Complexity of quadratic penalty methods with adaptive accuracy under a PL condition for the constraints

Florentin Goyens [1] and Geovani N. Grapiglia[1]

[1]ICTEAM Institute, UCLouvain, Louvain-la-Neuve, Belgium
{florentin.goyens,geovani.grapiglia}@uclouvain.be

November 26, 2025

Abstract

We study the quadratic penalty method (QPM) for smooth nonconvex optimization problems with equality constraints. Assuming the constraint violation satisfies the PL condition near the feasible set, we derive sharper worst-case complexity bounds for obtaining approximate first-order KKT points. When the objective and constraints are twice continuously differentiable, we show that QPM equipped with a suitable first-order inner solver requires at most $\mathcal{O}(\varepsilon_0^{-1}\varepsilon_1^{-2})$ first-order oracle calls to find an $(\varepsilon_0, \varepsilon_1)$-approximate KKT point—that is, a point that is ε_0-approximately feasible and ε_1-approximately stationary. Furthermore, when the objective and constraints are three times continuously differentiable, we show that QPM with a suitable second-order inner solver achieves an $(\varepsilon_0, \varepsilon_1)$-approximate KKT point in at most $\mathcal{O}\left(\varepsilon_0^{-1/2}\varepsilon_1^{-3/2}\right)$ second-order oracle calls. We also introduce an adaptive, feasibility-aware stopping criterion for the subproblems, which relaxes the stationarity tolerance when far from feasibility. This rule preserves all theoretical guarantees while substantially reducing computational effort in practice.
Keywords: Quadratic penalty method, worst-case complexity, second-order methods, adaptive subproblem tolerance.

Contents

1 Introduction **2**
- 1.1 Contributions . . . 3
- 1.2 Related literature . . . 3
- 1.3 Contents . . . 4

2 The Quadratic Penalty Method **5**

3 Outer iteration complexity without the PL condition **6**

4 Outer iteration complexity under the PL condition **7**

5 Total evaluation complexity with first-order inner solver **11**
- 5.1 Lemmas on first-order inner solvers . . . 11
- 5.2 First-order inner solver without the PL condition . . . 13
- 5.3 First-order inner solver under the PL condition . . . 13

6 Total evaluation complexity with second-order inner solver **14**
6.1 Lemmas on second-order inner solvers . . . 14
6.2 Second-order inner solver without the PL condition . . . 16
6.3 Second-order inner solver under the PL condition . . . 16

7 Illustrative numerical results **17**

1 Introduction

We consider the equality-constrained optimization problem

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ f(x) \text{ subject to } c(x)=0, \tag{P}$$

where $f\colon \mathbb{R}^n \to \mathbb{R}$ and $c\colon \mathbb{R}^n \to \mathbb{R}^m$ are smooth and possibly nonconvex functions with $m \le n$. The idea of penalty methods is to transform (P) into a sequence of unconstrained optimization problems. At its kth iteration, the Quadratic Penalty Method (QPM) computes x_{k+1} by approximately minimizing the penalized objective

$$Q_{\beta_k}(x) = f(x) + \frac{\beta_k}{2}\|c(x)\|^2, \tag{1.1}$$

where $\beta_k > 0$ is a penalty parameter that determines the weight given to the constraint violation. Each inner minimization (also called subproblem) uses an increasingly larger value of β_k. This raises the cost of infeasibility, thereby pushing iterates closer to the feasible region. As $\beta_k \to \infty$, minimizers of Q_{β_k} ideally approach feasible points of the constrained problem (P). The quadratic penalty method has a long history (Zangwill, 1967; Bertsekas, 1997) and remains an important topic in nonconvex optimization, in part because it serves as a building block for augmented Lagrangian methods.

Traditionally, the theoretical analysis of optimization methods has focused on asymptotic convergence—that is, identifying conditions under which the iterates converge (or have accumulation points converging) to a KKT point of problem (P). More recently, attention has shifted toward *worst-case complexity*, aiming to answer the question:

> *How many iterations (and function evaluations) are required, in the worst case, to generate the first approximate KKT point?*

Given tolerances $\varepsilon_0, \varepsilon_1 > 0$, an $(\varepsilon_0, \varepsilon_1)$-approximate KKT point for (P) is such that for some $\lambda \in \mathbb{R}^m$,

$$\|c(x)\| \le \varepsilon_0, \qquad \text{and} \qquad \|\nabla f(x) - \sum_{i=1}^{m} \lambda_i \nabla c_i(x)\| \le \varepsilon_1. \tag{$(\varepsilon_0,\varepsilon_1)$-KKT}$$

The *outer iteration complexity* of a penalty method refers to the number of subproblems solved (or equivalently, the number of updates of the penalty parameter) before reaching an $(\varepsilon_0, \varepsilon_1)$-KKT point. At each outer iteration, an unconstrained optimization method approximately minimizes the penalized function Q_{β_k}, and the *inner evaluation complexity* is the number of evaluations of f, c and their derivatives needed for the minimization of a subproblem. These two concepts come together in the *total evaluation complexity* of a penalty method, which is the total number of evaluations of f, c, and their derivatives required to find an $(\varepsilon_0, \varepsilon_1)$-KKT point.

Understanding and improving these bounds is central to the design of optimization algorithms, as the insight derived from the complexity analysis may give clues to design methods with better practical performance.

1.1 Contributions

Our work improves complexity bounds of the QPM for obtaining approximate first-order KKT points for (P). We analyze a simple, practical version of the QPM (Algorithm 1), without introducing any auxiliary mechanisms designed to facilitate the theoretical analysis.

Our main contributions are as follows:

1. We introduce an adaptive *feasibility-aware tolerance* for the subproblems, which relaxes the stationarity requirement when far from feasibility, thus avoiding wasted computation in early iterations. This substantially improves practical performance.

2. We improve the best-known *outer iteration complexity* bound of QPM. If the constraint violation satisfies the Polyak–Łojasiewicz (PL) condition near the feasible set (Assumption A4), we show that $\beta_k \geq \mathcal{O}(\varepsilon_0^{-1})$ ensures $\|c(x_{k+1})\| \leq \varepsilon_0$ (Lemma 4.3), thus improving on the existing $\beta_k \geq \mathcal{O}(\varepsilon_0^{-2})$ bound (Grapiglia, 2023).

3. We establish *total evaluation complexity* bounds for the QPM. Under A4, when the objective and constraints are twice continuously differentiable, we show that QPM equipped with a suitable first-order inner solver requires at most $\tilde{\mathcal{O}}(\varepsilon_0^{-1}\varepsilon_1^{-2})$ first-order oracle calls to find an $(\varepsilon_0, \varepsilon_1)$-KKT point. Furthermore, under A4, when the objective and constraints are three times continuously differentiable, we show that QPM with a suitable second-order inner solver achieves an $(\varepsilon_0, \varepsilon_1)$-KKT point in at most $\tilde{\mathcal{O}}\left(\varepsilon_0^{-1/2}\varepsilon_1^{-3/2}\right)$ second-order oracle calls. We do not assume Lipschitz continuity of the derivatives of f and c on a set that contains the iterates. Instead, we only assume that f has a bounded sublevel set (Assumption A5).

   Table 1 summarizes the complexity results established in this paper.

QPM: target $(\varepsilon_0, \varepsilon_1)$-KKT	Without PL	Under PL (A4)
Penalty parameter $\beta_{\max}$	$\mathcal{O}(\varepsilon_0^{-2})$	$\mathcal{O}(\varepsilon_0^{-1})$
Outer iteration complexity	$\mathcal{O}\big(\log\big(\beta_0^{-1}\varepsilon_0^{-2}\big)\big)$	$\mathcal{O}\big(\log\big(\beta_0^{-1}\varepsilon_0^{-1}\big)\big)$
Total evaluation complexity with first-order inner solver (A6)	$\mathcal{O}\big(\log\big(\beta_0^{-1}\varepsilon_0^{-2}\big)\,\varepsilon_0^{-2}\,\varepsilon_1^{-2}\big)$	$\mathcal{O}\big(\log\big(\beta_0^{-1}\varepsilon_0^{-1}\big)\,\varepsilon_0^{-1}\,\varepsilon_1^{-2}\big)$
Total evaluation complexity with second-order inner solver (A9)	$\mathcal{O}\Big(\log\big(\beta_0^{-1}\varepsilon_0^{-2}\big)\,\varepsilon_0^{-1}\,\varepsilon_1^{-3/2}\Big)$	$\mathcal{O}\big(\log\big(\beta_0^{-1}\varepsilon_0^{-1}\big)\,\varepsilon_0^{-1/2}\,\varepsilon_1^{-3/2}\big)$

Table 1: Summary of our complexity results for the quadratic penalty method, with and without the PL condition on the constraint violation.

1.2 Related literature

The worst-case evaluation complexity of methods for constrained nonconvex problems has been widely studied in recent years. For first-order schemes, Cartis et al. (2011) analyzed an exact penalty method, showing a bound of at most $\mathcal{O}(\varepsilon^{-5})$ problem evaluations to reach an ε-KKT point, which corresponds to an $(\varepsilon_0, \varepsilon_1)$-KKT point with $\varepsilon_0 = \varepsilon_1 = \varepsilon$ in our definition. This rate improves to $\mathcal{O}(\varepsilon^{-2})$ under the assumption that the penalty parameters remain bounded. Comparable $\mathcal{O}(\varepsilon^{-2})$ bounds were obtained for a two-phase method (Cartis et al., 2014), inexact restoration methods (Bueno and Martínez, 2020), and using Fletcher's augmented Lagrangian (Goyens et al., 2024).

For sequential quadratic programming schemes, Facchinei et al. (2021) established an $\mathcal{O}(\varepsilon^{-4})$ bound and Curtis et al. (2024) established an $\mathcal{O}(\varepsilon^{-2})$ bound. Second-order methods offer improved complexity: adaptive cubic regularization and trust-funnel schemes attain a worst-case rate of $\mathcal{O}(\varepsilon^{-3/2})$ (Cartis et al., 2013; Curtis et al., 2018), while higher-order models achieve $\mathcal{O}(\varepsilon^{-(p+1)/p})$ for p-th order methods ($p \geq 2$), as shown in (Birgin et al., 2016; Martínez, 2017; Cartis et al., 2022).

Inexact penalty methods, such as the quadratic penalty and augmented Lagrangian methods have also received a lot of attention. Under a non-singularity condition on the constraints, Li et al. (2021) showed a $\tilde{\mathcal{O}}(\varepsilon^{-3})$ bound for an inexact augmented Lagrangian method.

Birgin and Martínez (2020) give complexity results for the Algencan algorithm (Andreani et al., 2008), an augmented Lagrangian method designed for practical performance. They show that the number of outer iteration is logarithmic in several circumstances, and combine it with an inner solver of arbitrary complexity. Grapiglia (2023) shows that the quadratic penalty method (Algorithm 1) with non-adaptive subproblem accuracy $\tau(x) \equiv \varepsilon_1$ and gradient descent with an Armijo linesearch in the subproblems requires at most $\tilde{\mathcal{O}}(\varepsilon^{-4})$ problem evaluations to find an ε-KKT point. For f and c weakly convex, Lin et al. (2020) show a $\tilde{\mathcal{O}}(\varepsilon^{-3})$ for a quadratic penalty method.

For linearly constrained problems, Kong et al. (2019) proved a $\mathcal{O}(\varepsilon^{-3})$ bound for a quadratic penalty method with an accelerated first-order inner solver. Under a Slater-type condition, proximal-point or augmented Lagrangian variants reach $\tilde{\mathcal{O}}(\varepsilon^{-5/2})$, e.g., (Li and Xu, 2021; Lin et al., 2020; Melo et al., 2020). Grapiglia and Yuan (2021) obtained a $\tilde{\mathcal{O}}(\varepsilon^{-(p+1)/p})$ bound for an augmented Lagrangian method with a p-th order inner solver, matching the order of the proximal augmented Lagrangian bound of Xie and Wright (2021). He et al. (2023) later improved this bound, giving a result of $\tilde{\mathcal{O}}(\varepsilon^{-7/2})$ under an LICQ condition and $\tilde{\mathcal{O}}(\varepsilon^{-11/2})$ without LICQ.

Outside the realm of penalty methods, Riemannian optimization methods are designed for instances of (P) where the feasible set is a smooth manifold, with a convenient expression for the tangent space and projection onto the feasible set. Under a Lipschitz continuity assumption, Riemannian gradient descend finds a $(0, \varepsilon_1)$-KKT point in at most $\mathcal{O}(\varepsilon_1^{-2})$ iterations (Boumal et al., 2019). A similar bound holds for a second-order Riemannian trust-region, but is improved to $\mathcal{O}(\log\log(\varepsilon^{-1}))$ for a class of function with strict saddle points (Goyens and Royer, 2024).

Overall, comparing these results is nontrivial, as they address different problem classes, regularity assumptions, and requirements on the quality of the initial point. The present work improves the existing literature by establishing complexity bounds for a QPM that allows inexact subproblem solutions.

1.3 Contents

The paper is organized as follows. In Section 2, we define the Quadratic Penalty Method (QPM). In Section 3, we show that the outer iteration complexity result from (Grapiglia, 2023) still applies when the accuracy in the subproblems uses the adaptive rule (2.4). In Section 4, we derive an improved outer iteration complexity bound under a PL condition on the constraint violation (A4). In Section 5, we give total evaluation complexity bounds for the QPM when a first-order method performs the inner minimization. In Section 6, we give total evaluation complexity bounds when a second-order methods performs the inner minimization. In Section 7, we illustrate our findings empirically and show that the adaptive accuracy in the subproblems improves practical performance.

Notations

Let $\|\cdot\|$ to denote the usual 2-norm in $\mathbb{R}^n$. Let $J(x) \in \mathbb{R}^{m\times n}$ denote the Jacobian of the function c, and $\mathbb{N}^*$ denote the set of positive natural numbers. We use $\mathrm{conv}(A)$ to denote the convex hull of a set A.

2 The Quadratic Penalty Method

We consider the Quadratic Penalty Method described in Algorithm 1. A classical result about the quadratic penalty method (Bertsekas, 1997, Prop. 4.2.1) shows that, under mild assumptions, global minimizers of Q_{β_k} converge to the global minimizer of (P) as $\beta_k \to \infty$. This is of little practical value since it is undesirable to have β_k grow arbitrarily large, as it makes the subproblems badly conditioned. Furthermore, it is in general not possible to find a global minimizer of the quadratic penalty. This raises two important questions:

(i) How rapidly should β_k grow?

(ii) To what accuracy should each subproblem be solved?

At iteration k, we update the penalty parameter as $\beta_{k+1} = \alpha\beta_k$, where $\alpha > 1$ is a parameter of the QPM. We require that the approximate minimizer of Q_{β_k} yields a reduction in the value of Q_{β_k} compared to the current point x_k and the initial point x_0 (2.3); and we also require that it be approximately first-order stationary (2.4).

In (Grapiglia, 2023), at each iteration, the subproblem is solved up to the target accuracy ε_1. However, it is customary in optimization softwares to use a tolerance larger than ε_1 for the stationarity in the early iterations. Various heuristics are used in practice, such as starting with $\sqrt{\varepsilon_1}$ and dividing the tolerance by a factor 10 at each outer iteration (Algencan, Andreani et al. (2008)); another strategy is given in (Eckstein and Silva, 2013). We propose an adaptive and practical tolerance on the gradient norm. The accuracy is proportional to the current level of feasibility, which is a natural way to spare computational resources when the iterates are far from the feasible set.

Definition 2.1 (Feasibility-aware tolerance)**.** *Given tolerances* $\varepsilon_0 > 0, \varepsilon_1 > 0$, *a* feasibility-aware tolerance *is a function* $\tau : \mathbb{R}^n \to \mathbb{R}_+$ *given, for some* $\tau_{\text{cap}} \geq \varepsilon_1$ *by*

$$\tau(x) := \max\Big\{\varepsilon_1,\ \min\Big(\tau_{\text{cap}},\ \tfrac{\varepsilon_1}{\varepsilon_0}\,\|c(x)\|\Big)\Big\}. \tag{2.1}$$

This ensures

$$\tau(x) \geq \varepsilon_1 \quad \text{for all } x \in \mathbb{R}^n, \qquad \text{and} \qquad \tau(x) = \varepsilon_1 \text{ whenever } \|c(x)\| \leq \varepsilon_0. \tag{2.2}$$

Consequently, as soon as $\|c(x_{k+1})\| \leq \varepsilon_0$, the point x_{k+1} satifies $\|\nabla Q_{\beta_k}(x_{k+1})\| \leq \varepsilon_1$, which makes it an $(\varepsilon_0, \varepsilon_1)$-KKT point. In our numerical experiments (Section 7), we compare the choice

$$\tau(x) := \max\left(\varepsilon_1, \frac{\varepsilon_1}{\varepsilon_0}\|c(x)\|\right),$$

corresponding to $\tau_{\text{cap}} = +\infty$; with the non-adaptive choice $\tau(x) \equiv \varepsilon_1$, given by $\tau_{\text{cap}} = 0$.

Algorithm 1 Quadratic Penalty Method (QPM)

Given: $\varepsilon_0 > 0$, $\varepsilon_1 > 0$, $\alpha > 0$, $\beta_0 \geq 1$, feasibility-aware tolerance $\tau\colon \mathbb{R}^n \to \mathbb{R}_+$, and $x_0 \in \mathbb{R}^n$

$k \leftarrow 0$

while true **do**

  Find $x_{k+1} \in \mathbb{R}^n$ an approximate minimizer of Q_{β_k} that satisfies

$$Q_{\beta_k}(x_{k+1}) \leq \min\left\{Q_{\beta_k}(x_k), Q_{\beta_k}(x_0)\right\} \tag{2.3}$$

  and

$$\|\nabla Q_{\beta_k}(x_{k+1})\| \leq \tau(x_{k+1}). \tag{2.4}$$

  if $\|c(x_{k+1})\| \leq \varepsilon_0$ **then**

    Stop and Return x_{k+1}, an $(\varepsilon_0, \varepsilon_1)$-KKT point with Lagrange multipliers $-\beta_k c(x_{k+1})$.

  end if

  $\beta_{k+1} = \alpha\beta_k$.

  $k \leftarrow k + 1$

end while

3 Outer iteration complexity without the PL condition

This section establishes an outer iteration complexity bound for the QPM (Algorithm 1). The main result of this section already appears in (Grapiglia, 2023) with the choice $\tau(x) \equiv \varepsilon_1$. We extend the analysis to include an adaptive feasibility-aware tolerance in (2.4).

This section makes no regularity assumption on the Jacobian of the constraints. Section 4 shows improved outer iteration complexity guarantees under the assumption that the constraint violation satisfies the PL condition.

We state standard assumptions.

A1. *The functions f and c are differentiable.*

A2. *There exists $f_{\text{low}} \in \mathbb{R}$ such that $f(x) \geq f_{low}$ for all $x \in \mathbb{R}^n$.*

A3. *The initial iterate $x_0 \in \mathbb{R}^n$ is such that $\|c(x_0)\| \leq \dfrac{\varepsilon_0}{\sqrt{2}}$.*

The following lemma states (in contrapositive form) that if, at some iteration k, the penalty parameter β_k is sufficiently large—specifically, if $\beta_k \geq \mathcal{O}(\varepsilon_0^{-2})$—, then the QPM terminates; i.e., $\|c(x_{k+1})\| \leq \varepsilon_0$.

Lemma 3.1 (Grapiglia (2023), Corollary 3.3)**.** *Under A1, A2, A3, if $\|c(x_k)\| > \varepsilon_0$ for $k = 1, \ldots, T-1$, then*

$$\beta_k < 4(f(x_0) - f_{\text{low}})\varepsilon_0^{-2}, \qquad \text{for } k = 0, \ldots, T-2. \tag{3.1}$$

Recall that $\mathbb{N}^*$ denotes the set of positive natural numbers. We now show that the number of outer iterations of the QPM is at most a logarithmic factor of the upper bound on the penalty parameter given in Lemma 3.1.

Theorem 3.2 (Outer iterations of QPM without the PL condition)**.** *Under A1, A2, A3, define*

$$T(\varepsilon_0) := \inf\{k \in \mathbb{N}^*\colon\ \|c(x_k)\| \leq \varepsilon_0\}. \tag{3.2}$$

Then,

$$T(\varepsilon_0) < \hat{T}(\varepsilon_0) := 2 + \log_\alpha \left(4(f(x_0) - f_{\text{low}})\beta_0^{-1}\varepsilon_0^{-2}\right), \tag{3.3}$$

and the point $x_{T(\varepsilon_0)}$ generated by Algorithm 1 is an $(\varepsilon_0, \varepsilon_1)$-KKT point of (P), satisfying

$$\left\|c(x_{T(\varepsilon_0)})\right\| \leq \varepsilon_0 \qquad \text{and} \qquad \left\|\nabla f(x_{T(\varepsilon_0)}) + \beta_{T(\varepsilon_0)-1}\sum_{i=1}^m c_i(x_{T(\varepsilon_0)})\nabla c_i(x_{T(\varepsilon_0)})\right\| \leq \varepsilon_1. \tag{3.4}$$

Proof. By definition of $T(\varepsilon_0)$, we have $\|c(x_k)\| > \varepsilon_0$ for $k = 1, \ldots, T(\varepsilon_0) - 1$. By Lemma 3.1, this gives

$$\alpha^k \beta_0 = \beta_k < 4(f(x_0) - f_{\text{low}})\varepsilon_0^{-2} \quad \text{for } k = 0, \ldots, T(\varepsilon_0) - 2. \tag{3.5}$$

Therefore, $T(\varepsilon_0)$ is finite and in particular

$$\alpha^{T(\varepsilon_0)-2}\beta_0 < 4(f(x_0) - f_{\text{low}})\varepsilon_0^{-2}. \tag{3.6}$$

This gives

$$T(\varepsilon_0) < 2 + \log_\alpha \left(4(f(x_0) - f_{\text{low}})\beta_0^{-1}\varepsilon_0^{-2}\right) =: \hat{T}(\varepsilon_0). \tag{3.7}$$

Finally we have $\left\|c(x_{T(\varepsilon_0)})\right\| \leq \varepsilon_0$ by construction, which in turn ensures

$$\left\|\nabla Q_{\beta_{T(\varepsilon_0)-1}}(x_{T(\varepsilon_0)})\right\| \leq \tau(x_{T(\varepsilon_0)}) = \varepsilon_1. \tag{3.8}$$

To conclude, the identity

$$\left\|\nabla Q_{\beta_{T(\varepsilon_0)-1}}(x_{T(\varepsilon_0)})\right\| = \left\|\nabla f(x_{T(\varepsilon_0)}) + \beta_{T(\varepsilon_0)-1}\sum_{i=1}^m c_i(x_{T(\varepsilon_0)})\nabla c_i(x_{T(\varepsilon_0)})\right\|, \tag{3.9}$$

implies that $x_{T(\varepsilon_0)}$ is an $(\varepsilon_0, \varepsilon_1)$-KKT point with Lagrange multipliers $-\beta_{T(\varepsilon_0)-1}c(x_{T(\varepsilon_0)})$. □

In the next section, we make an additional assumption—that the constraint violation satisfies a PL condition near the feasible set. This allows to improve the complexity result.

4 Outer iteration complexity under the PL condition

In this section, we show a new complexity result for the QPM, which improves the bound of Theorem 3.2. The analysis of the previous section merely uses the decrease of the penalty at each subproblem (2.3). To obtain a sharper bound, we leverage the approximate stationarity condition (2.4). The most natural way to do so is to work under an additional regularity condition on the Jacobian of the constraints.

A4. *There exists positive constants $R > \varepsilon_0$ and $\sigma_{\min}$ such that for all x in the set*

$$\mathcal{C}_R := \{x \in \mathbb{R}^n \colon \|c(x)\| \leq R\}, \tag{4.1}$$

we have

$$\left\|J(x)^\top c(x)\right\| \geq \sigma_{\min} \|c(x)\|, \tag{4.2}$$

where $J(x) \in \mathbb{R}^{m\times n}$ denotes the Jacobian of c.

Note that (4.2) is the Polyak–Łojasiewicz (PL) condition for the constraint violation

$$P(x) = \tfrac{1}{2}\|c(x)\|^2.$$

Indeed, we have

$$P(x) \le \frac{1}{2\sigma_{\min}^2}\,\|\nabla P(x)\|^2 \qquad \text{for all } x \in \mathcal{C}_R. \tag{4.3}$$

It is weaker than the usual LICQ condition at a point x which demands that $J(x)$ be full rank, but we require a global lower bound $\sigma_{\min} > 0$ in a tubular neighbourhood of the feasible set. We also require that a sublevel set of f be bounded.

A5. *The set $\mathcal{L}_f\left(2f(x_0) - f_{\text{low}}\right)$ is bounded.*

First, we provide several examples which satisfy our assumption.

Example 1. *The problem*

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x) \\ \textit{subject to} \quad & Ax = b, \end{aligned} \tag{4.4}$$

with $\operatorname{rank}(A) = m$ *and f coercive (i.e., $\lim_{\|x\|\to\infty} f(x) = +\infty$). This problem satisfies A4 and A5.*

Example 2. *The problem*

$$\begin{aligned} \min_{X\in\mathbb{R}^{n\times p}} \quad & f(x) \\ \textit{subject to} \quad & X^\top X = \mathrm{I}_p, \end{aligned} \tag{4.5}$$

where f is coercive, satisfies A4 and A5.

Proof. See (Goyens et al., 2022, p.4), which states that for any $1 \le p \le n$, the function $X \mapsto c(X) = X^\top X - \mathrm{I}_p$ satisfies A4 for any $\varepsilon_0 < R < 1$ and $\sigma_{\min} \le 2\sqrt{1-R}$. □

Example 3 (Binary constraints satisfy A4.)**.** *For the binary constraint $x_i \in \{0,1\}$, consider the constraint mapping $c\colon \mathbb{R}^n \to \mathbb{R}^n$ with*

$$c_i(x) = x_i(1-x_i) \qquad \textit{for } i = 1,\dots,n. \tag{4.6}$$

Then, A4 holds with any R such that $\varepsilon_0 < R < \frac{1}{4}$ and $\sigma_{\min} = \sqrt{1-4R} > 0$.

Proof. The Jacobian of c is diagonal, $J(x) = \operatorname{diag}\big(1-2x_1,\dots,1-2x_n\big)$. Using the identity

$$(1-2x_i)^2 = 1 - 4c_i(x),$$

we obtain, for any x such that $\|c(x)\| \le R$:

$$\begin{aligned} \|J(x)^\top c(x)\|^2 &= \sum_{i=1}^n (1-2x_i)^2\, c_i(x)^2 && (4.7)\\ &= \sum_{i=1}^n (1-4c_i)\, c_i(x)^2 && (4.8)\\ &\ge (1-4R)\sum_{i=1}^n c_i(x)^2 && (4.9)\\ &= (1-4R)\|c(x)\|^2. && (4.10) \end{aligned}$$

Thus,

$$\|J(x)^\top c(x)\| \ge \sqrt{1-4R}\,\|c(x)\| \qquad \text{for all } x \in \mathcal{C}_R.$$

□

Furthermore, combining several nonsingular constraints gives a constraint function that satisfies A4.

Proposition 4.1 (Goyens et al. (2024), *Prop.1.1*)**.** *For $i = 1, 2, \dots, k$, consider k functions $c_i\colon \mathbb{R}^n \to \mathbb{R}^{m_i}$ that satisfy assumptions A4 with constants R_i and $\underline{\sigma}_i$. Then, the function $c\colon \mathbb{R}^n \to \mathbb{R}^m$ with $m = m_1 + \cdots + m_k$ defined by $c(x) = (c_1(x), \dots, c_k(x))^\top$ satisfies A4 with constants $R = \min(R_1, \dots, R_k)$ and $\underline{\sigma} = \min(\underline{\sigma}_1, \dots, \underline{\sigma}_k)$.*

We now turn to the complexity analysis under A4, and begin with a lemma stating that the iterates of the QPM are contained in a sublevel set of the function f.

Lemma 4.2. *Under A1, A2, A3, if $\|c(x_k)\| > \varepsilon_0$ for $k = 1, \dots, T-1$, then*

$$x_k \in \mathcal{L}_f(2f(x_0) - f_{\text{low}}) := \{x \in \mathbb{R}^n : f(x) \le 2f(x_0) - f_{\text{low}}\}, \tag{4.11}$$

for $k = 0, \dots, T-1$.

Proof. Clearly $x_0 \in \mathcal{L}_f(2f(x_0) - f_{\text{low}})$. Let $k \in \{1, \dots, T-1\}$. The decrease condition $Q_{\beta_{k-1}}(x_k) \le Q_{\beta_{k-1}}(x_0)$ (2.3), combined with Lemma 3.1 and $\|c(x_0)\| \le \frac{\varepsilon_0}{\sqrt{2}}$ gives

$$\begin{aligned} f(x_k) &\le f(x_k) + \frac{\beta_{k-1}}{2}\|c(x_k)\|^2 && (4.12)\\ &\le f(x_0) + \frac{\beta_{k-1}}{2}\|c(x_0)\|^2 && (4.13)\\ &\le f(x_0) + 2(f(x_0) - f_{\text{low}})\varepsilon_0^{-2}\frac{\varepsilon_0^2}{2} && (4.14)\\ &= 2f(x_0) - f_{\text{low}}. \end{aligned}$$

□

The next lemma, shows that if $\beta_k \ge \mathcal{O}(\varepsilon_0^{-1})$, then the QPM terminates; i.e., $\|c(x_{k+1})\| \le \varepsilon_0$.

Lemma 4.3. *Under A1, A2, A3, A4, A5, if $\|c(x_k)\| > \varepsilon_0$ for $k = 1, \dots, T-1$, then*

$$\beta_k < \max\left\{\frac{L_{f,0} + \varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0) - f_{\text{low}})}{R}\right\}\varepsilon_0^{-1} \quad \text{for } k = 0, 1, \dots, T-2, \tag{4.15}$$

where $L_{f,0} \ge 0$ is such that

$$\|\nabla f(x)\| \le L_{f,0} \quad \text{for all } x \in \mathcal{L}_f(2f(x_0) - f_{\text{low}}). \tag{4.16}$$

Proof. Note that $L_{f,0} \ge 0$ is well defined since $\mathcal{L}_f(2f(x_0) - f_{\text{low}})$ is bounded (A5). Suppose by contradiction that there exists $k \in \{0, \dots, T-2\}$ such that $\|c(x_{k+1})\| > \varepsilon_0$ and

$$\beta_k \ge \max\left\{\frac{L_{f,0} + \varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0) - f_{\text{low}})}{R}\right\}\varepsilon_0^{-1}. \tag{4.17}$$

The condition $Q_{\beta_k}(x_{k+1}) \le Q_{\beta_k}(x_0)$ yields

$$f(x_{k+1}) + \frac{\beta_k}{2}\|c(x_{k+1})\|^2 \le f(x_0) + \frac{\beta_k}{2}\|c(x_0)\|^2. \tag{4.18}$$

Since $\varepsilon_0 < R$, condition (4.17) implies that $\beta_k \ge 4(f(x_0) - f_{\text{low}})/R^2$. Using $\|c(x_0)\|^2 \le \frac{\varepsilon_0^2}{2}$, it follows

$$\begin{aligned} \|c(x_{k+1})\|^2 &\le \frac{2(f(x_0) - f_{\text{low}})}{\beta_k} + \|c(x_0)\|^2 && (4.19)\\ &\le \frac{R^2}{2} + \frac{R^2}{2} \le R^2. && (4.20) \end{aligned}$$

Thus, x_{k+1} belongs to the regular region $\mathcal{C}_R$, which ensures

$$\left\|J(x_{k+1})^\top c(x_{k+1})\right\| \geq \sigma_{\min} \|c(x_{k+1})\| . \tag{4.21}$$

Notice that $\nabla Q_\beta(x) = \nabla f(x) + \beta J(x)^\top c(x)$. Therefore, the first-order condition (2.4) gives

$$\begin{aligned}
\tau(x_{k+1}) &\geq \|\nabla Q_{\beta_k}(x_{k+1})\| & (4.22)\\
&= \left\|\nabla f(x_{k+1}) + \beta_k J(x_{k+1})^\top c(x_{k+1})\right\| & (4.23)\\
&\geq \beta_k \left\|J(x_{k+1})^\top c(x_{k+1})\right\| - \|\nabla f(x_{k+1})\| & (4.24)\\
&\geq \beta_k \sigma_{\min} \|c(x_{k+1})\| - \|\nabla f(x_{k+1})\| , & (4.25)
\end{aligned}$$

which gives

$$\sigma_{\min}\beta_k \|c(x_{k+1})\| \leq \tau(x_{k+1}) + \|\nabla f(x_{k+1})\| . \tag{4.26}$$

By Lemma 4.2, $x_{k+1} \in \mathcal{L}_f(2f(x_0) - f_{\text{low}})$, and therefore $\|\nabla f(x_{k+1})\| \leq L_{f,0}$. Since $\|c(x_{k+1})\| > \varepsilon_0$,

$$\tau(x_{k+1}) = \max\left\{\varepsilon_1, \min\left(\tau_{\text{cap}}, \frac{\varepsilon_1}{\varepsilon_0}\|c(x_{k+1})\|\right)\right\} \leq \max\left\{\varepsilon_1, \frac{\varepsilon_1}{\varepsilon_0}\|c(x_{k+1})\|\right\} = \frac{\varepsilon_1}{\varepsilon_0}\|c(x_{k+1})\| .$$

This gives

$$\sigma_{\min}\beta_k \|c(x_{k+1})\| \leq \frac{\varepsilon_1}{\varepsilon_0}\|c(x_{k+1})\| + L_{f,0}, \tag{4.27}$$

that is,

$$\|c(x_{k+1})\| \leq \frac{L_{f,0}}{\sigma_{\min}\beta_k - \frac{\varepsilon_1}{\varepsilon_0}}. \tag{4.28}$$

Since $\beta_k \geq \frac{L_{f,0}+\varepsilon_1}{\sigma_{\min}\varepsilon_0}$, we have $\|c(x_{k+1})\| \leq \varepsilon_0$, which is a contradiction. □

We now state our outer iteration bound for the QPM under the PL condition on the constraint violation.

Theorem 4.4 (Outer iterations of QPM under the PL condition)**.** *Under A1, A2, A3, A4, A5, let*

$$T(\varepsilon_0) := \inf\{k \in \mathbb{N}^* \colon \|c(x_k)\| \leq \varepsilon_0\}. \tag{4.29}$$

Then,

$$T(\varepsilon_0) \leq \tilde{T}(\varepsilon_0) := 2 + \log_\alpha\left(\max\left\{\frac{L_{f,0}+\varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0) - f_{\text{low}})}{R}\right\}\beta_0^{-1}\varepsilon_0^{-1}\right), \tag{4.30}$$

and the point $x_{T(\varepsilon_0)}$ generated by Algorithm 1 is an $(\varepsilon_0, \varepsilon_1)$-KKT point of (P)*, satisfying*

$$\left\|c(x_{T(\varepsilon_0)})\right\| \leq \varepsilon_0 \quad \textit{and} \quad \left\|\nabla f(x_{T(\varepsilon_0)}) + \beta_{T(\varepsilon_0)-1}\sum_{i=1}^m c_i(x_{T(\varepsilon_0)})\nabla c_i(x_{T(\varepsilon_0)})\right\| \leq \varepsilon_1. \tag{4.31}$$

Proof. By definition of $T(\varepsilon_0)$, we have $\|c(x_k)\| > \varepsilon_0$ for $k = 1, \ldots, T(\varepsilon_0) - 1$. Lemma 4.3 gives

$$\alpha^k \beta_0 = \beta_k < \beta_{\max} := \max \left\{ \frac{L_{f,0} + \varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0) - f_{\text{low}})}{R} \right\} \varepsilon_0^{-1} \quad \text{for } k = 1, \ldots, T(\varepsilon_0) - 2.$$

Therefore, $T(\varepsilon_0)$ is finite and in particular

$$\alpha^{T(\varepsilon_0)-2} \beta_0 = \beta_{T(\varepsilon_0)-2} < \beta_{\max}. \tag{4.32}$$

This gives

$$T(\varepsilon_0) \leq 2 + \log_\alpha \left(\frac{\beta_{\max}}{\beta_0} \right) =: \tilde{T}(\varepsilon_0). \tag{4.33}$$

Finally we have $\|c(x_{T(\varepsilon_0)})\| \leq \varepsilon_0$ by construction, which ensures that $x_{T(\varepsilon_0)}$ is an $(\varepsilon_0, \varepsilon_1)$-KKT point with multipliers $-\beta_{T(\varepsilon_0)-1} c(x_{T(\varepsilon_0)})$ (see the proof of Theorem 3.2). □

Remark 4.1. *As the proof of Lemma 4.3 suggests, we can replace the requirement $\|c(x_0)\| \leq \varepsilon_0/\sqrt{2}$ by $x_0 \in \mathcal{C}_R$ and maintain convergence, provided that the norm of ∇f remains bounded over the iterates.*

5 Total evaluation complexity with first-order inner solver

In this section, we derive total evaluation complexity bounds for the QPM (Algorithm 1), which give an upper bound on the number of evaluations of f, c and their derivatives. In fact, the inner evaluation complexity is measured in evaluations of Q_{β_k}, ∇Q_{β_k}, and $\nabla^2 Q_{\beta_k}$; each such evaluation requires evaluating f, c, and their corresponding derivatives.

We present results where the inner solver of the quadratic penalty method is a first-order method. We detail the results with and without the PL condition on the constraint violation.

5.1 Lemmas on first-order inner solvers

We suppose that a first-order method—called $\mathcal{A}_1$—minimizes Q_{β_k} in each subproblem of the QPM. The following assumption on the (inner) evaluation complexity is typical of first-order methods for unconstrained nonconvex optimization.

A6. *Given a continuously differentiable function $F : \mathbb{R}^n \to \mathbb{R}$ with*

$$\mathcal{L}_F(\tilde{x}) = \{z \in \mathbb{R}^n \mid F(z) \leq F(\tilde{x})\}$$

bounded for some $\tilde{x} \in \mathbb{R}^n$. The method $\mathcal{A}_1$ starting from $\tilde{x}$ needs at most

$$C_{\mathcal{A}_1} L_1 \big(F(\tilde{x}) - F_{low}\big)\, \varepsilon^{-2}$$

evaluations of F and ∇F to generate a point satisfying $\|\nabla F(x)\| \leq \varepsilon$, where F_{low} is a lower bound of F, the positive constant $C_{\mathcal{A}_1}$ depends only on the method $\mathcal{A}_1$, and ∇F is L_1-Lipschitz continuous on $\operatorname{conv}(\mathcal{L}_F(F(\tilde{x})))$.

The traditional Gradient Descent Method with Armijo line search satisfies A6 (Grapiglia, 2023, Appendix 1).

We make the following smoothness assumption.

A7. *The functions f and c are twice continuously differentiable.*

Recall that $P(x) = \frac{1}{2}\|c(x)\|^2$.

Lemma 5.1. *Under A5 and A7, there exists constants $L_{f,1}$ and $L_{P,1} > 0$ such that*

$$\left\|\nabla^2 f(x)\right\| \le L_{f,1} \qquad \textit{and} \qquad \left\|\nabla^2 P(x)\right\| \le L_{P,1} \tag{5.1}$$

for all $x \in \operatorname{conv}(\mathcal{L}_f(2f(x_0) - f_{\text{low}}))$.

In addition, for any $\beta > 0$, we have that ∇Q_β is $(L_{f,1} + \beta L_{P,1})$-Lipschitz continuous on $\operatorname{conv}(\mathcal{L}_f(2f(x_0) - f_{\text{low}}))$.

Proof. By A5, the sublevel set $\mathcal{L}_f(2f(x_0) - f_{\text{low}})$ is bounded, so its convex hull is bounded. By A7, the Hessians $\nabla^2 f$ and $\nabla^2 P$ are continuous, hence there exists constants $L_{f,1}$ and $L_{P,1} > 0$ such that (5.1) holds. Thus, for all x in $\operatorname{conv}(\mathcal{L}_f(2f(x_0) - f_{\text{low}}))$ and any $\beta > 0$,

$$\left\|\nabla^2 Q_\beta(x)\right\| \le \left\|\nabla^2 f(x)\right\| + \beta \left\|\nabla^2 P(x)\right\| \tag{5.2}$$
$$\le L_{f,1} + \beta L_{P,1}. \tag{5.3}$$

Therefore, ∇Q_β is $(L_{f,1} + \beta L_{P,1})$-Lipschitz continuous on $\operatorname{conv}(\mathcal{L}_f(2f(x_0) - f_{\text{low}}))$. □

The next lemma establishes the inner evaluation complexity of the QPM with a first-order solver.

Lemma 5.2 (Inner evaluation complexity of $\mathcal{A}_1$)**.** *Assume A2, A5 and A7. Suppose that $\|c(x_k)\| > \varepsilon_0$ and a monotone first-order method $\mathcal{A}_1$ is used to minimize Q_{β_k} starting from*

$$\tilde{x}_{k,0} := \operatorname*{argmin}_{x \in \{x_k, x_0\}} Q_{\beta_k}(x). \tag{5.4}$$

If $\mathcal{A}_1$ satisfies A6, the method $\mathcal{A}_1$ generates x_{k+1} satisfying the subproblem conditions (2.3) and (2.4) in at most

$$2C_{\mathcal{A}_1}(L_{f,1} + \beta_k L_{P,1})(f(x_0) - f_{\text{low}})\varepsilon_1^{-2} \tag{5.5}$$

evaluations of f, c, and their first-order derivatives, where $L_{f,1}$ and $L_{P,1}$ are defined in Lemma 5.1.

Proof. First note that, by A2, Q_{β_k} is has the lower bound:

$$Q_{\beta_k}(x) = f(x) + \frac{\beta_k}{2}\|c(x)\|_2^2 \ge f(x) \ge f_{\text{low}}, \quad \text{for all } x \in \mathbb{R}^n. \tag{5.6}$$

Let $x \in \mathcal{L}_{Q_{\beta_k}}(\tilde{x}_{k,0})$, since $\|c(x_k)\| > \varepsilon_0$, Lemma 3.1 gives

$$f(x) \le Q_{\beta_k}(x) \le Q_{\beta_k}(\tilde{x}_{k,0}) \le Q_{\beta_k}(x_0) = f(x_0) + \frac{\beta_k}{2}\|c(x_0)\|^2 \le 2f(x_0) - f_{\text{low}}, \tag{5.7}$$

which shows

$$\mathcal{L}_{Q_{\beta_k}}(\tilde{x}_{k,0}) \subset \mathcal{L}_f(2f(x_0) - f_{\text{low}}).$$

Therefore, by Lemma 5.1, ∇Q_{β_k} is $(L_{f,1} + \beta_k L_{P,1})$-Lipschitz continuous on $\mathcal{L}_{Q_{\beta_k}}(\tilde{x}_{k,0})$. By A6, $\mathcal{A}_1$ generates x_{k+1} satisfying $\|\nabla Q_{\beta_k}(x_{k+1})\| \le \varepsilon_1 \le \tau(x_{k+1})$ and $Q_{\beta_k}(x_{k+1}) \le Q_{\beta_k}(\tilde{x}_{k,0})$ in at most

$$C_{\mathcal{A}_1}(L_{f,1} + \beta_k L_{P,1})(Q_{\beta_k}(\tilde{x}_{k,0}) - f_{\text{low}})\varepsilon_1^{-2} \tag{5.8}$$

evaluations of f, c, and their first-order derivatives. By (5.7), it holds $Q_{\beta_k}(\tilde{x}_{k,0}) - f_{\text{low}} \le 2(f(x_0) - f_{\text{low}})$. Since $\varepsilon_1 \le \tau(x_{k+1})$, the point x_{k+1} satisfies (2.3) and (2.4). □

5.2 First-order inner solver without the PL condition

Based on the inner evaluation complexity of Lemma 5.2, we have the following total evaluation complexity for Algorithm 1, without the PL condition on the constraint violation.

Theorem 5.3 (Total evaluation complexity with first-order inner solver without PL). *Under A1, A2, A3, A5, A7, suppose that at each iteration of QPM (Algorithm 1), the point x_{k+1} is computed using a monotone first-order method $\mathcal{A}_1$ initialized at*

$$\tilde{x}_{k,0} = \operatorname{argmin}_{x \in \{x_k, x_0\}} Q_{\beta_k}(x).$$

If the method $\mathcal{A}_1$ satisfies A6, then QPM generates an $(\varepsilon_0, \varepsilon_1)$-KKT point in at most

$$8\hat{T}(\varepsilon_0)\alpha C_{\mathcal{A}_1}(f(x_0) - f_{\text{low}})^2(L_{f,1} + L_{P,1})\varepsilon_0^{-2}\varepsilon_1^{-2} \tag{5.9}$$

evaluations of f, c, and their first-order derivatives, where $\hat{T}(\varepsilon_0)$ is defined in (3.3).

Proof. By Theorem 3.2, the number of outer iterations $T(\varepsilon_0)$ is upper bounded by $\hat{T}(\varepsilon_0)$. Since $\|c(x_k)\| > \varepsilon_0$ for $k = 1, \ldots, T(\varepsilon_0) - 1$, Lemma 5.2 gives that the kth iteration of Algorithm 1 requires at most

$$2C_{\mathcal{A}_1}(L_{f,1} + \beta_k L_{P,1})(f(x_0) - f_{\text{low}})\varepsilon_1^{-2}$$

evaluations of f, c, and their first-order derivatives. Lemma 3.1 gives $\beta_k < 4(f(x_0) - f_{\text{low}})\varepsilon_0^{-2}$ for $k = 0, \ldots, T(\varepsilon_0) - 2$. Therefore, $1 \leq \beta_k \leq 4\alpha(f(x_0) - f_{\text{low}})\varepsilon_0^{-2}$ for $k = 0, \ldots T(\varepsilon_0) - 1$. It follows that the total number of first-order oracle calls is bounded by

$$\sum_{k=0}^{T(\varepsilon_0)-1} 2C_{\mathcal{A}_1}(L_{f,1} + \beta_k L_{P,1})(f(x_0) - f_{\text{low}})\varepsilon_1^{-2} \tag{5.10}$$

$$\leq \sum_{k=0}^{T(\varepsilon_0)-1} 2C_{\mathcal{A}_1}\beta_k(L_{f,1} + L_{P,1})(f(x_0) - f_{\text{low}})\varepsilon_1^{-2} \tag{5.11}$$

$$\leq \hat{T}(\varepsilon_0)8\alpha C_{\mathcal{A}_1}(L_{f,1} + L_{P,1})(f(x_0) - f_{\text{low}})^2\varepsilon_0^{-2}\varepsilon_1^{-2}. \tag{5.12}$$

□

From Theorem 5.3, in the absence of the PL condition on the constraint violation, QPM equipped with a first-order inner solver takes at most $\mathcal{O}\left(|\log(\beta_0^{-1}\varepsilon_0^{-2})|\varepsilon_0^{-2}\varepsilon_1^{-2}\right)$ calls to the first-order oracle to find an $(\varepsilon_0, \varepsilon_1)$-KKT point. In particular, when $\beta_0 = O(\varepsilon_0^{-2})$, the complexity bound reduces to $\mathcal{O}(\varepsilon_0^{-2}\varepsilon_1^{-2})$.

5.3 First-order inner solver under the PL condition

The total evaluation complexity improves by a factor ε_0^{-1} under the PL condition on the constraint violation (A4).

Theorem 5.4 (Total evaluation complexity with first-order inner solver under PL). *Under A1, A2, A3, A4, A5, A7, suppose that at each iteration of QPM (Algorithm 1), the point x_{k+1} is computed using a monotone first-order method $\mathcal{A}_1$ initialized at*

$$\tilde{x}_{k,0} = \operatorname{argmin}_{x \in \{x_k, x_0\}} Q_{\beta_k}(x).$$

If the method $\mathcal{A}_1$ satisfies A6, then QPM generates an $(\varepsilon_0, \varepsilon_1)$-KKT point in at most

$$2\alpha\tilde{T}(\varepsilon_0)C_{\mathcal{A}_1}(L_{f,1} + L_{P,1})(f(x_0) - f_{\text{low}}) \max\left\{\frac{L_{f,0} + \varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0) - f_{\text{low}})}{R}\right\}\varepsilon_0^{-1}\varepsilon_1^{-2} \tag{5.13}$$

evaluations of f, c, and their first-order derivatives, where $L_{f,1}$ and $L_{P,1}$ are defined in Lemma 5.1, $\sigma_{\min}$ and R are defined in A4, $\tilde{T}(\varepsilon_0)$ is defined in (4.30) and $L_{f,0}$ is defined in Lemma 4.3.

Proof. By Theorem 4.4, the number of outer iterations $T(\varepsilon_0)$ is upper bounded by $\tilde{T}(\varepsilon_0)$. Since $\|c(x_k)\| > \varepsilon_0$ for $k = 1, \ldots, T(\varepsilon_0) - 1$, by Lemma 5.2, iteration k of Algorithm 1 requires at most

$$2C_{\mathcal{A}_1}(L_{f,1} + \beta_k L_{P,1})(f(x_0) - f_{\text{low}})\varepsilon_1^{-2}$$

evaluations of f, c, and their first-order derivatives. Lemma 4.3 gives

$$1 \leq \beta_k < \alpha \max\left\{\frac{L_{f,0} + \varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0) - f_{\text{low}})}{R}\right\}\varepsilon_0^{-1} \quad \text{for } k = 0, 1, \ldots, T(\varepsilon_0) - 1. \tag{5.14}$$

Therefore, the total number of first-order oracle calls is bounded by

$$\sum_{k=0}^{T(\varepsilon_0)-1} 2C_{\mathcal{A}_1}(L_{f,1} + \beta_k L_{P,1})(f(x_0) - f_{\text{low}})\varepsilon_1^{-2} \tag{5.15}$$

$$\leq \sum_{k=0}^{T(\varepsilon_0)-1} 2C_{\mathcal{A}_1}(L_{f,1} + L_{P,1})\beta_k(f(x_0) - f_{\text{low}})\varepsilon_1^{-2} \tag{5.16}$$

$$\leq T(\varepsilon_0)2C_{\mathcal{A}_1}(L_{f,1} + L_{P,1})(f(x_0) - f_{\text{low}})\alpha \max\left\{\frac{L_{f,0} + \varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0) - f_{\text{low}})}{R}\right\}\varepsilon_0^{-1}\varepsilon_1^{-2} \tag{5.17}$$

$$\leq 2\alpha\tilde{T}(\varepsilon_0)C_{\mathcal{A}_1}(L_{f,1} + L_{P,1})(f(x_0) - f_{\text{low}}) \max\left\{\frac{L_{f,0} + \varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0) - f_{\text{low}})}{R}\right\}\varepsilon_0^{-1}\varepsilon_1^{-2}. \qquad \square$$

From Theorem 5.4, under the PL condition on the constraint violation (A4), QPM equipped with a first-order inner solver takes at most $\mathcal{O}\left(\left|\log(\beta_0^{-1}\varepsilon_0^{-1})\right|\varepsilon_0^{-1}\varepsilon_1^{-2}\right)$ calls to the first-order oracle to find an $(\varepsilon_0, \varepsilon_1)$-KKT point. In particular, when $\beta_0 = \mathcal{O}(\varepsilon_0^{-1})$, the complexity bound reduces to $\mathcal{O}(\varepsilon_0^{-1}\varepsilon_1^{-2})$.

6 Total evaluation complexity with second-order inner solver

We consider the total evaluation complexity of the QPM when a second-order method minimizes the subproblem. We detail the results with and without the PL condition on the constraint violation.

6.1 Lemmas on second-order inner solvers

We consider the following assumption on the problem and the (inner) evaluation complexity of the second-order solver.

A8. *The functions f and c are three times continuously differentiable.*

A9. *Given a three times continuously differentiable function $F\colon \mathbb{R}^n \to \mathbb{R}$, consider that the method $\mathcal{A}_2$ is applied to minimize F starting from $\tilde{x} \in \mathbb{R}^n$, where*

$$\mathcal{L}_F(\tilde{x}) = \{z \in \mathbb{R}^n | F(z) \leq F(\tilde{x})\}$$

is bounded. There exists an algorithmic constant of $\mathcal{A}_2$, called $\Delta_{\max}$, such that all trial points and iterates visited by $\mathcal{A}_2$ belong to

$$\Omega_F = \{x + d \in \mathbb{R}^n | x \in \mathcal{L}_F(\tilde{x}) \text{ and } \|d\| \leq \Delta_{\max}\},$$

and the method $\mathcal{A}_2$ takes at most

$$C_{\mathcal{A}_2}(F(\tilde{x}) - F_{\text{low}})L^{1/2}\varepsilon^{-3/2}$$

evaluations of F, ∇F, and $\nabla^2 F$ to produce a point $x \in \mathbb{R}^n$ with $\|\nabla F(x)\| \leq \varepsilon$, where $\nabla^2 F$ is L-Lipschitz continuous on Ω_F.

The development of second-order algorithms for unconstrained minimization with optimal worst-case complexity is a very active field of research. For example, the trust-region method from (Hamad and Hinder, 2025), once augmented with an upper bound $\Delta_{\max}$ on the trust-region radius, satisfies A9. Under A5 and A9, we define the bounded set

$$\Omega_* := \operatorname{conv}\{x + d \in \mathbb{R}^n : x \in \mathcal{L}_f(2f(x_0) - f_{\text{low}}) \text{ and } \|d\| \le \Delta_{\max}\}. \tag{6.1}$$

Additionally, under A8, there exists finite constants

$$L_{f,2} := \sup_{x\in\Omega_*} \|D^3 f(x)\| \qquad \text{and} \qquad L_{P,2} := \sup_{x\in\Omega_*} \|D^3 P(x)\|. \tag{6.2}$$

In the following lemma, we show that, for all $\beta \ge 0$, the Hessian $\nabla^2 Q_\beta$ is Lipschitz continuous on Ω_*. We emphasize that we do not assume Lipschitz continuity of some derivative of f and c, merely that f has a bounded sublevel set (A5).

Lemma 6.1. *Under A5 and A8, for any $\beta > 0$, the Hessian $\nabla^2 Q_\beta$ is $(L_{f,2} + \beta L_{P,2})$-Lipschitz continuous on the set Ω_* (6.1).*

Proof. The set Ω_* defined in (6.1) is compact by A5 and the constants $L_{f,2}$ and $L_{P,2}$ in (6.2) are well defined and finite. We find that, for all $x \in \Omega_*$ and $\beta > 0$,

$$\begin{aligned} \left\|D^3 Q_\beta(x)\right\| &\le \left\|D^3 f(x) + \beta D^3 P(x)\right\| &(6.3)\\ &\le \left\|D^3 f(x)\right\| + \beta \left\|D^3 P(x)\right\| &(6.4)\\ &\le L_{f,2} + \beta L_{P,2}. &(6.5)\end{aligned}$$

Thus, $\left\|D^3 Q_\beta(x)\right\| \le L_{f,2} + \beta L_{P,2}$ for all $x \in \Omega_*$, and therefore $\nabla^2 Q_\beta$ is $(L_{f,2} + \beta L_{P,2})$-Lipschitz continuous on Ω_*. □

This allows to derive total evaluation complexity bounds for the QPM. We begin with an inner evaluation complexity for the second-order subproblem solver.

Lemma 6.2 (Inner evaluation complexity of $\mathcal{A}_2$)**.** *Under the conditions of Lemma 6.1, consider iteration k of Algorithm 1 where a monotone second-order method $\mathcal{A}_2$ minimizes Q_{β_k} with starting point*

$$\tilde{x}_{k,0} = \operatorname{argmin}_{x\in\{x_k,x_0\}} Q_{\beta_k}(x).$$

If $\mathcal{A}_2$ satisfies A9, the method $\mathcal{A}_2$ generates x_{k+1} satisfying the subproblem conditions (2.3) and (2.4) in at most

$$2C_{\mathcal{A}_2}(f(x_0) - f_{\text{low}})(L_{f,2} + \beta_k L_{P,2})^{\frac{1}{2}} \varepsilon_1^{-\frac{3}{2}} \tag{6.6}$$

evaluations of f, c, and their derivatives up to second order, where $L_{f,2}$ and $L_{P,2}$ are defined in (6.2).

Proof. By A2, we have $Q_{\beta_k}(x) \ge f_{\text{low}}$ for all $x \in \mathbb{R}^n$. We also have that $\mathcal{L}_{Q_{\beta_k}}(\tilde{x}_{k,0}) \subset \mathcal{L}_f(2f(x_0) - f_{\text{low}})$ (5.7). Therefore, all iterates and trial points of $\mathcal{A}_2$ applied to Q_{β_k} starting from $\tilde{x}_{k,0}$ remain in the set Ω_*. Furthermore, the Hessian $\nabla^2 Q_{\beta_k}$ is $(L_{f,2} + \beta_k L_{P,2})$-Lipschitz continuous on Ω_* by Lemma 6.1.

Therefore, $\mathcal{A}_2$ generates x_{k+1} satisfying $\|\nabla Q_{\beta_k}(x_{k+1})\| \le \varepsilon_1 \le \tau(x_{k+1})$ and $Q_{\beta_k}(x_{k+1}) \le Q_{\beta_k}(\tilde{x}_{k,0})$ in at most

$$C_{\mathcal{A}_2}(Q_{\beta_k}(\tilde{x}_{k,0}) - f_{\text{low}})(L_{f,2} + \beta_k L_{P,2})^{\frac{1}{2}} \varepsilon_1^{-\frac{3}{2}} \tag{6.7}$$

evaluations of f, c, and their derivatives up to second order. The conclusion follows from $Q_{\beta_k}(\tilde{x}_{k,0}) \le 2f(x_0) - f_{\text{low}}$ (5.7). □

Considering this result, we derive total evaluation complexity bounds for Algorithm 1.

6.2 Second-order inner solver without the PL condition

We give a total evaluation complexity bound for the QPM with a second-order inner solver and without the PL condition on the constraint violation.

Theorem 6.3 (Total evaluation complexity with second-order inner solver without PL)**.** *Under A1, A2, A3, A5, A8, suppose that at each iteration of QPM (Algorithm 1), the point x_{k+1} is computed using a monotone second-order method $\mathcal{A}_2$ initialized at*

$$\tilde{x}_{k,0} = \text{argmin}_{x\in\{x_k,x_0\}} Q_{\beta_k}(x).$$

If the method $\mathcal{A}_2$ satisfies A9, then QPM generates an $(\varepsilon_0, \varepsilon_1)$-KKT point in at most

$$4\hat{T}(\varepsilon_0)C_{\mathcal{A}_2}\sqrt{\alpha}(f(x_0) - f_{\text{low}})^{3/2}(L_{f,2} + L_{P,2})^{\frac{1}{2}}\varepsilon_0^{-1}\varepsilon_1^{-3/2} \tag{6.8}$$

evaluations of f, c, and their derivatives up to second order, where $L_{f,2}$ and $L_{P,2}$ are defined in (6.2), and $\hat{T}(\varepsilon_0)$ is defined in (3.3).

Proof. By Theorem 3.2, the number of outer iterations $T(\varepsilon_0)$ is upper bounded by $\hat{T}(\varepsilon_0)$. Since $\|c(x_k)\| > \varepsilon_0$ for $k = 1, \ldots, T(\varepsilon_0) - 1$, Lemma 6.2 gives that iteration k of Algorithm 1 requires at most

$$2C_{\mathcal{A}_2}(f(x_0) - f_{\text{low}})(L_{f,2} + \beta_k L_{P,2})^{\frac{1}{2}}\varepsilon_1^{-\frac{3}{2}} \tag{6.9}$$

evaluations of f, c, and their derivatives up to second order. Lemma 3.1 gives $1 \le \beta_k \le 4\alpha(f(x_0) - f_{\text{low}})\varepsilon_0^{-2}$ for $k = 0, \ldots T(\varepsilon_0) - 1$. Therefore, the total number of second-order oracle calls is bounded by

$$\begin{aligned}
&\sum_{k=0}^{T(\varepsilon_0)-1} 2C_{\mathcal{A}_2}(f(x_0) - f_{\text{low}})(L_{f,2} + \beta_k L_{P,2})^{\frac{1}{2}}\varepsilon_1^{-\frac{3}{2}} && (6.10)\\
&\le \sum_{k=0}^{T(\varepsilon_0)-1} 2C_{\mathcal{A}_2}(f(x_0) - f_{\text{low}})(L_{f,2} + L_{P,2})^{\frac{1}{2}}\beta_k^{\frac{1}{2}}\varepsilon_1^{-\frac{3}{2}} && (6.11)\\
&\le 2T(\varepsilon_0)C_{\mathcal{A}_2}(f(x_0) - f_{\text{low}})(L_{f,2} + L_{P,2})^{\frac{1}{2}}(4\alpha(f(x_0) - f_{\text{low}})\varepsilon_0^{-2})^{\frac{1}{2}}\varepsilon_1^{-\frac{3}{2}} && (6.12)\\
&\le 4\hat{T}(\varepsilon_0)C_{\mathcal{A}_2}\sqrt{\alpha}(f(x_0) - f_{\text{low}})^{3/2}(L_{f,2} + L_{P,2})^{\frac{1}{2}}\varepsilon_0^{-1}\varepsilon_1^{-\frac{3}{2}}. && (6.13)
\end{aligned}$$

□

From Theorem 6.3, in the absence of the PL condition on the constraint violation, QPM equipped with a second-order inner solver takes at most $\mathcal{O}\left(|\log(\beta_0^{-1}\varepsilon_0^{-2})|\varepsilon_0^{-1}\varepsilon_1^{-3/2}\right)$ calls to the first-order oracle to find an $(\varepsilon_0, \varepsilon_1)$-KKT point. In particular, when $\beta_0 = \mathcal{O}\left(\varepsilon_0^{-2}\right)$, the complexity bound reduces to $\mathcal{O}\left(\varepsilon_0^{-1}\varepsilon_1^{-3/2}\right)$.

6.3 Second-order inner solver under the PL condition

We now show a total evaluation complexity bound under the PL condition on the constraint violation with a second-order solver in the subproblems.

Theorem 6.4 (Total evaluation complexity with second-order inner solver under PL)**.** *Under A1, A2, A3, A4, A5, A8, suppose that at each iteration of QPM (Algorithm 1), the point x_{k+1} is computed using a monotone second-order method $\mathcal{A}_2$ initialized at*

$$\tilde{x}_{k,0} = \operatorname{argmin}_{x \in \{x_k, x_0\}} Q_{\beta_k}(x).$$

If the method $\mathcal{A}_2$ satisfies 9, then QPM generates an $(\varepsilon_0, \varepsilon_1)$-KKT in at most

$$2\tilde{T}(\varepsilon_0) C_{\mathcal{A}_2} (f(x_0) - f_{\text{low}})(L_{f,2} + L_{P,2})^{\frac{1}{2}} \sqrt{\alpha} \max\left\{ \frac{L_{f,0} + \varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0) - f_{\text{low}})}{R} \right\}^{\frac{1}{2}} \varepsilon_0^{-\frac{1}{2}} \varepsilon_1^{-\frac{3}{2}} \quad (6.14)$$

evaluations of f, c, and their derivatives up to second order, where $\tilde{T}(\varepsilon_0)$ is defined in (4.30), $L_{f,2}$ and $L_{P,2}$ are defined in (6.2), $\sigma_{\min}$ and R are defined in A4, and $L_{f,0}$ is defined in Lemma 4.3.

Proof. By Theorem 4.4, the number of outer iterations $T(\varepsilon_0)$ is upper bounded by $\tilde{T}(\varepsilon_0)$. Lemma 4.3 gives

$$1 \le \beta_k < \alpha \max\left\{ \frac{L_{f,0} + \varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0) - f_{\text{low}})}{R} \right\} \varepsilon_0^{-1} \quad \text{for } k = 0, 1, \ldots, T(\varepsilon_0) - 1. \quad (6.15)$$

Lemma 6.2 gives that the iteration k of Algorithm 1 requires at most

$$2C_{\mathcal{A}_2}(f(x_0) - f_{\text{low}})(L_{f,2} + \beta_k L_{P,2})^{\frac{1}{2}} \varepsilon_1^{-\frac{3}{2}} \quad (6.16)$$

evaluations of f, c, and their derivatives up to second order. Therefore, the total number of second-order oracle calls is bounded by

$$\begin{aligned}
&\sum_{k=0}^{T(\varepsilon_0)-1} 2C_{\mathcal{A}_2}(f(x_0) - f_{\text{low}})(L_{f,2} + \beta_k L_{P,2})^{\frac{1}{2}} \varepsilon_1^{-\frac{3}{2}} \\
&\le \sum_{k=0}^{T(\varepsilon_0)-1} 2C_{\mathcal{A}_2}(f(x_0) - f_{\text{low}})(L_{f,2} + L_{P,2})^{\frac{1}{2}} \beta_k^{\frac{1}{2}} \varepsilon_1^{-\frac{3}{2}} \\
&\le 2\tilde{T}(\varepsilon_0) C_{\mathcal{A}_2}(f(x_0) - f_{\text{low}})(L_{f,2} + L_{P,2})^{\frac{1}{2}} \sqrt{\alpha} \max\left\{ \frac{L_{f,0} + \varepsilon_1}{\sigma_{\min}}, \frac{4(f(x_0) - f_{\text{low}})}{R} \right\}^{\frac{1}{2}} \varepsilon_0^{-\frac{1}{2}} \varepsilon_1^{-\frac{3}{2}}. \square
\end{aligned}$$

From Theorem 6.4, under the PL condition on the constraint violation (A4), QPM equipped with a second-order inner solver takes at most $\mathcal{O}\left(\left|\log(\beta_0^{-1}\varepsilon_0^{-1})\right| \varepsilon_0^{-1/2} \varepsilon_1^{-3/2}\right)$ calls to the second-order oracle to find an $(\varepsilon_0, \varepsilon_1)$-KKT point. In particular, when $\beta_0 = \mathcal{O}(\varepsilon_0^{-1})$, the complexity bound reduces to $\mathcal{O}\left(\varepsilon_0^{-1/2} \varepsilon_1^{-3/2}\right)$.

7 Illustrative numerical results

In this section, we illustrate our theoretical findings. We demonstrate numerically the gain in performance induced by the feasibility-aware tolerance in the subproblems, and we also compared the performance of first- and second-order solvers in the subproblems.

Our experiments use a Julia implementation of Algorithm 1. The code to reproduce the results is available at `www.github.com/flgoyens/QPM`. Our implementation includes first- and second-order methods for the subproblems. The first-order method, called QPM-GD, uses a gradient descent method with an Armijo linesearch for the inner minimizations. The second-order method, called QPM-TR, uses a trust region method with exact Hessian for the inner minimizations. The trust-region subproblem is solved via the truncated conjugate gradient method (Conn et al., 2000). In our experiments, we use the values $\alpha = 1.2$ and $\beta_0 = 1$ unless specified otherwise.

In the experiments below, we compare the performance of two choices for the subproblem tolerance: the non-adaptive rule

$$\|\nabla Q_{\beta_k}(x_{k+1})\| \leq \varepsilon_1, \tag{7.1}$$

against the adaptive tolerance

$$\|\nabla Q_{\beta_k}(x_{k+1})\| \leq \tau_k := \max\left(\varepsilon_1, \frac{\varepsilon_1}{\varepsilon_0}\|c(x_{k+1})\|\right), \tag{7.2}$$

where τ_k is used in our plots to denote the adaptive tolerance at iteration k.

Our test problem is the minimization of the extended Rosenbrock function over the unit sphere: for n even,

$$\begin{aligned} \underset{x\in\mathbb{R}^n}{\text{minimize}} \quad & \sum_{i=1}^{n/2}\left[100(x_{2i} - x_{2i-1}^2)^2 + (1 - x_{2i-1})^2\right] \\ \text{subject to} \quad & \|x\|^2 = 1. \end{aligned} \tag{7.3}$$

The problem is smooth A8, the cost function is lower bounded A2, coercive A5, and the constraint function satisfies A4 as a particular case of Example 2 with $p = 1$. The starting point is chosen as

$$x_0 = \sqrt{\frac{1}{n}\left(1 + \frac{\varepsilon_0}{\sqrt{2}}\right)}\,\mathbf{1}_n,$$

in order to ensure $\|c(x_0)\| = \varepsilon_0/\sqrt{2}$ (A3).

First-order inner minimization: QPM-GD

Figure 1 shows two variants of QPM-GD applied to Problem (7.3) in dimension $n = 1000$. The blue curve represents the adaptive tolerance condition in the subproblem ((7.2), labelled τ_k), and the orange curve represent the constant tolerance ((7.1), labelled ε_1). It appears clearly that the adaptive tolerance performs less inner iterations in each subproblem. To reach an accuracy of 10^{-3}, the adaptive QPM uses 1570 gradients steps and the non-adaptive QPM uses 3259 gradient steps. Table 2 shows that both methods terminate with the same accuracy when the tolerance is set to 10^{-6}, and reports the markedly smaller number of oracle calls for the adaptive version.

Method	$\|c(x_{\text{final}})\|$	$f(x_{\text{final}})$	Q_β eval	∇Q_β eval
QPM-GD (τ_k)	7.1×10^{-7}	515.76	8441	4583
QPM-GD (ε_1)	7.1×10^{-7}	515.76	12079	7771

Table 2: QPM-GD with τ_k vs ε_1 on Problem (7.3) with $n = 1000$, $\varepsilon_0 = \varepsilon_1 = 10^{-6}$.

Second-order inner minimization: QPM-TR

Figure 2 shows the performance of QPM-GD and QPM-TR on the same instance of Problem (7.3) with $n = 1000$, both with adaptive tolerance in the subproblems. It is apparent that QPM-TR takes much less time to solve the problem. Additionally, Table 3 shows that the second-order method terminates with a smaller function value than the first-order method.

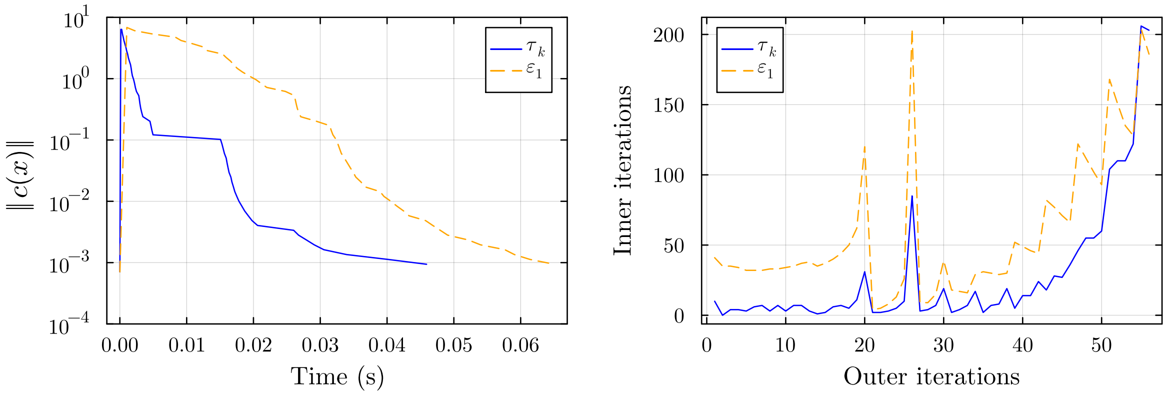

Figure 1: Comparing QPM-GD with adaptive and fixed subproblem tolerance on Problem (7.3) with $\beta_0 = 1$, $\alpha = 1.2$, $n = 10^3$, $\varepsilon_0 = \varepsilon_1 = 10^{-3}$. Total number of inner iterations: 1570 for τ_k and 3259 for ε_1.

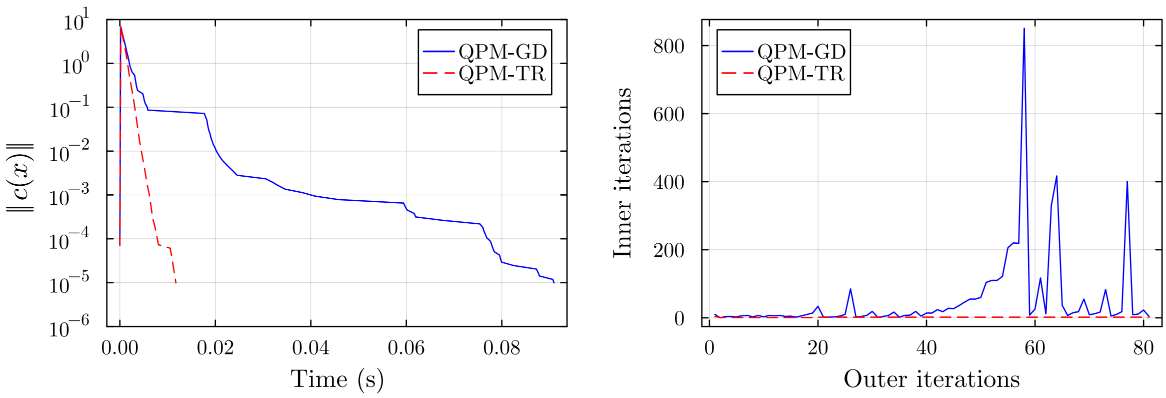

Figure 2: Comparing QPM-GD and QPM-TR, both with adaptive subproblem tolerance on Problem (7.3) with $\beta_0 = 1$, $\alpha = 1.2$, $n = 10^3$, $\varepsilon_0 = \varepsilon_1 = 10^{-5}$. Total number of inner iterations: 4316 for QPM-GD and 146 for QPM-TR.

Method	$\|c(x_{\text{final}})\|$	$f(x_{\text{final}})$	Q_β eval	∇Q_β eval	$\nabla^2 Q_\beta$ eval
QPM-GD (τ_k)	7.1×10^{-7}	515.76	8441	4583	0
QPM-TR (τ_k)	9.2×10^{-7}	456.31	548	265	262

Table 3: QPM-GD vs QPM-TR on Problem (7.3) with $n = 1000$, $\varepsilon_0 = \varepsilon_1 = 10^{-6}$ and $\tau(x) = \tau_k$.

Conclusions

In this work, we analyzed the worst-case oracle complexity of the Quadratic Penalty Method (QPM) for smooth, nonconvex, equality-constrained optimization problems, both with and without the PL condition on the constraint violation. In the absence of PL, we established complexity bounds of $\tilde{O}(\varepsilon_0^{-2}\varepsilon_1^{-2})$ and $\tilde{\mathcal{O}}(\varepsilon_0^{-1}\varepsilon_1^{-3/2})$ for obtaining $(\varepsilon_0, \varepsilon_1)$-KKT points when QPM is equipped with suitable first- and second-order inner solvers, respectively. Under the PL condition on the constraint violation, these bounds improve to $\mathcal{O}(\varepsilon_0^{-1}\varepsilon_1^{-2})$ and $\tilde{O}(\varepsilon_0^{-1/2}\varepsilon_1^{-3/2})$, reflecting the sharper dependence on feasibility accuracy afforded by the regularity assumption. In both regimes, we further showed that the logarithmic dependence on ε_0 vanishes when the initial penalty parameter is chosen proportional to a suitable power of ε_0^{-1}. Our analysis accommodates variants of QPM that use relaxed stopping criteria for the subproblems. Leveraging this flexibility, we proposed a feasibility-aware stopping rule that adaptively loosens the stationarity accuracy when far from feasibility. This criterion preserves all theoretical guarantees and can yield substantial practical speedups, as illustrated in our preliminary numerical experiments.

References

Andreani, R., Birgin, E. G., Martínez, J. M., and Schuverdt, M. L. (2008). On Augmented Lagrangian Methods with General Lower-Level Constraints. *SIAM Journal on Optimization*, 18(4):1286–1309.

Bertsekas, D. P. (1997). Nonlinear programming. *Journal of the Operational Research Society*, 48(3):334–334.

Birgin, E. G., Gardenghi, J. L., Martínez, J. M., Santos, S. A., and Toint, Ph. L. (2016). Evaluation Complexity for Nonlinear Constrained Optimization Using Unscaled KKT Conditions and High-Order Models. *SIAM Journal on Optimization*, 26(2):951–967.

Birgin, E. G. and Martínez, J. M. (2020). Complexity and performance of an Augmented Lagrangian algorithm. *Optimization Methods and Software*, 35(5):885–920.

Boumal, N., Absil, P.-A., and Cartis, C. (2019). Global rates of convergence for nonconvex optimization on manifolds. *IMA Journal of Numerical Analysis*, 39(1):1–33.

Bueno, L. F. and Martínez, J. M. (2020). On the Complexity of an Inexact Restoration Method for Constrained Optimization. *SIAM Journal on Optimization*, 30(1):80–101.

Cartis, C., Gould, N. I., and Toint, P. L. (2011). On the evaluation complexity of composite function minimization with applications to nonconvex nonlinear programming. *SIAM Journal on Optimization*, 21(4):1721–1739.

Cartis, C., Gould, N. I., and Toint, P. L. (2013). On the evaluation complexity of cubic regularization methods for potentially rank-deficient nonlinear least-squares problems and its relevance to constrained nonlinear optimization. *SIAM Journal on Optimization*, 23(3):1553–1574.

Cartis, C., Gould, N. I., and Toint, P. L. (2014). On the complexity of finding first-order critical points in constrained nonlinear optimization. *Mathematical Programming*, 144(1):93–106.

Cartis, C., Gould, N. I. M., and Toint, P. L. (2022). Evaluation complexity of algorithms for nonconvex optimization. *MOS-SIAM Series on Optimization.*

Conn, A. R., Gould, N. I., and Toint, P. L. (2000). *Trust Region Methods*, volume 1. SIAM.

Curtis, F. E., O'Neill, M. J., and Robinson, D. P. (2024). Worst-case complexity of an SQP method for nonlinear equality constrained stochastic optimization. *Mathematical Programming*, 205(1):431–483.

Curtis, F. E., Robinson, D. P., and Samadi, M. (2018). Complexity Analysis of a Trust Funnel Algorithm for Equality Constrained Optimization. *SIAM Journal on Optimization*, 28(2):1533–1563.

Eckstein, J. and Silva, P. J. S. (2013). A practical relative error criterion for augmented Lagrangians. *Mathematical Programming*, 141(1-2):319–348.

Facchinei, F., Kungurtsev, V., Lampariello, L., and Scutari, G. (2021). Ghost Penalties in Nonconvex Constrained Optimization: Diminishing Stepsizes and Iteration Complexity. *Mathematics of Operations Research*, 46(2):595–627.

Goyens, F., Eftekhari, A., and Boumal, N. (2022). Computing second-order points under equality constraints: Revisiting Fletcher's augmented Lagrangian. *arXiv preprint arXiv:2204.01448*, https://arxiv.org/abs/2204.01448.

Goyens, F., Eftekhari, A., and Boumal, N. (2024). Computing Second-Order Points Under Equality Constraints: Revisiting Fletcher's Augmented Lagrangian. *Journal of Optimization Theory and Applications*, 201(3):1198–1228.

Goyens, F. and Royer, C. W. (2024). Riemannian trust-region methods for strict saddle functions with complexity guarantees. *Mathematical Programming*.

Grapiglia, G. N. (2023). Worst-case evaluation complexity of a quadratic penalty method for nonconvex optimization. *Optimization Methods and Software*, 38(4):781–803.

Grapiglia, G. N. and Yuan, Y.-x. (2021). On the complexity of an augmented Lagrangian method for nonconvex optimization. *IMA Journal of Numerical Analysis*, 41(2):1508–1530.

Hamad, F. and Hinder, O. (2025). A simple and practical adaptive trust-region method.

He, C., Lu, Z., and Pong, T. K. (2023). A Newton-CG Based Augmented Lagrangian Method for Finding a Second-Order Stationary Point of Nonconvex Equality Constrained Optimization with Complexity Guarantees. *SIAM Journal on Optimization*, 33(3):1734–1766.

Kong, W., Melo, J. G., and Monteiro, R. D. C. (2019). Complexity of a Quadratic Penalty Accelerated Inexact Proximal Point Method for Solving Linearly Constrained Nonconvex Composite Programs. *SIAM Journal on Optimization*, 29(4):2566–2593.

Li, Z., Chen, P.-Y., Liu, S., Lu, S., and Xu, Y. (2021). Rate-improved inexact augmented Lagrangian method for constrained nonconvex optimization. In *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, pages 2170–2178. PMLR.

Li, Z. and Xu, Y. (2021). Augmented Lagrangian based first-order methods for convex-constrained programs with weakly-convex objective.

Lin, Q., Ma, R., and Xu, Y. (2020). Inexact Proximal-Point Penalty Methods for Constrained Non-Convex Optimization.

Martínez, J. M. (2017). On High-order Model Regularization for Constrained Optimization. *SIAM Journal on Optimization*, 27(4):2447–2458.

Melo, J. G., Monteiro, R. D. C., and Wang, H. (2020). Iteration-complexity of an inexact proximal accelerated augmented Lagrangian method for solving linearly constrained smooth nonconvex composite optimization problems.

Xie, Y. and Wright, S. J. (2021). Complexity of proximal augmented Lagrangian for nonconvex optimization with nonlinear equality constraints. *Journal of Scientific Computing*, 86(3):1–30.

Zangwill, W. I. (1967). Non-Linear Programming Via Penalty Functions. *Management Science*, 13(5):344–358.