lipschitz continuous assumption mentioned above. Using FBS to solve Poisson inverse regularized problems was first proposed in [4] via the idea of Bregman divergence. Recently, Salzo [40] proved that the FBS method with an appropriate line search enjoys a complexity of $o(k^{-1})$ when it is applied to solve Poisson inverse regularized problems. In this paper, we advance this direction by showing that the convergence rate of FBS method with Beck–Teboulle’s line search is indeed Q-linear in solving Poisson inverse regularized problems.

It is worth noting that linear convergence of the sequence generated by the FBS in solving some structured optimization problems was also studied in [12, 28, 32, 33, 41] when the nonsmooth function is *partly smooth relative to a manifold* by using the idea of finite support identification. The latter notion introduced by Lewis [29] allows Liang, Fadili, and Peyré [32, 33] to cover many important problems such as the total variation semi-norm, the ℓ_1 -norm, the ℓ_∞ -norm, and the nuclear norm problems. In their paper, a second-order condition was introduced to guarantee the Q-local linear convergence of the FBS sequence under the non-degeneracy assumption [29]. When considering the ℓ_1 -regularized problem, we are able to avoid the non-degeneracy assumption. Under the setting of [8], this allows us to improve the well-known work of Hale, Yin, and Zhang [28] in two aspects: (a) We completely drop the aforementioned non-degeneracy assumption. (b) Our second-order condition is strictly weaker than the one in [28, Theorem 4.10]. The wider view is that when considering particular optimization

problems listed in the spirit of [32, 33], the assumption of non-degeneracy may not be necessary. Furthermore, we revisit the *iterative shrinkage thresholding algorithm* (ISTA) [7, 18], which is indeed FBS for solving Lasso problem [42]. It is well known that the complexity of this algorithm is $\mathcal{O}(k^{-1})$; however, recent works [32, 41] indicate the local linear convergence of ISTA. The stronger conclusion in this direction is obtained lately by Bolte, Nguyen, Peypouquet, and Suter [10, 25] that: the iterative sequence of ISTA is R -linearly convergent, and its corresponding cost sequence is globally Q -linearly convergent, but the rate may depend on the initial point. Inspired by these achievements, we provide two new information under the setting of [8]: (c) The iterative sequence of ISTA is indeed *globally* Q -linearly convergent. (d) The iterative sequence of ISTA is eventually Q -linearly convergent to an optimal solution with a uniform rate that does not depend on the initial point.

In order to obtain the linear convergence of ISTA, several papers make the assumption that the optimal solution to Lasso is unique; see, e.g., [12, 24, 28, 41]. Although solution uniqueness is not necessary, as discussed above, it is an important property with immediate implications for recovering sparse signals in compressed sensing; see, e.g., [13, 23, 24, 27, 36, 43, 44, 48, 50, 51] and the references therein. As a direct consequence of our analysis on the ℓ_1 -regularized problem, we fully characterize solution uniqueness to Lasso. To the best of our knowledge, Fuchs [23] initialized this direction by introducing a simple sufficient condition for this property, which has been extended in other cited papers. Then, in [43], Tibshirani showed that a sufficient condition closely related to Fuchs' condition is also necessary almost everywhere. The full characterization for this property has been obtained recently in [50, 51] by using results of strong duality in linear programming. This characterization, which is based on an existence of a vector satisfying a system of linear equations and inequalities, allows [50, 51] to recover the aforementioned sufficient conditions and provide some situations in which these conditions turn necessary. Some related results have been developed in [27, 36]. Our approach to solution uniqueness is new and different. We also derive several new full characterizations in terms of *positively linear independence* and *Slater-type conditions*, which can be easily verifiable.

The outline of our paper is as follows. Section 2 briefly presents some second-order characterization for quadratic growth condition in terms of *subgradient graphical derivative* [39] and recalls some convergence analysis from our part I [8]. Section 3 devotes to the study of the quadratic growth condition in some structured optimization problems involving Poisson inverse regularized, ℓ_1 -regularized, and ℓ_1 -regularized least square optimization problems. In Sect. 4, we obtain several new full characterizations to the uniqueness of optimal solution to Lasso problem. Section 5 gives some conclusions and potential future works in this direction.

2 Preliminary Results on Metric Subregularity of the Subdifferential and Quadratic Growth Condition

Throughout the paper, \mathbb{R}^n is the usual Euclidean space with dimension n where $\|\cdot\|$ and $\langle \cdot, \cdot \rangle$ denote the corresponding Euclidean norm and inner product in \mathbb{R}^n . We use $\Gamma_0(\mathbb{R}^n)$ to denote the set of proper, lower semicontinuous, and convex functions on

\mathbb{R}^n . Let $h \in \Gamma_0(\mathbb{R}^n)$, we write $\text{dom } h := \{x \in \mathbb{R}^n \mid h(x) < +\infty\}$. The subdifferential of h at $\bar{x} \in \text{dom } h$ is defined by

$$\partial h(\bar{x}) := \{v \in \mathbb{R}^n \mid \langle v, x - \bar{x} \rangle \leq h(x) - h(\bar{x}), \text{ for all } x \in \mathbb{R}^n\}. \quad (1)$$

We say h satisfies the *quadratic growth condition* at \bar{x} with modulus $\kappa > 0$ if there exist $\varepsilon > 0$ such that

$$h(x) \geq h(\bar{x}) + \frac{\kappa}{2} d^2(x; (\partial h)^{-1}(0)) \quad \text{for all } x \in \mathbb{B}_\varepsilon(\bar{x}). \quad (2)$$

Here, for a set S , $d(x; S)$ denotes the distance from x to S , and $\mathbb{B}_\varepsilon(\bar{x})$ denotes the ball centered at \bar{x} with radius ε . Moreover, if (2) and $(\partial h)^{-1}(0) = \{\bar{x}\}$ are both satisfied, then we say *strong quadratic growth condition* holds for h at \bar{x} with modulus κ .

Some relationship between the quadratic growth condition and the so-called metric subregularity of the subdifferential could be found in [1–3, 10, 22] even for the case of nonconvex functions. The quadratic growth condition (2) is also called *quadratic functional growth* property in [38] when h is continuously differentiable over a closed convex set. In [25, 26], h is said to be 2-conditioned on $\mathbb{B}_\varepsilon(\bar{x})$ if it satisfies the quadratic growth condition (2).

The following proposition, a slight improvement in [2, Corollary 3.7], provides a useful characterization for strong quadratic growth condition via the subgradient graphical derivative [39, Chapter 13].

Proposition 2.1 (Characterization of strong quadratic growth condition) *Let $h \in \Gamma_0(\mathbb{R}^n)$ and \bar{x} be an optimal solution, i.e., $0 \in \partial h(\bar{x})$. The following are equivalent:*

- (i) h satisfies the strong quadratic growth condition at \bar{x} .
- (ii) $D(\partial h)(\bar{x}|0)$ is positive-definite in the sense that

$$\langle v, u \rangle > 0 \quad \text{for all } v \in D(\partial h)(\bar{x}|0)(u), u \in \mathbb{R}^n, u \neq 0, \quad (3)$$

where $D(\partial h)(\bar{x}|0) : \mathbb{R}^n \rightrightarrows \mathbb{R}^n$ is the subgradient graphical derivative of ∂h at \bar{x} for 0 defined by

$$D(\partial h)(\bar{x}|0)(u) := \{v \in \mathbb{R}^n \mid \exists (u_n, v_n) \rightarrow (u, v), t_n \downarrow 0 \\ \text{such that } t_n v_n \in \partial h(\bar{x} + t_n u_n)\} \text{ for any } u \in \mathbb{R}^n.$$

Moreover, if (ii) is satisfied then

$$\ell := \inf \left\{ \frac{\langle v, u \rangle}{\|u\|^2} \mid v \in D(\partial h)(\bar{x}|0)(u), u \in \mathbb{R}^n \right\} > 0 \quad (4)$$

with convention $\frac{0}{0} = \infty$ and h satisfies the strong quadratic growth condition at \bar{x} with any modulus $\kappa < \ell$.

Proof The implication [(i)⇒(ii)] follows from [2, Theorem 3.6 and Corollary 3.7]. If (ii) is satisfied, we obtain from (3) that $\|v\| \geq \ell\|u\|$. Combining [20, Theorem 4C.1] and [22, Corollary 3.3] tells us that h satisfies the strong quadratic growth condition at \bar{x} with any modulus $\kappa < \ell$. The proof is complete. \square

Next, let us recall here some main results from our part I [8] regarding the convergence of forward–backward splitting method (FBS) for solving the following optimization problem:

$$\min_{x \in \mathbb{R}^n} F(x) := f(x) + g(x), \tag{5}$$

where $f, g : \mathbb{R}^n \rightarrow \mathbb{R} \cup \{\infty\}$ are proper, lower semi-continuous, and convex functions.¹ The standing assumptions on the initial data for (5) used throughout the paper:

- A1** $f, g \in \Gamma_0(\mathbb{R}^n)$ and $\text{int}(\text{dom } f) \cap \text{dom } g \neq \emptyset$.
- A2** f is continuously differentiable at any point in $\text{int}(\text{dom } f) \cap \text{dom } g$
- A3** For any $x \in \text{int}(\text{dom } f) \cap \text{dom } g$, the sublevel set $\{F \leq F(x)\}$ is contained in $\text{int}(\text{dom } f) \cap \text{dom } g$.

The forward–backward splitting methods for solving (5) is described by

$$x^{k+1} = \text{prox}_{\alpha_k g}(x^k - \alpha_k \nabla f(x^k)) := (\text{Id} + \alpha_k \partial g)^{-1}(x^k - \alpha_k \nabla f(x^k)) \tag{6}$$

with the proximal operator $\text{prox}_g : \mathbb{R}^n \rightarrow \text{dom } g$ given by

$$\text{prox}_g(z) := (\text{Id} + \partial g)^{-1}(z) \text{ for all } z \in \mathbb{R}^n, \tag{7}$$

and the stepsize $\alpha_k > 0$ determined from the Beck–Teboulle’s line search as follows:

Linesearch BT (Beck–Teboulle’s line search)
 Given $\sigma > 0$ and $\theta \in (0, 1)$. With $\alpha_{-1} := \sigma$ and $x^0 \in \text{int}(\text{dom } f) \cap \text{dom } g$:
Input. Set $\alpha_k = \alpha_{k-1}$ and $J(x^k, \alpha_k) = \text{prox}_{\alpha_k g}(x^k - \alpha_k \nabla f(x^k))$.

While $f(J(x^k, \alpha_k)) > f(x^k) + \langle \nabla f(x^k), J(x^k, \alpha_k) - x^k \rangle + \frac{1}{2\alpha_k} \|x^k - J(x^k, \alpha_k)\|^2$,
do $\alpha_k = \theta \alpha_k$.
End While

Output. α_k

In [8, Proposition 3.1 and Corollary 3.1], we show that the linesearch above terminates after finite steps, the FBS sequence $(x^k)_{k \in \mathbb{N}} \subset \text{int}(\text{dom } f) \cap \text{dom } g$ is well defined, and thus f is differentiable at x^k by assumption **A2**. The global convergence [8, Theorem 3.1] is recalled here.

¹ In [8], we examined FBS in the more general (possibly infinite dimensional) Hilbert space setting. On the other hand, for the purpose of discussing the structured optimization problems later on, we restrict ourselves to the finite-dimensional setting here.

Theorem 2.1 (Global convergence of FBS method) *Let $(x^k)_{k \in \mathbb{N}}$ be the sequence generated from FBS method. Suppose that the solution set is not empty. Then, $(x^k)_{k \in \mathbb{N}}$ converges to an optimal solution point. Moreover, $(F(x^k))_{k \in \mathbb{N}}$ also converges to the optimal value.*

When the cost function F satisfies the quadratic growth condition and ∇f is locally Lipschitz continuous, our [8, Theorem 4.1] shows that both iterative and cost sequences of FBS are Q -linearly convergent.

Theorem 2.2 (Q -linear convergence under quadratic growth condition) *Let $(x^k)_{k \in \mathbb{N}}$ be the sequence generated from FBS method. Suppose that the optimal solution set S^* to problem (5) is nonempty, and let $x^* \in S^*$ be the limit point of $(x^k)_{k \in \mathbb{N}}$. Suppose further that ∇f is locally Lipschitz continuous around x^* with constant $L > 0$. If F satisfies the quadratic growth condition at x^* with modulus $\kappa > 0$, there exists $K \in \mathbb{N}$ such that*

$$\|x^{k+1} - x^*\| \leq \frac{1}{\sqrt{1 + \frac{\alpha\kappa}{4}}} \|x^k - x^*\| \quad (8)$$

$$|F(x^{k+1}) - F(x^*)| \leq \frac{\sqrt{1 + \alpha\kappa} + 1}{2\sqrt{1 + \alpha\kappa}} |F(x^k) - F(x^*)| \quad (9)$$

for any $k > K$, where $\alpha := \min \left\{ \alpha_K, \frac{\theta}{L} \right\}$.

If, in addition, ∇f is globally Lipschitz continuous on $\text{int}(\text{dom } f) \cap \text{dom } g$ with constant $L > 0$, α could be chosen as $\min \left\{ \sigma, \frac{\theta}{L} \right\}$.

Under the strong quadratic growth condition, a sharper rate is obtained in [8, Corollary 4.1].

Corollary 2.1 (Sharper Q -linear convergence rate under strong quadratic growth condition) *Let $(x^k)_{k \in \mathbb{N}}$ be the sequence generated from FBS method. Suppose that the solution set S^* is not empty, and let $x^* \in S^*$ be the limit point of $(x^k)_{k \in \mathbb{N}}$ as in Theorem 2.1. Suppose further that ∇f is locally Lipschitz continuous around x^* with constant $L > 0$. If F satisfies the strong quadratic growth condition at x^* with modulus $\kappa > 0$, then there exists some $K \in \mathbb{N}$ such that for any $k > K$ we have*

$$\|x^{k+1} - x^*\| \leq \frac{1}{\sqrt{1 + \alpha\kappa}} \|x^k - x^*\| \quad \text{with } \alpha := \min \left\{ \alpha_K, \frac{\theta}{L} \right\}.$$

Additionally, if ∇f is globally Lipschitz continuous on $\text{int}(\text{dom } f) \cap \text{dom } g$ with constant $L > 0$, α above could be chosen as $\min \left\{ \sigma, \frac{\theta}{L} \right\}$.

3 Quadratic Growth Conditions and Linear Convergence of Forward-Backward Splitting Method in Some Structured Optimization Problems

In this section, we mainly show that quadratic growth condition is automatic or can be fulfilled under mild assumptions in several important classes of convex optimization.

3.1 Poisson Linear Inverse Problem

This subsection devotes to the study of the *eventually* linear convergence of FBS when solving the following standard Poisson regularized problem [16, 47]

$$\min_{x \in \mathbb{R}_+^n} \sum_{i=1}^m \{b_i \log \frac{b_i}{(Ax)_i} + (Ax)_i - b_i\}, \tag{10}$$

where $A \in \mathbb{R}_+^{m \times n}$ is an $m \times n$ matrix with nonnegative entries and nontrivial rows, and $b \in \mathbb{R}_+^m$ is a positive vector. This problem is usually used to recover a signal $x \in \mathbb{R}_+^n$ from the measurement b corrupted by Poisson noise satisfying $Ax \simeq b$. The problem (10) could be written in terms of (5) in which

$$f(x) := h(Ax), \quad g(x) = \delta_{\mathbb{R}_+^n}(x), \quad \text{and} \quad F_1(x) := h(Ax) + g(x), \tag{11}$$

where h is the Kullback–Leibler divergence defined by

$$h(y) = \begin{cases} \sum_{i=1}^m \{b_i \log \frac{b_i}{y_i} + y_i - b_i\} & \text{if } y \in \mathbb{R}_{++}^m, \\ +\infty & \text{if } y \in \mathbb{R}_+^m \setminus \mathbb{R}_{++}^m. \end{cases} \tag{12}$$

Note from (11) and (12) that $\text{dom } f = A^{-1}(\mathbb{R}_{++}^m)$, which is an open set. Moreover, since $A \in \mathbb{R}_+^{m \times n}$, we have $\text{dom } f \cap \text{dom } g = A^{-1}(\mathbb{R}_{++}^m) \cap \mathbb{R}_+^n \neq \emptyset$ and f is continuously differentiable at any point on $\text{dom } f \cap \text{dom } g$. The standing assumptions **A1** and **A2** are satisfied for Problem (10). Moreover, since the function F_1 is bounded below and coercive, the optimal solution set of problem (10) is always nonempty.

It is worth noting further that ∇f is locally Lipschitz continuous at any point $\text{int}(\text{dom } f) \cap \text{dom } g$ but not globally Lipschitz continuous on $\text{int}(\text{dom } f) \cap \text{dom } g$. [40, Section 4] and our [8, Theorem 3.2] show that FBS is applicable to solving (10) with global convergence rate $o(\frac{1}{k})$. In the recent work [4], a new algorithm on a variant of FBS was designed with applications to solving (10). However, the theory developed in [4] could not guarantee the global convergence of the sequence $(x^k)_{k \in \mathbb{N}}$ generated by the algorithm in solving (10). This is because their assumptions on the closedness of the domain of their auxiliary Legendre function in [4, Theorem 2] are not satisfied for (10). Our intent here is to reveal the Q-linear convergence of our method when solving (10) in the sense of Theorem 2.2. In order to do so, we need to verify the quadratic growth condition for F_1 at any optimal minimizer for 0. Note further that the Kullback–Leibler divergence h is not strongly convex and ∇f is not globally Lipschitz continuous; hence, standing assumptions in [21] are not satisfied. Proving the quadratic growth condition for F_1 at an optimal solution via the approach of [21] needs to be proceeded with caution.

Lemma 3.1 *Let \bar{x} be an optimal solution to problem (10). Then, for any $R > 0$, we have*

$$F_1(x) - F_1(\bar{x}) \geq \frac{\nu}{2} d^2(x; S^*) \text{ for all } x \in \mathbb{B}_R(\bar{x}) \tag{13}$$

with some constant $\nu > 0$.

Proof Pick any $R > 0$ and $x \in \mathbb{B}_R(\bar{x})$. We only need to prove (13) for the case that $x \in \text{dom } F_1 \cap \mathbb{B}_R(\bar{x})$, i.e., $x \in A^{-1}(\mathbb{R}_{++}^n) \cap \mathbb{R}_+^n \cap \mathbb{B}_R(\bar{x})$. Note that

$$\nabla f(x) = \sum_{i=1}^m \left[1 - \frac{b_i}{\langle a_i, x \rangle} \right] a_i \text{ and } \langle \nabla^2 f(x) d, d \rangle = \sum_{i=1}^m b_i \frac{\langle a_i, d \rangle^2}{\langle a_i, x \rangle^2} \text{ for all } d \in \mathbb{R}^n,$$

where a_i is the i th row of A . Define $\bar{y} := A\bar{x}$, for any $x, u \in \mathbb{B}_R(\bar{x}) \cap \text{dom } f$ we have $[x, u] \subset \mathbb{B}_R(\bar{x}) \cap \text{dom } f$ and obtain from the mean-value theorem that

$$\begin{aligned} f(x) - f(u) - \langle \nabla f(u), x - u \rangle &= \frac{1}{2} \int_0^1 \langle \nabla^2 f(u + t(x - u)) x - u, x - u \rangle dt \\ &= \frac{1}{2} \int_0^1 \sum_{i=1}^m b_i \frac{\langle a_i, x - u \rangle^2}{\langle a_i, u + t(x - u) \rangle^2} dt \\ &\geq \frac{1}{2} \int_0^1 \sum_{i=1}^m b_i \frac{\langle a_i, x - u \rangle^2}{[\langle a_i, \bar{x} \rangle + \|a_i\|(\|u - \bar{x}\| + t\|x - u\|)]^2} dt \\ &\geq \frac{1}{2} \sum_{i=1}^m \frac{b_i}{[\langle a_i, \bar{x} \rangle + 3\|a_i\|R]^2} \langle a_i, x - u \rangle^2. \end{aligned}$$

Similarly, we have

$$f(u) - f(x) - \langle \nabla f(x), u - x \rangle \geq \frac{1}{2} \sum_{i=1}^m \frac{b_i}{[\langle a_i, \bar{x} \rangle + 3\|a_i\|R]^2} \langle a_i, u - x \rangle^2$$

for $x, u \in \mathbb{B}_R(\bar{x}) \cap \text{dom } f$. (14)

Adding the above two inequalities gives us that

$$\langle \nabla f(x) - \nabla f(u), x - u \rangle \geq \sum_{i=1}^m \frac{b_i}{[\langle a_i, \bar{x} \rangle + 3\|a_i\|R]^2} \langle a_i, x - u \rangle^2$$

for all $x, u \in \mathbb{B}_R(\bar{x}) \cap \text{dom } f$. (15)

We claim that the optimal solution set S^* to problem (10) satisfies that

$$S^* = A^{-1}(\bar{y}) \cap (\partial g)^{-1}(-\nabla f(\bar{x})) \text{ with } \bar{y} = A\bar{x}. \tag{16}$$

Pick another optimal solution $\bar{u} \in S^*$, we have $\bar{u}_t := \bar{x} + t(\bar{x} - \bar{u}) \in S^* \subset \text{dom } f$ for any $t \in [0, 1]$ due to the convexity of S^* . By choosing t sufficiently small, we have

$\bar{u}_t \in \mathbb{B}_R(\bar{x}) \cap \text{dom } f$. Note further that $-\nabla f(\bar{u}_t) \in \partial g(\bar{u}_t)$ and $-\nabla f(\bar{x}) \in \partial g(\bar{x})$. Since ∂g is a monotone operator, we obtain that

$$0 \geq \langle \nabla f(\bar{x}) - \nabla f(\bar{u}_t), \bar{x} - \bar{u}_t \rangle.$$

This together with (15) tells us that $\langle a_i, \bar{x} - \bar{u}_t \rangle = 0$ for all $i = 1, \dots, m$. Hence, $A\bar{x} = A\bar{u} = \bar{y}$ for any $\bar{u} \in S^*$, which also implies that

$$\nabla f(\bar{u}) = A^T \nabla h(A\bar{u}) = A^T \nabla h(A\bar{x}) = \nabla f(\bar{x}). \tag{17}$$

This verifies the inclusion “ \subset ” in (16). The opposite inclusion is trivial. Indeed, take any u satisfying that $Au = \bar{y}$ and $-\nabla f(\bar{x}) \in \partial g(u)$, similarly to (17) we have $-\nabla f(u) = -\nabla f(\bar{x}) \in \partial g(u)$. This shows that $0 \in \nabla f(u) + \partial g(u)$, i.e., $u \in S^*$. The proof for equality (16) is completed.

Note from (16) that the optimal solution set S^* is a polyhedral with the following format

$$S^* = \{u \in \mathbb{R}^n \mid Au = \bar{y} = A\bar{x}, \langle \nabla f(\bar{x}), u \rangle = 0, u \in \mathbb{R}_+^n\}$$

due to the fact that $(\partial g)^{-1}(-\nabla f(\bar{x})) = \{u \in \mathbb{R}_+^n \mid \langle \nabla f(\bar{x}), u \rangle = 0 = \langle \nabla f(\bar{x}), \bar{x} \rangle\}$. Thanks to Hoffman’s lemma, there exists a constant $\gamma > 0$ such that

$$d(x; S^*) \leq \gamma(\|Ax - A\bar{x}\| + |\langle \nabla f(\bar{x}), x - \bar{x} \rangle|) \quad \text{for all } x \in \mathbb{R}_+^n. \tag{18}$$

Fix any $x \in \mathbb{B}_R(\bar{x}) \cap \mathbb{R}_+^n$, (14) tells us that

$$f(x) - f(\bar{x}) - \langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq \frac{1}{2}\alpha\|Ax - A\bar{x}\|^2, \tag{19}$$

Where $\alpha = \min_{1 \leq i \leq m} \left[\frac{b_i}{\|(\langle a_i, \bar{x} \rangle + 3\|a_i\|R)^2} \right]$. Since $-\nabla f(\bar{x}) \in \partial g(\bar{x})$, we have $\langle \nabla f(\bar{x}), x - \bar{x} \rangle \geq 0$. This together with (19) implies that

$$\begin{aligned} F_1(x) - F_1(\bar{x}) &\geq \frac{1}{2}\alpha\|Ax - A\bar{x}\|^2 + \langle \nabla f(\bar{x}), x - \bar{x} \rangle \\ &\geq \frac{1}{2}\alpha\|Ax - A\bar{x}\|^2 \\ &\quad + \frac{1}{(\|\nabla f(\bar{x})\| + 1)\|x - \bar{x}\|} \langle \nabla f(\bar{x}), x - \bar{x} \rangle^2 \\ &\geq \min \left\{ \frac{1}{2}\alpha \frac{1}{(\|\nabla f(\bar{x})\| + 1)R} \right\} \\ &\quad [\|Ax - A\bar{x}\|^2 + \langle \nabla f(\bar{x}), x - \bar{x} \rangle^2] \\ &\geq \frac{1}{2} \min \left\{ \frac{1}{2}\alpha, \frac{1}{(\|\nabla f(\bar{x})\| + 1)R} \right\} \\ &\quad [\|Ax - A\bar{x}\| + |\langle \nabla f(\bar{x}), x - \bar{x} \rangle|]^2 \end{aligned}$$

$$\geq \frac{1}{2\gamma^2}\alpha, \frac{1}{(\|\nabla f(\bar{x})\| + 1)R} \} d^2(x; S^*),$$

where the fourth inequality follows from the elementary inequality that $\frac{(a+b)^2}{2} \leq a^2 + b^2$ with $a, b \geq 0$, and the last inequality is from (18). This clearly ensures (13). □

When applying FBS to solving problem (10), we have

$$x^{k+1} = \mathbb{P}_{\mathbb{R}_+^n} \left(x^k - \alpha_k \sum_{i=1}^m \left[1 - \frac{b_i}{\langle a_i, x^k \rangle} \right] a_i \right) \quad \text{with } x^0 \in A^{-1}(\mathbb{R}_{++}^n) \cap \mathbb{R}_+^n, \tag{20}$$

where $\mathbb{P}_{\mathbb{R}_+^n}(\cdot)$ is the projection mapping to \mathbb{R}_+^n .

Corollary 3.1 (Q-linear convergence of method (20)) *Let $(x^k)_{k \in \mathbb{N}}$ be the sequence generated from (20) with $x^0 \in A^{-1}(\mathbb{R}_+^n) \cap \mathbb{R}_+^n$ for solving the Poisson regularized problem (10). Then, the sequences $(x^k)_{k \in \mathbb{N}}$ and $(F_1(x^k))_{k \in \mathbb{N}}$ are Q-linearly convergent to an optimal solution and the optimal value to (10), respectively.*

Proof Since both functions f and g in problem (10) satisfy our standing assumptions **A1** and **A2**, and problem (10) always has optimal solutions, the sequence $(x^k)_{k \in \mathbb{N}}$ converges to an optimal solution \bar{x} to problem (10) by Theorem 2.1. Since ∇f is locally Lipschitz continuous around \bar{x} , the combination of Theorem 2.2 and Lemma 3.1 tells us that $(x^k)_{k \in \mathbb{N}}$ is Q-linearly convergent to \bar{x} . □

Using a similar line of argument as above, one can show that quadratic growth condition in Lemma 3.1 is also valid for the following Poisson inverse problem with sparse regularization [4]:

$$\min_{x \in \mathbb{R}_+^n} \sum_{i=1}^m \left\{ b_i \log \frac{b_i}{(Ax)_i} + (Ax)_i - b_i \right\} + \mu \|x\|_1, \tag{21}$$

where $\mu > 0$ is the penalty parameter. Indeed, noting that $\|x\|_1 = \langle e, x \rangle$ for $x \in \mathbb{R}_+^n$, with $e = (1, 1, \dots, 1) \in \mathbb{R}^n$. The objective function of (21) can be written as $p(x) + g(x)$ where $p(x) := f(x) + \mu \langle e, x \rangle$, and f, g are given as in (11). Then, the FBS method for solving (21) can proceed by replacing the function $f(x)$ in (11) by $p(x)$. Let $\hat{x} \in \text{dom } p = \text{dom } f$ be a minimizer to (21). Observe that the function p also satisfies the similar inequality as in (14)

$$\begin{aligned}
 p(u) - p(x) - \langle \nabla p(x), u - x \rangle &= f(u) - f(x) - \langle \nabla f(x), u - x \rangle \\
 &\geq \frac{1}{2} \sum_{i=1}^m \frac{b_i}{[|\langle a_i, \bar{x} \rangle| + 3\|a_i\|R]^2} \langle a_i, u - x \rangle^2 \\
 &\quad \text{for } x, u \in \mathbb{B}_R(\hat{x}) \cap \text{dom } p.
 \end{aligned}$$

As (14) plays the central role in the proof of Lemma 3.1, we can repeat all the steps in this proof by replacing the function f there by p and \bar{x} by \hat{x} to prove the quadratic growth condition of problem (21). This together with Corollary 3.1 shows that FBS (10) solves (21) linearly.

3.2 ℓ_1 -Regularized Optimization Problems

In this section, we consider the ℓ_1 -regularized optimization problems

$$\min_{x \in \mathbb{R}^n} F_2(x) := f(x) + \mu \|x\|_1, \tag{22}$$

where $\|x\|_1$ denotes the ℓ_1 -norm of x .

In order to use Proposition 2.1 for characterizing the strong quadratic growth condition for F_2 , we need the following calculation of subgradient graphical derivative of $\partial(\mu\|\cdot\|_1)$.

Proposition 3.1 (Subgradient graphical derivative $\partial(\mu\|\cdot\|_1)$) *Suppose that $\bar{s} \in \partial(\mu\|\cdot\|_1)(x^*)$. Define $I := \{j \in \{1, \dots, n\} \mid |\bar{s}_j| = \mu\}$, $J := \{j \in I \mid x_j^* \neq 0\}$, $K := \{j \in I \mid x_j^* = 0\}$, and $H(x^*) := \{u \in \mathbb{R}^n \mid u_j = 0, j \notin I \text{ and } u_j \bar{s}_j \geq 0, j \in K\}$. Then, $D\partial\mu\|\cdot\|_1(x^*|\bar{s})(u)$ is nonempty if and only if $u \in H(x^*)$. Furthermore, we have*

$$\begin{aligned}
 &D\partial(\mu\|\cdot\|_1)(x^*|\bar{s})(u) \\
 &= \left\{ v \in \mathbb{R}^n \mid \begin{array}{l} v_j = 0, j \in J \\ u_j v_j = 0, \bar{s}_j v_j \leq 0, j \in K \end{array} \right\} \text{ for all } u \in H(x^*). \tag{23}
 \end{aligned}$$

Proof For any $x \in \mathbb{R}^n$, note that

$$\partial\mu\|x\|_1 = \left\{ s \in \mathbb{R}^n \mid \begin{array}{l} s_j = \mu \operatorname{sgn}(x_j) \text{ if } x_j \neq 0 \\ s_j \in [-\mu, \mu] \text{ if } x_j = 0 \end{array} \right\}, \tag{24}$$

where $\operatorname{sgn} : \mathbb{R} \rightarrow \{-1, 1\}$ is the *sign function*. Take any $v \in D\partial\|\cdot\|_1(x^*|\bar{s})(u)$, there exists sequence $t^k \downarrow 0$ and $(u^k, v^k) \rightarrow (u, v)$ such that $(x^*, \bar{s}) + t^k(u^k, v^k) \in \operatorname{gph} \partial\mu\|\cdot\|_1$. Let us consider three partitions of j described below:

Partition 1.1 $j \notin I$, i.e., $|\bar{s}_j| < \mu$. It follows from (24) that $x_j^* = 0$. For sufficiently large k , we have $|(\bar{s} + t^k v^k)_j| < \mu$ and thus $|(x^* + t^k u^k)_j| = 0$ by (24) again. Hence, $u_j^k = 0$, which implies that $u_j = 0$ for all $j \notin I$.

Partition 1.2 $j \in J$, i.e., $|\bar{s}_j| = \mu$ and $x_j^* \neq 0$. When k is sufficiently large, we have $(x^* + t^k u^k)_j \neq 0$ and derive from (24) that

$$(\bar{s} + t^k v^k)_j = \mu \operatorname{sgn}(x^* + t^k u^k)_j = \mu \operatorname{sgn} x_j^* = \bar{s}_j,$$

which implies that $v_j = 0$ for all $j \in J$.

Partition 1.3 $j \in K$, i.e., $|\bar{s}_j| = \mu$ and $x_j^* = 0$. If there is a subsequence of $(x^*, \bar{s})_j + t^k(u^k, v^k)_j$ (without relabeling) such that $|(\bar{s} + t^k v^k)_j| < \mu = |\bar{s}_j|$, we have $\bar{s}_j v_j^k < 0$ and $(x^* + t^k u^k)_j = 0$ by (24). It follows that $u_j^k = 0$. Letting $k \rightarrow \infty$, we have $u_j = 0$ and $\bar{s}_j v_j \leq 0$. Otherwise, we find some $L > 0$ such that $|(\bar{s} + t^k v^k)_j| = \mu = |\bar{s}_j|$ for all $k > L$, which yields $v_j^k = 0$. Taking $k \rightarrow \infty$ gives us that $v_j = 0$. Furthermore, by (24) again, we have

$$\begin{aligned} \bar{s}_j &= (\bar{s} + t^k v^k)_j = \mu \operatorname{sgn}(x^* + t^k u^k)_j = \mu \operatorname{sgn}(u_j^k) \quad \text{or} \\ 0 &= (x^* + t^k u^k)_j = t^k u_j^k, \quad \text{i.e., } u_j^k = 0, \end{aligned}$$

which imply that $\bar{s}_j u_j \geq 0$ after passing the limit $k \rightarrow \infty$.

Combining the conclusions in three cases above gives us that $u \in H(x^*)$ and also verifies the inclusion “ \subset ” in (23). To justify the converse inclusion “ \supset ”, take $u \in H(x^*)$ and any $v \in \mathbb{R}^n$ with $v_j = 0$ for $j \in J$ and $u_j v_j = 0, \bar{s}_j v_j \leq 0$ for $j \in K$. For any $t^k \downarrow 0$, we prove that $(x^*, \bar{s}) + t^k(u, v) \in \operatorname{gph} \partial \mu \|\cdot\|_1$ and thus verify that $v \in D\partial \mu \|\cdot\|_1(x^*|\bar{s})(u)$. For any $t \in \mathbb{R}$, define the set-valued mapping:

$$\operatorname{SGN}(t) := \partial|t| = \begin{cases} \operatorname{sgn}(t) & \text{if } t \neq 0 \\ [-1, 1] & \text{if } t = 0. \end{cases}$$

Similarly to the proof of “ \subset ” inclusion, we consider three partitions of j as follows:

Partition 2.1 $j \notin I$, i.e., $|\bar{s}_j| < \mu$. Since $u \in H(x^*)$, we have $u_j = 0$. Note also that $x_j^* = 0$. Hence, we get $(x^* + t^k u)_j = 0$ and $(\bar{s} + t^k v)_j \in [-\mu, \mu]$ when k is sufficiently large, which means $(\bar{s} + t^k v)_j \in \mu \operatorname{SGN}(x^* + t^k u)_j$.

Partition 2.2 $j \in J$, i.e., $|\bar{s}_j| = \mu$ and $x_j^* \neq 0$. Since $v_j = 0$, we have

$$\operatorname{sgn}(\bar{s} + t^k v)_j = \operatorname{sgn} \bar{s}_j = \operatorname{sgn}(x_j^*) = \operatorname{sgn}(x^* + t^k u)_j$$

and $(x^* + t^k u)_j \neq 0$ when k is large. It follows that $(\bar{s} + t^k v)_j \in \mu \operatorname{SGN}(x^* + t^k u)_j$.

Partition 2.3 $j \in K$, i.e., $|\bar{s}_j| = \mu$ and $x_j^* = 0$. If $u_j = 0$, we have $(x^* + t^k u)_j = 0$ and $|(\bar{s} + t^k v)_j| \leq |\bar{s}_j| \leq \mu$ for sufficiently large k , since $\bar{s}_j v_j \leq 0$. If $u_j \neq 0$, we have $v_j = 0$ and

$$(\bar{s} + t^k v)_j = \bar{s}_j = \operatorname{sgn}(u_j) = \operatorname{sgn}(x^* + t^k u)_j$$

when k is large, since $u_j \bar{s}_j \geq 0$. In both cases, we have $(\bar{s} + t^k v)_j \in \mu \operatorname{SGN}(x^* + t^k u)_j$.

From those cases, we always have $(x^*, \bar{s}) + t^k(u, v) \in \text{gph } \partial\mu \| \cdot \|_1$ and thus $v \in D\partial\mu \| \cdot \|_1(x^*|\bar{s})(u)$. □

As a consequence, we establish a characterization of strong quadratic growth condition for F_2 .

Theorem 3.1 (Characterization of strong quadratic growth condition for F_2) *Let x^* be an optimal solution to problem (22). Suppose that ∇f is differentiable at x^* . Define $\mathcal{E} := \{j \in \{1, \dots, n\} \mid |(\nabla f(x^*))_j| = \mu\}$, $K := \{j \in \mathcal{E} \mid x_j^* = 0\}$, $\mathcal{U} := \{u \in \mathbb{R}^{\mathcal{E}} \mid u_j(\nabla f(x^*))_j \leq 0, j \in K\}$ with $\mathbb{R}^{\mathcal{E}} = \{u = (u_j)_{j \in \mathcal{E}} \in \mathbb{R}^{|\mathcal{E}|}\}$, and $\mathcal{H}_{\mathcal{E}}(x^*) := [\nabla^2 f(x^*)_{i,j}]_{i,j \in \mathcal{E}}$. Then, the following statements are equivalent:*

- (i) F_2 satisfies the strong quadratic growth condition at x^* .
- (ii) $\mathcal{H}_{\mathcal{E}}(x^*)$ is positive definite over \mathcal{U} in the sense that

$$\langle \mathcal{H}_{\mathcal{E}}(x^*)u, u \rangle > 0 \quad \text{for all } u \in \mathcal{U} \setminus \{0\}. \tag{25}$$

- (iii) $\mathcal{H}_{\mathcal{E}}(x^*)$ is nonsingular over \mathcal{U} in the sense that

$$\ker \mathcal{H}_{\mathcal{E}}(x^*) \cap \mathcal{U} = \{0\}. \tag{26}$$

Moreover, if (25) is satisfied, then F_2 satisfies the strong quadratic growth condition with any positive modulus $\kappa < \ell$ with

$$\ell := \min \left\{ \frac{\langle \mathcal{H}_{\mathcal{E}}(x^*)u, u \rangle}{\|u\|^2} \mid u \in \mathcal{U} \right\} > 0 \tag{27}$$

with the convention $\frac{0}{0} = \infty$.

Proof First let us verify the equivalence between (i) and (ii) by using Proposition 2.1. Indeed, for any $v \in D(\partial F_2)(x^*|0)(u)$ we have get from the sum rule [20, Proposition 4A.2] that

$$v - \nabla^2 f(x^*)u \in D\partial\mu \| \cdot \|_1(x^*| - \nabla f(x^*))(u).$$

Define $\mathcal{V} := \{u \in \mathbb{R}^n \mid u_j = 0, j \notin \mathcal{E}, u_j(\nabla f(x^*))_j \leq 0, j \in K\}$. Thanks to Proposition 3.1, we have

$$\langle v - \nabla^2 f(x^*)u, u \rangle = 0 \quad \text{for all } u \in \mathcal{V}. \tag{28}$$

This tells us that (25) is the same with (3) when $h = F_2$. By Proposition 2.1, (i) and (ii) are equivalent. Moreover, F_2 satisfies the strong quadratic growth condition with any positive modulus $\kappa < \ell$.

Finally, the equivalence between (ii) and (iii) is trivial due to the fact that f is convex and thus $\mathcal{H}_{\mathcal{E}}(x^*)$ is positive semi-definite. □

Corollary 3.2 (Linear convergence of FBS method for ℓ_1 -regularized problems) *Let $(x^k)_{k \in \mathbb{N}}$ be the sequences generated from FBS method for problem (22). Suppose that the solution set S^* is not empty, $(x^k)_{k \in \mathbb{N}}$ is converging to some $x^* \in S^*$, and that f is C^2 around x^* . If condition (25) holds, then $(x^k)_{k \in \mathbb{N}}$ and $(F_2(x^k))_{k \in \mathbb{N}}$ are Q-linearly convergent to x^* and $F_2(x^*)$, respectively, with rates determined in Corollary 2.1, where κ is any positive number smaller than ℓ in (27).*

Proof Since f is C^2 around x^* , ∇f is locally Lipschitz continuous around x^* . The result follows from Corollary 2.1 and Theorem 3.1. □

Remark 3.1 It is worth noting that condition (26) is strictly weaker than the assumption used in [28] that $H_{\mathcal{E}}(x^*)$ has full rank to obtain the linear convergence of FBS for (22). Indeed, let us take into account the case $n = 2$, $\mu = 1$, and $f(x_1, x_2) = \frac{1}{2}(x_1 + x_2)^2 + x_1 + x_2$. Note that $x^* = (0, 0)$ is an optimal solution to problem (22). Moreover, direct computation gives us that $\nabla f(x^*) = (1, 1)$, $\mathcal{E} = \{1, 2\}$, and $H_{\mathcal{E}}(x^*) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$. It is clear that $H_{\mathcal{E}}(x^*)$ does not have full rank, but condition (25) and its equivalence (26) hold.

3.3 Global Q-linear Convergence of ISTA on Lasso Problem

In this section, we study the linear convergence of ISTA for Lasso problem

$$\min_{x \in \mathbb{R}^n} F_3(x) := \frac{1}{2} \|Ax - b\|^2 + \mu \|x\|_1, \tag{29}$$

where A is an $m \times n$ real matrix and b is a vector in \mathbb{R}^m .

The following lemma taken from [10, Lemma 10] plays an important role in our proof.

Lemma 3.2 (Global error bound) *Fix any $R > \frac{\|b\|^2}{2\mu}$. Suppose that x^* is an optimal solution to problem (29). Then, we have*

$$F_3(x) - F_3(x^*) \geq \frac{\gamma_R}{2} d^2(x; S^*) \text{ for all } \|x\|_1 \leq R,$$

where

$$\gamma_R^{-1} := v^2 \left(1 + \frac{\sqrt{5}}{2} \mu R + (R\|A\| + \|b\|)(4R\|A\| + \|b\|) \right) \tag{30}$$

while v is the Hoffman constant defined in [10, Definition 1] and only depends on the initial data A, b, μ .

Global R-linear convergence of $(x^k)_{k \in \mathbb{N}}$ from ISTA and Q-linear convergence of $(F_3(x^k))_{k \in \mathbb{N}}$ for solving Lasso problem were obtained in [25, Theorem 4.2 and Remark 4.3] and also [26, Theorem 4.8]. Here, we add another feature: The iterative sequence $(x^k)_{k \in \mathbb{N}}$ is also globally Q-linearly convergent.

Theorem 3.2 (Global Q-linear convergence of ISTA) *Let $(x^k)_{k \in \mathbb{N}}$ be the sequence generated by ISTA for problem (29) that converges to an optimal solution $x^* \in S^*$. Then, $(x^k)_{k \in \mathbb{N}}$ and $(F_3(x^k))_{k \in \mathbb{N}}$ are globally Q-linearly convergent to x^* and $F_3(x^*)$, respectively:*

$$\|x^{k+1} - x^*\| \leq \frac{1}{\sqrt{1 + \frac{\alpha\gamma_R}{4}}} \|x^k - x^*\| \tag{31}$$

$$|F_3(x^{k+1}) - F_3(x^*)| \leq \frac{2\sqrt{1 + \alpha\gamma_R}}{\sqrt{1 + \alpha\gamma_R} + 1} |F_3(x^k) - F_3(x^*)| \tag{32}$$

for all $k \in \mathbb{N}$, where R is any number bigger than $\|x^0\| + \frac{\|b\|^2}{\mu}$ and γ_R is given as in (30) while $\alpha := \min \left\{ \sigma, \frac{\theta}{\lambda_{\max}(A^T A)} \right\}$.

Proof Note that Lasso always has optimal solutions. With $x^* \in S^*$, we have

$$F_3(0) = \frac{1}{2} \|b\|^2 \geq F_3(x^*) \geq \mu \|x^*\|_1,$$

which implies that $\|x^*\| \leq \|x^*\|_1 \leq \frac{1}{2\mu} \|b\|^2$. It follows from [8, Corollary 3.1] that

$$\begin{aligned} \|x^k\| &\leq \|x^k - x^*\| + \|x^*\| \leq \|x^0 - x^*\| + \|x^*\| \leq \|x^0\| + 2\|x^*\| \\ &\leq \|x^0\| + \frac{\|b\|^2}{\mu} < R \end{aligned}$$

for all $k \in \mathbb{N}$. Thanks to Lemma 3.2, [8, Corollary 3.1], and [8, Proposition 3.2], we have

$$\|x^k - x^*\|^2 - \|x^{k+1} - x^*\|^2 \geq \alpha\gamma_R d^2(x^{k+1}; S^*) \tag{33}$$

with $\alpha = \min \left\{ \sigma, \frac{\theta}{\lambda_{\max}(A^T A)} \right\}$ and the note that $\lambda_{\max}(A^T A)$ is the global Lipschitz constant of the gradient of $\frac{1}{2} \|Ax - b\|^2$. The proof of (31) and (32) is quite similar to the one of (8) and (9) in Theorem 2.2; see [8, Theorem 4.1] for further details. \square

Remark 3.2 (Linear convergence rate comparisons for ISTA) In this remark, we provide some comparisons for the derived linear convergence rate for ISTA with the existing results in the literature.

- For the sequence $(x^k)_{k \in \mathbb{N}}$ generated by ISTA, we first note that our derived Q-linear convergence in Theorem 3.2 is $\frac{1}{\sqrt{1 + \frac{\gamma_R}{4\lambda_{\max}(A^T A)}}$ according to (31). This result is new to the literature. In [10, Theorem 25 and Remark 26], R-linear convergence for this sequence via γ_R was obtained. In the case of constant step size, by setting $\sigma = \frac{1}{\lambda_{\max}(A^T A)}$ and $\theta = 1$, we have $\alpha_k = \alpha = \sigma$; see [8, Remark 4.1].

In this case, the corresponding R-linear convergence rate given in [10] reads as $\frac{1}{\sqrt{1 + \frac{\gamma_R}{3\lambda_{\max}(A^T A)}}}$. On the other hand, using [8, Proposition 4.1(i)] and Lemma 3.2,

one can deduce that the R-linear rate for $(x^k)_{k \in \mathbb{N}}$ is $\frac{1}{\sqrt{1 + \frac{\gamma_R}{\lambda_{\max}(A^T A)}}}$, which is

sharper than the rate $\frac{1}{\sqrt{1 + \frac{\gamma_R}{3\lambda_{\max}(A^T A)}}}$ given in [10].

- For the R-linear convergence rate for $(F_3(x^k))_{k \in \mathbb{N}}$, from [8, Proposition 4.1(ii)] and Lemma 3.2, one can deduce that the rate is $\frac{1}{1 + \frac{\gamma_R}{\lambda_{\max}(A^T A)}}$. This rate

is sharper than the one $\frac{1}{1 + \frac{\gamma_R}{3\lambda_{\max}(A^T A)}}$ derived in [10, Remark 26]. How-

ever, the Q-linear rate $\frac{1}{1 + \frac{\gamma_R}{4\lambda_{\max}(A^T A)}}$ for $(F_3(x^k))_{k \in \mathbb{N}}$ obtained by combining [25, Theorem 4.2(iii)] and Lemma 3.2 is better than our rate given in (32); see also [8, Remark 4.1] for related comparisons. How to improve the Q-linear convergence rate for $(F_3(x^k))_{k \in \mathbb{N}}$ is an interesting future direction of research.

Observe further that the linear rates in Theorem 3.2 depend on the initial point x^0 ; see also [26, Theorem 4.8]. Next, we show that the local linear rates around optimal solutions are uniform and independent of the choice of x^0 .

Corollary 3.3 (Local Q-linear convergence of ISTA with uniform rate) *Let $(x^k)_{k \in \mathbb{N}}$ be the sequence generated by ISTA for problem (29) that converges to an optimal solution $x^* \in S^*$. Then, (31) and (32) are satisfied when k is sufficiently large, where $\alpha = \min \left\{ \sigma, \frac{\theta}{\lambda_{\max}(A^T A)} \right\}$ and R is any number bigger than $\frac{\|b\|^2}{2\mu}$.*

Proof Note from the proof of Theorem 3.2 that $\|x^*\| \leq \frac{\|b\|^2}{2\mu} < R$. By Lemma 3.2, there exists some $\varepsilon \in (0, R - \|x^*\|)$ such that the quadratic growth condition holds at x^* :

$$F_3(x) - F_3(x^*) \geq \frac{\gamma_R}{2} d^2(x; S^*) \quad \text{for all } x \in \mathbb{B}_\varepsilon(x^*).$$

The corollary follows directly from the second part of Theorem 2.2. □

3.4 Discussions on Nuclear Norm Regularized Least Square Optimization Problems

Another important optimization problem, which has received a lot of attention, is that so-called *nuclear norm regularized least square* optimization problem

$$\min_{X \in \mathbb{R}^{p \times q}} h(X) := \frac{1}{2} \|AX - B\|^2 + \mu \|X\|_* \tag{34}$$

Here, $\mathcal{A} : \mathbb{R}^{p \times q} \rightarrow \mathbb{R}^{m \times n}$ is a linear operator, $B \in \mathbb{R}^{m \times n}$, and $\|X\|_*$ is the nuclear norm of X which is defined as the sum of its singular values.

Similar to the development in [8], Q-linear convergence can be derived by assuming (strong) quadratic growth conditions. On the other hand, the following example shows that, different from Lasso problem (29) studied in Sect. 3.3, the (strong) quadratic growth condition is no longer automatically true for the nuclear norm regularized least square optimization problem (34), even when the underlying problem admits a unique solution.

Example 3.1 (Failure of quadratic growth condition for nuclear norm regularized optimization problems) Consider the following optimization problem:

$$\min_{X \in \mathbb{R}^{2 \times 2}} h(X) := \frac{1}{2} \left[(X_{11} + X_{22} - 2)^2 + (X_{12} - X_{21} + X_{22})^2 \right] + \|X\|_*, \tag{35}$$

which is a particular case of (34) with $\mathcal{A}(X) = \begin{bmatrix} X_{11} + X_{22} \\ X_{12} - X_{21} + X_{22} \end{bmatrix}$ for any $X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix} \in \mathbb{R}^{2 \times 2}$, $B = \begin{bmatrix} 2 \\ 0 \end{bmatrix}$, and $\mu = 1$. For $X := \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, let σ_1 and σ_2 be the singular value of X , we have

$$\|X\|_* = \sigma_1 + \sigma_2 = \sqrt{\sigma_1^2 + \sigma_2^2 + 2\sigma_1\sigma_2} = \sqrt{\|X\|_F^2 + 2|\det(X)|}. \tag{36}$$

Given $X = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, it follows that

$$\begin{aligned} h(X) &= \frac{1}{2} \left[(a + d - 2)^2 + (b - c + d)^2 \right] + \sqrt{a^2 + b^2 + c^2 + d^2 + 2|ad - bc|} \\ &\geq \frac{1}{2} \left[(a + d - 2)^2 + (b - c + d)^2 \right] + \sqrt{a^2 + b^2 + c^2 + d^2 + 2(ad - bc)} \\ &\geq \frac{1}{2} (a + d - 2)^2 + \sqrt{(a + d)^2 + (b - c)^2} \\ &\geq \frac{1}{2} (a + d - 2)^2 + (a + d) \\ &= \frac{1}{2} (a + d - 1)^2 + \frac{3}{2}. \end{aligned}$$

Moreover, $h(X) = \frac{3}{2}$ if $a + d = 1$, $ad - bc \geq 0$, $b - c = 0$, $b - c + d = 0$, which means $b = c = d = 0$ and $a = 1$. Thus, $\bar{X} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$ is the unique optimal solution

to problem (35). Choose $X_\varepsilon = \begin{bmatrix} 1 - \varepsilon^{1.5} & \varepsilon - \varepsilon^{1.5} \\ \varepsilon & \varepsilon^{1.5} \end{bmatrix}$ with $\varepsilon > 0$ sufficiently small and note that

$$\begin{aligned} h(X_\varepsilon) &= \frac{1}{2} + \sqrt{(1 - \varepsilon^{1.5})^2 + (\varepsilon - \varepsilon^{1.5})^2 + \varepsilon^2 + \varepsilon^3 + 2|(1 - \varepsilon^{1.5})\varepsilon^{1.5} - (\varepsilon - \varepsilon^{1.5})\varepsilon|} \\ &= \frac{1}{2} + \sqrt{(1 - \varepsilon^{1.5})^2 + (\varepsilon - \varepsilon^{1.5})^2 + \varepsilon^2 + \varepsilon^3 + 2(1 - \varepsilon^{1.5})\varepsilon^{1.5} - 2(\varepsilon - \varepsilon^{1.5})\varepsilon} \\ &= \frac{1}{2} + \sqrt{1 + \varepsilon^3} \\ &= h(\bar{X}) + \sqrt{1 + \varepsilon^3} - 1 = h(\bar{X}) + \mathcal{O}(\varepsilon^3). \end{aligned}$$

Observe further that $\|X_\varepsilon - \bar{X}\|_F^2 = \varepsilon^3 + (\varepsilon - \varepsilon^{1.5})^2 + \varepsilon^2 + \varepsilon^3 = \mathcal{O}(\varepsilon^2)$. This tells us that \bar{X} does not satisfy the strong quadratic growth condition for (35). Note that \bar{X} is the unique solution, we also see that the quadratic growth condition (2) fails at \bar{X} .

Remark 3.3 Moreover, by setting X^0 as the identity matrix, $\sigma = 1$, and $\theta = \frac{1}{2}$, we solve problem (35) numerically by FBS (6) with the Beck–Teboulle’s line search and store the quotients $\delta_k := \frac{h(X^{k+1}) - h(\bar{X})}{h(X^{k+1}) - h(\bar{X})}$ and $\eta_k := \frac{\|X^{k+1} - \bar{X}\|_F}{\|X^k - \bar{X}\|_F}$. After 276 iterations, both δ_k and η_k are close to 1 with error 10^{-14} . This suggests that Q-linear convergence unlikely occurs for both sequences $\{h(X^{k+1}) - h(\bar{X})\}$ and $\{\|X^k - \bar{X}\|_F\}$.

The quadratic growth condition of nuclear norm regularized problem was studied in [53] under the nondegeneracy condition [53]²: $0 \in \text{ri } \partial h(\bar{X})$, where $\text{ri } \partial h(\bar{X})$ is the *relative interior* of $\partial h(\bar{X})$. In general, although the nondegeneracy condition is an important property in matrix optimization, it can be restrictive for some applications. Without assuming the nondegeneracy condition for (34), the strong quadratic growth condition can be used to guarantee the linear convergence of FBS as in Corollary 2.1 and Sect. 3.2. The strong quadratic growth condition for problem (34) can be characterized via second-order analysis on nuclear norm [17, 52]. On the other hand, the corresponding characterizations are highly non-trivial and are presented in a rather complicated form which may not be able to be easily verified in general. How to obtain easily verifiable and computationally tractable condition ensuring (strong) quadratic growth condition for nuclear norm regularized optimization problem or more generally for matrix optimization problem deserves a separate study and is out of scope of this current paper.

4 Uniqueness of Optimal Solution to ℓ_1 -Regularized Least Square Optimization Problems

As discussed in Sect. 1, the linear convergence of ISTA for Lasso was sometimes obtained by imposing an additional assumption that Lasso has a unique optimal solution x^* ; see, e.g., [41]. Since F_3 satisfies the quadratic growth condition at x^* (3.2),

² Another explanations for the failure of quadratic growth condition (2) in problem (35) of Example 3.1 is that the nondegeneracy condition does not satisfy here

the uniqueness of x^* is equivalent to the strong quadratic growth condition of F_3 at x^* . This observation together with Theorem 3.1 allows us to characterize the uniqueness of optimal solution to Lasso in the next result. A different characterization for this property could be found in [51, Theorem 2.1]. Suppose that x^* is an optimal solution, which means $-A^T(Ax^* - b) \in \partial(\mu\|\cdot\|_1)(x^*)$. In the spirit of Proposition 3.1 with $f(x) = \frac{1}{2}\|Ax - b\|^2$, define

$$\begin{aligned} \mathcal{E} &:= \{j \in \{1, \dots, n\} \mid |(A^T(Ax^* - b))_j| = \mu\}, \\ K &:= \{j \in \mathcal{E} \mid x_j^* = 0\}, \quad J := \mathcal{E} \setminus K. \end{aligned} \tag{37}$$

Since $-A^T(Ax^* - b) \in \partial(\mu\|\cdot\|_1)(x^*)$, if $x_j^* \neq 0$, then $(A^T(Ax^* - b))_j = -\mu \operatorname{sign}(x_j^*)$. This tells us that $J = \{j \in \{1, \dots, n\} \mid x_j^* \neq 0\} := \operatorname{supp}(x^*)$. Furthermore, given an index set $I \subset \{1, \dots, n\}$, we denote A_I by the submatrix of A formed by its columns $A_i, i \in I$ and x_I by the subvector of $x \in \mathbb{R}^n$ formed by $x_i, i \in I$. For any $x \in \mathbb{R}^n$, we also define $\operatorname{sign}(x) := (\operatorname{sign}(x_1), \dots, \operatorname{sign}(x_n))^T$ and $\operatorname{Diag}(x)$ by the square diagonal matrix with the main entries x_1, x_2, \dots, x_n .

Theorem 4.1 (Uniqueness of optimal solution to Lasso problem) *Let x^* be an optimal solution to problem (29). The following statements are equivalent:*

- (i) x^* is the unique optimal solution to Lasso (29).
- (ii) The system $A_J x_J - A_K Q_K x_K = 0$ and $x_K \in \mathbb{R}_+^K$ has a unique solution $(x_J, x_K) = (0_J, 0_K) \in \mathbb{R}^J \times \mathbb{R}^K$, where $Q_K := \operatorname{Diag}[\operatorname{sign}(A_K^T(A_J x_J^* - b))]$.
- (iii) The submatrix A_J has full column rank and the columns of $A_J A_J^\dagger A_K Q_K - A_K Q_K$ are positively linearly independent in the sense that

$$\operatorname{Ker}(A_J A_J^\dagger A_K Q_K - A_K Q_K) \cap \mathbb{R}_+^K = \{0_K\}, \tag{38}$$

where $A_J^\dagger := (A_J^T A_J)^{-1} A_J^T$ is the Moore–Penrose pseudoinverse of A_J .

- (iv) The submatrix A_J has full column rank and there exists a Slater point $y \in \mathbb{R}^m$ such that

$$(Q_K A_K^T A_J A_J^\dagger - Q_K A_K^T) y < 0. \tag{39}$$

Proof Since F_3 satisfies the quadratic growth condition at x^* as in Lemma 3.2, (i) means that F_3 satisfies the strong quadratic growth condition at x^* . Thus, by Theorem 3.1, (i) is equivalent to

$$(\mathcal{H}_{\mathcal{E}}(x^*)u, u) > 0 \quad \text{for all } u \in \mathcal{U} \setminus \{0\} \tag{40}$$

with $f(x) = \frac{1}{2}\|Ax - b\|^2$ and $\mathcal{U} = \{u \in \mathbb{R}^{\mathcal{E}} \mid u_j(\nabla f(x^*))_j \leq 0, j \in K\}$. Note that $\mathcal{H}_{\mathcal{E}} = [\nabla^2 f(x^*)_{i,j}]_{i,j \in \mathcal{E}} = [(A^T A)_{i,j}]_{i,j \in \mathcal{E}} = A_{\mathcal{E}}^T A_{\mathcal{E}}$. Hence, (40) means the system

$$0 = A_{\mathcal{E}} u = A_J u_J + A_K u_K \quad \text{and} \quad u_K \in \mathcal{U}_K \tag{41}$$

has a unique solution $u = (u_J, u_K) = (0_J, 0_K) \in \mathbb{R}^J \times \mathbb{R}^K$, where \mathcal{U}_K is defined by

$$\mathcal{U}_K := \{u \in \mathbb{R}^K \mid u_k(A^T(Ax^* - b))_k \leq 0, k \in K\}.$$

As observed after (37), $J = \text{supp}(x^*)$, for each $k \in K$ we have

$$(A^T(Ax^* - b))_k = (A^T(A_Jx_J^* - b))_k = (A_K^T(A_Jx_J^* - b))_k.$$

It follows that $\mathcal{U}_K = -\mathcal{Q}_K(\mathbb{R}_+^K)$ and \mathcal{Q}_K is a nonsingular diagonal square matrix (each diagonal entry is either 1 or -1). Uniqueness of system (41) is equivalent to (ii). This verifies the equivalence between (i) and (ii).

Let us justify the equivalence between (ii) and (iii). To proceed, suppose that (ii) is valid, i.e., the system

$$A_Jx_J - A_K\mathcal{Q}_Kx_K = 0 \quad \text{with} \quad (x_J, x_K) \in \mathbb{R}^J \times \mathbb{R}_+^K. \quad (42)$$

has a unique solution $(0_J, 0_K) \in \mathbb{R}^J \times \mathbb{R}^K$. Choose $x_K = 0_K$, the latter tells us that equation $A_Jx_J = 0$ has a unique solution $x_J = 0$, i.e., A_J has full column rank. Thus, $A_J^T A_J$ is nonsingular. Furthermore, it follows from (42) that $A_J^T A_J x_J = A_J^T A_K \mathcal{Q}_K x_K$, which means

$$x_J = (A_J^T A_J)^{-1} A_J^T A_K \mathcal{Q}_K x_K = A_J^\dagger A_K \mathcal{Q}_K x_K. \quad (43)$$

This together with (42) tells us that the system

$$A_J A_J^\dagger A_K \mathcal{Q}_K x_K - A_K \mathcal{Q}_K x_K = (A_J A_J^\dagger A_K \mathcal{Q}_K - A_K \mathcal{Q}_K) x_K = 0, \quad x_K \in \mathbb{R}_+^K \quad (44)$$

has a unique solution $x_K = 0_K \in \mathbb{R}^K$, which clearly verifies (38) and thus (iii).

To justify the converse implication, suppose that (iii) is valid. Consider equation (42) in (ii), since A_J has the full rank column, we also have (43). Similar to the above justification, one sees that x_K satisfies equation (44). Thanks to (38) in (iii), we get from (44) that $x_K = 0_K$ and thus $x_J = 0_J$ by (43). This verifies that equation (42) in (ii) has a unique solution $(x_J, x_K) = (0_J, 0_K)$.

Finally, the equivalence between (iii) and (iv) follows from the well-known Gordan's lemma [11, Theorem 2.2.1] and the fact that the matrix $A_J A_J^\dagger$ is symmetric. \square

Next, let us discuss some known conditions relating the uniqueness of optimal solution to Lasso. In [23], Fuchs introduced a sufficient condition for the above property:

$$A_J^T (A_J x_J^* - b) = -\mu \text{sign}(x_J^*), \quad (45)$$

$$\|A_{J^c}^T (A_J x_J^* - b)\|_\infty < \mu, \quad (46)$$

$$A_J \text{ has full column rank.} \quad (47)$$

The first equality (45) indeed tells us that x^* is an optimal solution to Lasso problem. Inequality (46) means that $\mathcal{E} = J$, i.e., $K = \emptyset$ in Theorem 4.1. (47), is also present in our characterizations. Hence, Fuchs' condition implies (iii) in Theorem 4.1 and is clearly not a necessary condition for the uniqueness of optimal solution to Lasso problem, since in many situations the set K is not empty.

Furthermore, in the recent work [43] Tibshirani shows that the optimal solution x^* to problem (29) is unique when the matrix $A_{\mathcal{E}}$ has full column rank. This condition is sufficient for our (ii) in Theorem 4.1. Indeed, if (x_J, x_K) satisfies system (42) in (ii), we have $A_{\mathcal{E}} \begin{pmatrix} x_J \\ -Q_K x_K \end{pmatrix} = 0$, which implies that $x_J = 0$ and $Q_K x_K = 0$ when $\ker A_{\mathcal{E}} = 0$. Since Q_K is invertible, the latter tells us that $x_J = 0$ and $x_K = 0$, which clearly verifies (ii). Tibshirani's condition is also necessary for the uniqueness of optimal solution to Lasso problem for *almost all* b in (29), but it is not for all b ; a concrete example could be found in [51].

In the recent works [50, 51], the following useful characterization of unique solution to Lasso has been established under mild assumptions:

$$\begin{aligned} &\text{There exists } y \in \mathbb{R}^m \text{ satisfying } A_J^T y = \text{sign}(x_J^*) \text{ and } \|A_K^T y\|_{\infty} < 1, \\ &A_J \text{ has full column rank.} \end{aligned} \quad (48)$$

It is still open to us to connect this condition directly to those ones in Theorem 4.1, although they must be logically equivalent under the assumptions required in [50, 51]. However, our approach via second-order variational analysis is completely different and also provides several new characterizations for the uniqueness of optimal solution to Lasso. It is also worth mentioning here that the standing assumption in [51] that A has full row rank is relaxed in our study.

5 Conclusion

In this paper, we analyze quadratic growth conditions for some structured optimization problems using second-order variational analysis. This allows us to establish the Q-linear convergence of FBS for

- Poisson regularized optimization problems and Lasso problems with no assumption on the initial data;
- ℓ_1 -regularized optimization problems with mild assumptions via second-order conditions.

As a by-product, we also obtain full characterizations for the uniqueness of optimal solution to Lasso problem, which complements and extends recent important results in the literature.

Our results in this paper point out several interesting research questions, particularly for extending the approach in this paper for matrix optimization problems such as nuclear norm regularized optimization problems.

- Firstly, as we have seen in Example 3.1, for nuclear norm regularized optimization problem (34), (strong) quadratic growth condition can fail even for problems with

unique solutions. Thus, there is a gap between the uniqueness of the solution and strong quadratic growth condition for (34). How to characterize this gap for the nuclear norm regularized optimization problem or, more generally, for matrix optimization problems would be an important research topic to investigate.

In particular, solution uniqueness for problem (34) has been characterized by the so-called *descent cone* [13]. Evaluating the descent cone for nuclear norm will help us understand more about solution uniqueness to (34) and understand the gap between the uniqueness of solution with the strong quadratic growth condition for (34).

- Secondly, what is the tightest possible complexity of FBS in solving the nuclear norm minimization problem? Certainly, the complexity is at least $o(\frac{1}{k})$ as studied in [7–9, 40]. But FBS may fail to exhibit linear convergence when the quadratic growth condition fails, as discussed in Remark 3.3. Due to algebraic structure in nuclear norm, it is natural to conjecture that the complexity is $\mathcal{O}(\frac{1}{k^\beta})$ with some $\beta > 1$. Finding the optimal β is another research direction which deserves further study.

Acknowledgements The authors are indebted to both anonymous referees for their careful readings and thoughtful suggestions that allowed us to improve the original presentation significantly.

Funding Open Access funding enabled and organized by CAUL and its Member Institutions.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Aragón Artacho, F.J., Geoffroy, M.H.: Characterizations of metric regularity of subdifferentials. *J. Convex Anal.* **15**, 365–380 (2008)
2. Aragón Artacho, F.J., Geoffroy, M.H.: Metric subregularity of the convex subdifferential in Banach spaces. *J. Nonlinear Convex Anal.* **15**, 35–47 (2014)
3. Azé, D., Corvellec, J.-N.: Nonlinear local error bounds via a change of metric. *J. Fixed Point Theory Appl.* **16**, 251–372 (2014)
4. Bauschke, H.H., Bolte, J., Teboulle, M.: A descent lemma beyond Lipschitz gradient continuity: first-order methods revisited and applications. *Math. Oper. Res.* **42**, 330–348 (2017)
5. Bauschke, H.H., Combettes, P.L.: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, New York (2011)
6. Bauschke, H.H., Noll, D., Phan, H.M.: Linear and strong convergence of algorithms involving averaged nonexpansive operators. *J. Math. Anal. Appl.* **421**, 1–20 (2015)
7. Beck, A., Teboulle, M.: Gradient-based algorithms with applications to signal recovery problems. in *Convex Optimization in Signal Processing and Communications*, (D. Palomar and Y. Eldar, eds.) 42–88 University Press, Cambridge (2010)
8. Bello-Cruz, J.Y., Li, G., Nghia, T. T.A.: On the Q-linear convergence of forward-backward splitting method. Part I: Convergence analysis. *J. Optim. Theory Appl.* **188**, 378–401 (2021)

9. Bello Cruz, J.Y., Nghia, T.T.A.: On the convergence of the proximal forward-backward splitting method with line searches. *Optim. Method Softw.* **31**, 1209–1238 (2016)
10. Bolte, J., Nguyen, T.P., Peyrouquet, J., Suter, B.W.: From error bounds to the complexity of first-order descent methods for convex functions. *Math. Program.* **165**, 471–507 (2017)
11. Borwein, J., Lewis, A.S.: *Convex analysis and nonlinear optimization: Theory and Examples*. Springer Science & Business Media (2010)
12. Bredies, K., Lorenz, D.A.: Linear convergence of iterative soft-thresholding. *J. Fourier Anal. Appl.* **14**, 813–837 (2008)
13. Chandrasekaran, V., Recht, B., Parrilo, P.A., Willsky, A.S.: The convex geometry of linear inverse problems. *Found Comput Math* **12**, 805–849 (2012)
14. Combettes, P.L., Pesquet, J.-C.: Proximal splitting methods in signal processing. in *Fixed-Point Algorithms for Inverse Problems*. Science and Engineering. Springer Optimization and Its Applications **49**, 185–212 Springer, New York, (2011)
15. Combettes, P.L., Wajs, V.R.: Signal recovery by proximal forward-backward splitting. *Multiscale Model. Simul.* **4**, 1168–1200 (2005)
16. Csiszár, I.: Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Ann. Statist.* **19**, 2032–2066 (1991)
17. Cui, Y., Ding, C., Zhao, X.: Quadratic growth conditions for convex matrix optimization problems associated with spectral functions. *SIAM Journal on Optimization* **27**(4), 2332–2355 (2017)
18. Daubechies, I., Defrise, M., De Mol, D.: An iterative thresholding algorithm for linear inverse problems with a sparsity constraint. *Comm. Pure Appl. Math.* **57**, 1413–1457 (2004)
19. Davis, D., Yin, W.: *Convergence rate analysis of several splitting schemes*. Splitting Methods in Communications, Image Science, and Engineering. Scientific Computation, Springer, Cham, 2016
20. Dontchev, A.L., Rockafellar, R.T.: *Implicit functions and solution mappings. A View from Variational Analysis*. Springer, Dordrecht (2009)
21. Drusvyatskiy, D., Lewis, A.: Error bounds, quadratic growth, and linear convergence of proximal methods. *Math. Oper. Res.* **43**, 693–1050 (2018)
22. Drusvyatskiy, D., Mordukhovich, B.S., Nghia, T.T.A.: Second-order growth, tilt stability, and metric regularity of the subdifferential. *J. Convex Anal.* **21**, 1165–1192 (2014)
23. Fuchs, J.-J.: On sparse representations in arbitrary redundant bases. *IEEE Trans. Inform. Theory.* **50**, 1341–1344 (2004)
24. Grasmair, M., Haltmeier, M., Scherzer, O.: Necessary and sufficient conditions for linear convergence of ℓ_1 regularization. *Comm. Pure Applied Math.* **64**, 161–182 (2011)
25. Garrigos, G., Rosasco, L., and Villa, S.: Convergence of the forward-backward algorithm: beyond the worst case with the help of geometry. [arXiv:1703.09477](https://arxiv.org/abs/1703.09477) (2017)
26. Garrigos, G., Rosasco, L., and Villa, S.: Thresholding gradient methods in Hilbert spaces: support identification and linear convergence, *ESAIM: COCV* **26** (2020), <https://doi.org/10.1051/cocv/2019011>
27. Gilbert, J.C.: On the solution uniqueness characterization in the ℓ_1 norm and polyhedral gauge recovery. *J. Optim. Theory Appl.* **172**, 70–101 (2017)
28. Hale, E.T., Yin, W., Zhang, Y.: Fixed-point continuation for ℓ_1 -minimization: methodology and convergence. *SIAM J. Optim.* **19**, 1107–1130 (2008)
29. Lewis, A.S.: Active sets, nonsmoothness, and sensitivity. *SIAM J. Optim.* **23**, 702–725 (2002)
30. Lewis, A.S., Zhang, S.: Partial smoothness, tilt stability, and generalized Hessians. *SIAM J. Optim.* **23**, 74–94 (2013)
31. Li, G., Pong, T.K.: Calculus of the exponent of Kurdyka-Łojasiewicz inequality and its applications to linear convergence of first-order methods. *Found. Comp. Math.* **18**, 1199–1232 (2018)
32. Liang, J., Fadili, J., Peyré, G.: Local linear convergence of forward-backward under partial smoothness. *Adv. Neural Inf. Process Syst* (2014)
33. Liang, J., Fadili, J., Peyré, G.: Activity identification and local linear convergence of forward-backward type methods. *SIAM J. Optim.* **27**, 408–437 (2017)
34. Luo, Z.-Q., Tseng, P.: Error bounds and convergence analysis of feasible descent methods: a general approach. *Ann. Oper. Res.* **46**, 157–178 (1993)
35. Mordukhovich, B.S.: *Variational Analysis and Generalized Differentiation, I: Basic Theory, II: Applications*. Springer, Berlin (2006)
36. Mousavi, S. and Shen, J.: Solution uniqueness of convex piecewise affine functions based optimization with applications to constrained ℓ_1 -minimization, *ESAIM: Control Optim. Cal. Variations*, **25** (2019), <https://doi.org/10.1051/cocv/2018061>

37. Neal, P., Boyd, S.: Proximal algorithms. *Found. Trends in Optim.* **1**, 127–239 (2014)
38. Necoara, I., Nesterov, Yu., Glineur, F.: Linear convergence of first order methods for non-strongly convex optimization. *Math. Program.* **175**, 69–107 (2019)
39. Rockafellar, R.T., Wets, R.J.-B.: *Variational analysis*. Springer, Berlin (1998)
40. Salzo, S.: The variable metric forward-backward splitting algorithm under mild differentiability assumptions. *SIAM J. Optim.* **27**, 2153–2181 (2017)
41. Tao, S., Boley, D., Zhang, S.: Local linear convergence of ISTA and FISTA on the Lasso problem. *SIAM J. Optim.* **26**, 313–336 (2016)
42. Tibshirani, R.: Regression shrinkage and selection via the Lasso. *J. R. Stat. Soc.* **58**, 267–288 (1996)
43. Tibshirani, R.J.: The Lasso problem and uniqueness. *Electron. J. Stat.* **7**, 1456–1490 (2013)
44. Tropp, J.: Just relax: Convex programming methods for identifying sparse signals in noise. *IEEE Trans. Inform. Theory.* **52**, 1030–1051 (2006)
45. Tseng, P.: A modified forward-backward splitting method for maximal monotone mappings. *SIAM J. Control Optim.* **38**, 431–446 (2000)
46. Tseng, P., Yun, S.: A coordinate gradient descent method for nonsmooth separable minimization. *Math. Program.* **117**, 387–423 (2000)
47. Vardi, Y., Shepp, L.A., Kaufman, L.: A statistical model for positron emission tomography. *J. Amer. Statist. Assoc.* **80**, 8–37 (1985)
48. Wainwright, M.J.: Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1 -constrained quadratic programming (lasso). *IEEE Trans. Inform. Theory.* **55**, 2183–2202 (2009)
49. Yu, P., Li, G., Pong, T.K.: Kurdyka-Lojasiewicz exponent via inf-projection, to appear in *Found. Comput. Math.* (2021). <https://doi.org/10.1007/s10208-021-09528-6>
50. Zhang, H., Yan, M., Yin, W.: One condition for solution uniqueness and robustness of both ℓ_1 -synthesis and ℓ_1 -analysis minimizations. *Adv. Comput. Math.* **42**, 1381–1399 (2016)
51. Zhang, H., Yin, W., Cheng, L.: Necessary and sufficient conditions of solution uniqueness in 1-norm minimization. *J. Optim. Theory Appl.* **164**, 109–122 (2015)
52. Zhang, L., Zhang, N.: Xiao, X: On the second-order directional derivatives of singular values of matrices and symmetric matrix-valued functions. *Set-Valued and Variational Analysis* **21**(3), 557–586 (2013)
53. Zhou, Z., So, A.M.-C.: A unified approach to error bounds for structured convex optimization. *Math. Program.* **165**, 689–728 (2017)

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