Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 56-2 (2023) 1247–1252

Regularized Interior Point Methods for Constrained Optimization and Control

Alberto De Marchi

University of the Bundeswehr Munich, Department of Aerospace Engineering, Institute of Applied Mathematics and Scientific Computing, Neubiberg/Munich, Germany (e-mail: alberto.demarchi@unibw.de, ORCID: 0000-0002-3545-6898)

Abstract: Regularization and interior point approaches offer valuable perspectives to address constrained nonlinear optimization problems in view of control applications. This paper discusses the interactions between these techniques and proposes an algorithm that synergistically combines them. Building a sequence of closely related subproblems and approximately solving each of them, this approach inherently exploits warm-starting, early termination, and the possibility to adopt subsolvers tailored to specific problem structures. Moreover, by relaxing the equality constraints with a proximal penalty, the regularized subproblems are feasible and satisfy a strong constraint qualification by construction, allowing the safe use of efficient solvers. We show how regularization benefits the underlying linear algebra and a detailed convergence analysis indicates that limit points tend to minimize constraint violation and satisfy suitable optimality conditions. Finally, numerical results indicate that the combined approach compares favorably, in terms of robustness, against both interior point and augmented Lagrangian codes.

Keywords: Nonlinear programming, interior point methods, augmented Lagrangian, proximal algorithms, optimization-based control.

1. INTRODUCTION

Mathematical optimization plays an important role in model-based and data-driven control systems, forming the basis for advanced techniques such as optimal control, nonlinear model predictive control (MPC) and parameters estimation. Significant research effort on computationally efficient real-time optimization algorithms contributed to the success of MPC over the years and yet the demand for fast and reliable methods for a broad spectrum of applications is growing; see Sopasakis et al. (2020); Saraf and Bemporad (2022) and references therein. In order to tackle these challenges, it is desirable to have an algorithm that benefits from warm-starting, can cope with infeasibility, is robust to problem scaling, and exploits the problem structure. In order to reduce computations and increase robustness, a common approach is to relax the requirements on the solutions, in terms of optimality, constraint violation, or both (Diehl et al., 2009; Saraf and Bemporad, 2022). In this work, we propose to address such features by combining proximal regularization and interior point techniques, for developing a stabilized, efficient and robust numerical method. We advocate for this strategy by bringing together and combining a variety of ideas from the nonlinear programming literature.

Let us consider the constrained nonconvex problem

$$\underset{x}{\text{minimize}} \quad f(x) \qquad \text{subject to} \quad c(x) = 0, \qquad x \geq 0, \tag{P}$$

where functions $f\colon \mathbb{R}^n \to \mathbb{R}$ and $c\colon \mathbb{R}^n \to \mathbb{R}^m$ are (at least) continuously differentiable. Problems with general inequality constraints on x and $c(x)$ can be reformulated as (P) by introducing auxiliary variables. Nonlinear programming (NLP) problems such as (P) have been extensively studied and there exist several approaches for their numerical solution. Interior point (IP) (Vanderbei and Shanno, 1999; Wächter and Biegler, 2006), penalty and augmented Lagrangian (AL) (Conn et al., 1991; Andreani et al., 2008; Birgin and Martínez, 2014) and sequential programming (Fiacco and McCormick, 1968) schemes are predominant ideas and have been re-combined in many ways (Curtis, 2012; Birgin et al., 2016; Armand and Tran, 2019).

Starting from linear programming, IP methods had a significant impact on the field of mathematical optimization (Gondzio, 2012). By solving a sequence of barrier subproblems, they can efficiently handle inequality constraints and scale well with the problem size. The state-of-the-art solver Ipopt, described by Wächter and Biegler (2006), is an emblem of this remarkable success. However, relying on Newton-type schemes for approximately solving the subproblems, IP algorithms may suffer degeneracy and lack of constraint qualifications if suitable counter-mechanisms are not implemented. On the contrary, proximal techniques naturally cope with these scenarios thanks to their inherent regularizing action. Widely investigated in the convex setting (Rockafellar, 1976), their favorable properties have been exploited to design stabilized methods building on the proximal point algorithm (Friedlander and Orban, 2012; Liao-McPherson and Kolmanovsky, 2020; De Marchi, 2022). The analysis of their close connection with the AL framework (Rockafellar, 1974) led to the development of techniques applicable to more general problems (Ma et al., 2018; Potschka and Bock, 2021).

10.1016/j.ifacol.2023.10.1747

The combination of IP and proximal strategies has been successfully leveraged in the context of convex quadratic programming (Altman and Gondzio, 1999; Cipolla and Gondzio, 2022) and for linear (Dehghani et al., 2020) and nonlinear (Orban and Siqueira, 2020) least-squares problems. With this work we address general NLPs and devise a method for their numerical solution, which can be seen as an extension of a regularized Lagrange–Newton method to handle bound constraints via a barrier function (De Marchi, 2021), or as a proximally stabilized IP algorithm, generalizing the ideas put forward by Cipolla and Gondzio (2022).

Outline The paper is organized as follows. In Section 2 we provide and comment on some relevant optimality notions. The methodology is discussed in Section 3 detailing the proposed algorithm, whose convergence properties are investigated in Section 4. We report numerical results on benchmark problems in Section 5 and conclude the paper in Section 6.

Notation With $\mathbb{N}$, $\mathbb{R}$, and $\overline{\mathbb{R}} := \mathbb{R} \cup \{\infty\}$ we denote the natural, real and extended real numbers, respectively. We denote the set of real vectors of dimension $n \in \mathbb{N}$ as $\mathbb{R}^n$; a real matrix with $m \in \mathbb{N}$ rows and $n \in \mathbb{N}$ columns as $A \in \mathbb{R}^{m\times n}$ and its transpose as $A^\top \in \mathbb{R}^{n\times m}$. For a vector $a \in \mathbb{R}^n$, its i-th element is a_i and its squared Euclidean norm is $\|a\|^2 = a^\top a$. A vector or matrix with all zero elements is represented by 0. The gradient of a function $f\colon \mathbb{R}^n \to \mathbb{R}$ at a point $\bar{x} \in \mathbb{R}^n$ is denoted by $\nabla f(\bar{x}) \in \mathbb{R}^n$; the Jacobian of a vector function $c\colon \mathbb{R}^n \to \mathbb{R}^m$ by $\nabla c(\bar{x}) \in \mathbb{R}^{m\times n}$.

2. OPTIMALITY AND STATIONARITY

In this section we introduce some optimality concepts, following (Birgin and Martínez, 2014, Ch. 3) and (De Marchi and Themelis, 2022, Sec. 2).

Definition 1. (Feasibility). Relative to (P), we say a point $x^* \in \mathbb{R}^n$ is *feasible* if $x^* \geq 0$ and $c(x^*) = 0$; it is *strictly feasible* if additionally $x^* > 0$.

Definition 2. (Approximate KKT stationarity). Relative to (P), a point $x^* \in \mathbb{R}^n$ is *ε-KKT stationary* for some $\varepsilon \geq 0$ if there exist multipliers $y^* \in \mathbb{R}^m$ and $z^* \in \mathbb{R}^n$ such that

$$\|\nabla f(x^*) + \nabla c(x^*)^\top y^* + z^*\| \leq \varepsilon, \tag{1a}$$

$$\|c(x^*)\| \leq \varepsilon, \tag{1b}$$

$$x^* \geq -\varepsilon, \quad z^* \leq \varepsilon, \quad \min\{x^*, -z^*\} \leq \varepsilon. \tag{1c}$$

When $\varepsilon = 0$, the point x^* is said *KKT stationary*.

Notice that (1c) provides a condition modeling the approximate satisfaction of the (elementwise) complementarity condition $\min\{x, -z\} = 0$ within some tolerance $\varepsilon \geq 0$. The IP algorithm discussed in Section 3 satisfies a stronger version of these conditions, since the iterates it generates meet the constraints $x \geq 0$ and $z \leq 0$ by construction. Furthermore, we point out that the condition $\min\{x_i, -z_i\} \leq \varepsilon$ is analogous to $-x_i z_i \leq \varepsilon$, more typical for interior point methods, but does not depend on a specific barrier function, e.g., the logarithmic barrier in (Wächter and Biegler, 2006, Sec. 2.1).

We shall consider the limiting behavior of approximate KKT stationary points when the tolerance ε vanishes. In fact, having $x^k \to x^*$ with x^k ε_k-KKT stationary for (P) and $\varepsilon_k \searrow 0$ does not guarantee KKT stationarity of a limit point x^* of $\{x^k\}$. This issue raises the need for defining KKT stationarity in an asymptotic sense (Birgin and Martínez, 2014, Def. 3.1).

Definition 3. (Asymptotic KKT stationarity). Relative to (P), a feasible point $x^* \in \mathbb{R}^n$ is *AKKT stationary* if there exist sequences $\{x^k\}, \{z^k\} \subset \mathbb{R}^n$, and $\{y^k\} \subset \mathbb{R}^m$ such that $x^k \to x^*$ and

$$\nabla f(x^k) + \nabla c(x^k)^\top y^k + z^k \to 0, \tag{2a}$$

$$\min\{x^k, -z^k\} \to 0. \tag{2b}$$

Any local minimizer x^* for (P) is AKKT stationary, independently of constraint qualifications (Birgin and Martínez, 2014, Thm 3.1).

3. APPROACH AND ALGORITHM

The methodology presented in this section builds upon the AL framework, interpreted as a proximal point scheme in the nonconvex regime, and IP methods. The basic idea is to construct a sequence of proximally regularized subproblems and to approximately solve each of them as a single barrier subproblem, effectively merging the AL and IP outer loops. Reduced computational cost can be achieved with an effective warm-starting of the IP iterations and with the tight entanglement of barrier and proximal penalty strategies, by monitoring and updating the parameters' values alongside with the inner tolerance.

A classical approach is to consider a sequence of bound-constrained Lagrangian (BCL) subproblems (Conn et al., 1991; Birgin and Martínez, 2014)

$$\underset{x \geq 0}{\text{minimize}} \quad f(x) + \frac{1}{2\rho_k}\|c(x) + \rho_k \hat{y}^k\|^2 \tag{3}$$

where $\rho_k > 0$ and $\hat{y}^k \in \mathbb{R}^m$ are some given penalty parameter and dual estimate, respectively. The nonlinearly-constrained Lagrangian (NCL) scheme (Ma et al., 2018) considers equality-constrained subproblems by introducing an auxiliary variable $s \in \mathbb{R}^m$ and the constraint $c(x) = s$. Analogously, a proximal point perspective yields the equivalent reformulation

$$\underset{x,\,\lambda}{\text{minimize}} \quad f(x) + \frac{\rho_k}{2}\|\lambda\|^2 \tag{4}$$

$$\text{subject to} \quad c(x) + \rho_k(\hat{y}^k - \lambda) = 0, \qquad x \geq 0,$$

recovering the dual regularization term obtained, e.g., by Potschka and Bock (2021); De Marchi (2021, 2022). By construction, these regularized subproblems are always feasible and satisfy a strong constraint qualification, namely the LICQ, at all points.

The regularized subproblems (3)–(4) can be numerically solved via IP algorithms. Let us consider a barrier parameter $\mu_k > 0$ and barrier functions $b_i\colon \mathbb{R} \to \overline{\mathbb{R}}$, $i = 1, \ldots, n$, each with domain $\operatorname{dom} b_i = (0, \infty)$, and such that $b_i(t) \to \infty$ as $t \to 0^+$ and $b_i' \leq 0$. Exemplarily, the logarithmic function $x \mapsto -\ln(x)$ is one of such barrier functions. Other choices can be considered as well, e.g., to handle bilateral constraints (Bertolazzi et al., 2007). We collect these barrier functions to define $b\colon \mathbb{R}^n \to \overline{\mathbb{R}}$,

$b\colon x \mapsto \sum_{i=1}^{n} b_i(x_i)$, whose domain is $\operatorname{dom} b = (0,\infty)^n$. Thus, analogously to Armand and Tran (2019), a barrier counterpart for the BCL subproblem (3) reads

$$\underset{x}{\text{minimize}} \quad f(x) + \frac{1}{2\rho_k}\|c(x) + \rho_k \hat{y}^k\|^2 + \mu_k b(x), \tag{5}$$

whereas for the constrained subproblem (4) this leads to

$$\underset{x,\,\lambda}{\text{minimize}} \quad f(x) + \frac{\rho_k}{2}\|\lambda\|^2 + \mu_k b(x) \tag{6}$$
$$\text{subject to} \quad c(x) + \rho_k(\hat{y}^k - \lambda) = 0,$$

which is a regularized version of (Wächter and Biegler, 2006, Eq. 3) and reminiscent of (Birgin et al., 2016, Eq. 2). It should be stressed that, in stark contrast with classical AL and IP schemes, we intend to find an (approximate) solution to the regularized subproblem (4) by (approximately) solving only one barrier subproblem (6). Inspired by Curtis (2012); Cipolla and Gondzio (2022), our rationale is to drive ρ_k, μ_k and the inner tolerance ϵ_k concurrently toward zero, effectively knitting together proximal and barrier strategies.

It should be noted that a primal (Tikhonov-like) regularization term is not explicitly included in (3)–(6). In fact, the original objective f could be replaced by a (proximal) model of the form $x \mapsto f(x) + \frac{\sigma_k}{2}\|x - \hat{x}^k\|^2$, with some given primal regularization parameter $\sigma_k \geq 0$ and reference point $\hat{x}^k \in \mathbb{R}^n$. However, as this term can be interpreted as an inertia correction, we prefer the subsolver to account for its contribution; cf. (Wächter and Biegler, 2006, Sec. 3.1). In this way, the subsolver can surgically tune the primal correction term as needed, possibly improving the convergence speed, and surpassing the issue that suitable values for σ_k are unknown a priori.

Algorithm 3.1: Regularized interior point method for general nonlinear programs (P)

Data: $\epsilon_0, \rho_0, \mu_0 > 0$, $\kappa_\rho, \kappa_\mu, \kappa_\epsilon \in (0,1)$, $\theta_\rho, \theta_\mu \in [0,1)$, $Y \subset \mathbb{R}^m$ nonempty bounded, $\varepsilon > 0$

Result: ε-KKT stationary point x^* with y^*, z^*

1 **for** $k = 0, 1, 2, \ldots$ **do**
2   Select $\hat{y}^k \in Y$
3   Find an ϵ_k-KKT stationary point (x^k, λ^k) for (6), with multiplier y^k
4   Set $z^k \leftarrow \mu_k \nabla b(x^k)$
5   **if** (x^k, y^k, z^k) satisfies (1) **then**
6   **return** $(x^*, y^*, z^*) \leftarrow (x^k, y^k, z^k)$
7   Set $C^k \leftarrow \|c(x^k)\|$ and $V^k \leftarrow \|\min\{x^k, -z^k\}\|$
8   **if** $k = 0$ **or** $C^k \leq \max\{\varepsilon, \theta_\rho C^{k-1}\}$ **then**
9   set $\rho_{k+1} \leftarrow \rho_k$, **else** select $\rho_{k+1} \in (0, \kappa_\rho \rho_k]$
10   **if** $k = 0$ **or** $V^k \leq \max\{\varepsilon, \theta_\mu V^{k-1}\}$ **then**
11   set $\mu_{k+1} \leftarrow \mu_k$, **else** select $\mu_{k+1} \in (0, \kappa_\mu \mu_k]$
12   Set $\epsilon_{k+1} \leftarrow \max\{\varepsilon, \kappa_\epsilon \epsilon_k\}$

The overall procedure is detailed in Algorithm 3.1. At every outer iteration, indexed by k, Step 2 requires to compute an approximate stationary point, with the associated Lagrange multiplier, for the regularized barrier subproblem (6). As the dual estimate $\hat{y}^k$ is selected from some bounded set $Y \subset \mathbb{R}^m$ at Step 1, the AL scheme is *safeguarded* and has stronger global convergence properties (Birgin and Martínez, 2014, Ch. 4). The assignment of z^k at Step 3 follows from comparing and matching the stationarity conditions for (P) and (6). After checking termination, we monitor progress in constraint violation and complementarity, based on (1), and update parameters ρ_k and μ_k accordingly, as well as the inner tolerance ε_k. At Steps 8 and 10 we consider relaxed conditions for *satisfactory* feasibility and complementarity as it is preferable to have the sequences $\{\rho_k\}$, $\{\mu_k\}$, and $\{\epsilon_k\}$ bounded away from zero, in order to avoid unnecessary ill-conditioning and tight tolerances. Sufficient conditions to guarantee boundedness of the penalty parameter $\{\rho_k\}$ away from zero are given, e.g., by (Andreani et al., 2008, Sec. 5). Remarkably, as established by Lemma 5 in Section 4, there is no need for the barrier parameter μ_k to vanish in order to achieve ε-complementarity in the sense of (1c), for $\varepsilon > 0$.

We shall mention that considering equivalent yet different subproblem formulations may affect the practical performance of the subsolver. It is enlightening to pinpoint the effect of the dual regularization in (6) and to appreciate its interactions with the linear algebra routines used to solve the linear systems arising in Newton-type methods. Although (6) has more (possibly many more) variables than (5), a simple reordering yields matrices with the same structure (De Marchi, 2021; Potschka and Bock, 2021). Let us have a closer look. Defining the Lagrangian function $\mathcal{L}_k(x,y) := f(x) + \mu_k b(x) + \langle y, c(x)\rangle$, the stationarity condition for (5) reads $0 = \nabla_x \mathcal{L}_k(x, y_k(x))$, where $y_k(x) := \hat{y}^k + \rho_k^{-1} c(x)$, and the corresponding Newton system is

$$\left[H_k(x, y_k(x)) + \frac{1}{\rho_k}\nabla c(x)^\top \nabla c(x)\right]\delta x = -\nabla_x \mathcal{L}_k(x, y_k(x)),$$

where $H_k(x,y) \in \mathbb{R}^{n\times n}$ denotes the Hessian matrix $\nabla^2_{xx}\mathcal{L}_k(x,y)$ or a symmetric approximation thereof. A linear transformation yields the equivalent linear system

$$\begin{bmatrix} H_k(x, y_k(x)) & \nabla c(x)^\top \\ \nabla c(x) & -\rho_k I \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \end{bmatrix} = -\begin{bmatrix} \nabla_x \mathcal{L}_k(x, y_k(x)) \\ 0 \end{bmatrix}.$$

Analogous Newton systems for (6) read

$$\begin{bmatrix} H_k(x,y) & \cdot & \nabla c(x)^\top \\ \cdot & \rho_k I & -\rho_k I \\ \nabla c(x) & -\rho_k I & \cdot \end{bmatrix} \begin{bmatrix} \delta x \\ \delta \lambda \\ \delta y \end{bmatrix} = -\begin{bmatrix} \nabla_x \mathcal{L}_k(x,y) \\ \rho_k(\lambda - y) \\ c(x) + \rho_k(\hat{y}^k - \lambda) \end{bmatrix}$$

and formally solving for $\delta\lambda$ gives the condensed system

$$\begin{bmatrix} H_k(x,y) & \nabla c(x)^\top \\ \nabla c(x) & -\rho_k I \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \end{bmatrix} = -\begin{bmatrix} \nabla_x \mathcal{L}_k(x,y) \\ c(x) + \rho_k(\hat{y}^k - y) \end{bmatrix}. \tag{7}$$

The resemblances between these linear systems are apparent, as well as the differences. The AL relaxation in (5) introduces a dual regularization for both the linear algebra and nonlinear solver, whose *hidden* constraint $c(x) + \rho_k(\hat{y}^k - y) = 0$ holds pointwise due to the identity $y = y_k(x)$. We remark that, entering the (2,2)-block, the dual regularization prevents issues due to linear dependence. Furthermore, the primal regularization is left to the inertia correction strategy of the subsolver, affecting the (1,1)-block as in (Wächter and Biegler, 2006, Sec. 3.1). If the approximation $H_k(x,y)$ is positive definite, e.g., by adopting suitable quasi-Newton techniques, the matrix in (7) is symmetric quasi-definite and can be efficiently factorized with tailored linear algebra routines (Vanderbei, 1995).

4. CONVERGENCE ANALYSIS

In this section we analyze the asymptotic properties of the iterates generated by Algorithm 3.1 under the following blanket assumptions:

(A1) Functions $f\colon \mathbb{R}^n \to \mathbb{R}$ and $c\colon \mathbb{R}^n \to \mathbb{R}^m$ in (P) are continuously differentiable.
(A2) Subproblems (6) are well-posed for all parameters' values, namely for any $\mu_k \leq \mu_0$, $\rho_k \leq \rho_0$, and $\hat{y}^k \in Y$.

First, we characterize the iterates in terms of stationarity.

Lemma 4. Consider a sequence $\{x^k, y^k, z^k\}$ generated by Algorithm 3.1. Then, for all $k \in \mathbb{N}$, it is $x^k > 0$, $z^k \leq 0$, and the following conditions hold:

$$\|\nabla f(x^k) + \nabla c(x^k)^\top y^k + z^k\| \leq \epsilon_k, \tag{8a}$$
$$\|c(x^k) + \rho_k(\hat{y}^k - y^k)\| \leq 2\epsilon_k. \tag{8b}$$

Proof. Positivity of x^k follows from the barrier function b having domain $\operatorname{dom} b = (0,\infty)^n$, whereas nonpositivity of z^k is a consequence of $b_i' \leq 0$ for all i and $\mu_k > 0$. Based on Definition 2 and Step 3 of Algorithm 3.1, the ϵ_k-KKT stationarity of (x^k, λ^k) for (6), with multiplier y^k, yields (8a) along with

$$\rho_k\|\lambda^k - y^k\| \leq \epsilon_k, \tag{9a}$$
$$\|c(x^k) + \rho_k(\hat{y}^k - \lambda^k)\| \leq \epsilon_k. \tag{9b}$$

By the triangle inequality, (9a)–(9b) imply (8b).

Patterning (De Marchi and Themelis, 2022, Thm 4.2(ii)), we establish asymptotic complementarity.

Lemma 5. Consider a sequence $\{x^k, y^k, z^k\}$ of iterates generated by Algorithm 3.1 with $\varepsilon = 0$. Then, it holds $\lim_{k\to\infty} \min\{x^k, -z^k\} = 0$.

Proof. The algorithm can terminate in finite time only if the returned triplet (x^*, y^*, z^*) satisfies $\min\{x^*, -z^*\} = 0$. Excluding this ideal situation, we may assume that it runs indefinitely and that consequently $\mu_k \searrow 0$, owing to Steps 10 and 11 and recalling that $x^k > 0$ and $z^k \leq 0$ for all $k \in \mathbb{N}$ by Lemma 4. Consider now an arbitrary index $i \in \{1, \ldots, n\}$ and the two possible cases. If $x_i^k \to 0$, then the statement readily follows from $z_i^k \leq 0$. If instead a subsequence $\{x_i^k\}_K$ remains bounded away from zero, then $\{b_i'(x_i^k)\}_K$ is bounded and therefore $z_i^k = \mu_k b_i'(x_i^k) \to 0$ as $k \to_K \infty$, proving the statement since $x_i^k > 0$. The claim then follows from the arbitrarity of the index i and the subsequence.

Like all penalty-type methods in the nonconvex setting, Algorithm 3.1 may generate limit points that are infeasible for (P). Patterning standard arguments, the following result gives sufficient conditions for the feasibility of limit points; cf. (Birgin and Martínez, 2014, Ex. 4.12).

Proposition 6. Consider a sequence $\{x^k, y^k, z^k\}$ of iterates generated by Algorithm 3.1. Then, each limit point x^* of $\{x^k\}$ is feasible for (P) if one of the following conditions holds:

(i) the sequence $\{\rho_k\}$ is bounded away from zero, or
(ii) there exists some $B \in \mathbb{R}$ such that for all $k \in \mathbb{N}$

$$f(x^k) + \frac{1}{2\rho_k}\|c(x^k) + \rho_k \hat{y}^k\|^2 \leq B.$$

These conditions are generally difficult to check a priori. Nevertheless, in the situation where each iterate x^k is actually a (possibly inexact) global minimizer of (6), then limit points generated by Algorithm 3.1 have minimum constraint violation and tend to minimize the objective function subject to minimal infeasibility (Birgin and Martínez, 2014, Thm 5.1, Thm 5.3). In particular, limit points are indeed feasible if (P) admits feasible points. However, these properties cannot be expected by solving the subproblems only up to stationarity. Nonetheless, even in the case where a limit point is not necessarily feasible, the next result shows that it is at least a stationary point for a feasibility problem associated to (P).

Proposition 7. Consider a sequence $\{x^k, y^k, z^k\}$ generated by Algorithm 3.1 with $\varepsilon = 0$. Then each limit point x^* of $\{x^k\}$ is KKT stationary for the feasibility problem

$$\operatorname*{minimize}_{x \geq 0} \quad \frac{1}{2}\|c(x)\|^2.$$

Proof. We may consider two cases, depending on the sequence $\{\rho_k\}$. If $\{\rho_k\}$ remains bounded away from zero, then Steps 8 and 9 of Algorithm 3.1 imply that $\|c(x^k)\| \to 0$ for $k \to \infty$. Continuity of c and properties of norms yield $c(x^*) = 0$. Furthermore, by construction, we have $x^k > 0$ for all $k \in \mathbb{N}$, hence $x^* \geq 0$. Altogether, this shows that x^* is feasible for (P), namely a global minimizer for the feasibility problem and, therefore, a KKT stationary point thereof. Assume now that $\rho_k \to 0$. Define $\delta^k \in \mathbb{R}^n$ and $\eta^k \in \mathbb{R}^m$ as

$$\delta^k := \nabla f(x^k) + \nabla c(x^k)^\top y^k + z^k$$
$$\eta^k := c(x^k) + \rho_k(\hat{y}^k - y^k)$$

for all $k \in \mathbb{N}$. In view of Lemma 4, we have that $\|\delta^k\| \leq \epsilon_k$ and $\|\eta^k\| \leq 2\epsilon_k$ hold for all $k \in \mathbb{N}$. Multiplying δ^k by ρ_k, substituting y^k and rearranging, we obtain

$$\rho_k\delta^k = \rho_k\nabla f(x^k) + \nabla c(x^k)^\top\left[\rho_k\hat{y}^k + c(x^k) - \eta^k\right] + \rho_k z^k.$$

Now, let x^* be a limit point of $\{x^k\}$ and $\{x^k\}_K$ a subsequence such that $x^k \to_K x^*$. Then the sequence $\{\nabla f(x^k)\}_K$ is bounded, and so is $\{\hat{y}^k\}_K \subset Y$ by construction. Recalling from Lemma 4 that $x^k > 0$ and $z^k \leq 0$, and observing that $0 \leq \|\delta^k\|, \|\eta^k\| \leq 2\epsilon_k \to 0$, we shall now take the limit of $\rho_k\delta^k$ for $k \to_K \infty$, resulting in

$$0 = \nabla c(x^*)^\top c(x^*) + \tilde{z}^*$$

for some $\tilde{z}^* \leq 0$. As a limit point of $\{\rho_k z^k\}$, $\tilde{z}^*$ together with x^* satisfy $\min\{x^*, -\tilde{z}^*\} = 0$ by Lemma 5. Since we also have $x^* \geq 0$, it follows that x^* is KKT stationary for the feasibility problem, according to Definition 2.

Notice that requiring the sequence of dual estimates $\{\hat{y}^k\}$ to remain bounded is not strictly necessary, provided that $\rho_k\hat{y}^k \to 0$ as $\rho_k \searrow 0$ (Robinson, 2007, Ch. 4).

Finally, we qualify the output of Algorithm 3.1 in the case of feasible limit points. In particular, it is shown that any feasible limit point is AKKT stationary for (P) in the sense of Definition 3. Under some additional boundedness conditions, feasible limit points are KKT stationary, according to Definition 2.

Theorem 8. Let $\{x^k, y^k, z^k\}$ be a sequence of iterates generated by Algorithm 3.1 with $\varepsilon = 0$. Let x^* be a feasible limit point of $\{x^k\}$ and $\{x^k\}_K$ a subsequence such that $x^k \to_K x^*$. Then,

(i) x^* is an AKKT stationary point for (P).
(ii) If $\{y^k, z^k\}_K$ remain bounded, then x^* is KKT stationary for (P).

Proof. (i) Together with the fact that $\epsilon_k \to 0$, Lemma 4 ensures that the sequence $\{x^k\}_K$ satisfies condition (2a), whereas Lemma 5 implies (2b). Feasibility of x^* completes the proof.

(ii) By boundedness, the subsequences $\{y^k\}_K$ and $\{z^k\}_K$ admit some limit points y^* and z^*, respectively. Thus, from the previous assertion and with continuity arguments on f and c, it follows that x^* is KKT stationary for (P), not only asymptotically.

Provided that the iterates admit a feasible limit point, finite termination of Algorithm 3.1 with an ε-KKT stationary point can be established as a direct consequence of Theorem 8.

5. NUMERICAL RESULTS

In this section we test an instance of the proposed regularized interior point approach, denoted RegIP, on the CUTEst benchmark problems (Gould et al., 2015). RegIP is compared in terms of robustness against the IP solver Ipopt (Wächter and Biegler, 2006) and the AL solver Percival (dos Santos and Siqueira, 2020), which is based on a BCL method (Conn et al., 1991) coupled with a trust-region matrix-free solver (Lin and Moré, 1999) for the subproblems. We do not report runtimes nor iteration counts since a fair comparison would require close inspection of heuristics and fallbacks (Wächter and Biegler, 2006, Sec. 3).

We implemented RegIP in Julia and set up the numerical experiments adopting the JSO software infrastructure by Orban and Siqueira (2019). The IP solver Ipopt acts as subsolver to execute Step 2, warm-started at the current primal (x^{k-1}, y^{k-1}) and dual (y^{k-1}, z^{k-1}) estimates. We use its parameter `tol` to set the (inner) tolerance ϵ_k, disabling other termination conditions, and let Ipopt control the barrier parameter as needed to approximately solve the regularized subproblem.We let the safeguarding set be $Y := \{v \in \mathbb{R}^m \,|\, \|v\|_\infty \leq 10^{20}\}$ and choose $\hat{y}^k$ by projecting the current estimate y^{k-1} onto Y. We set the initial penalty parameter to $\rho_0 = 10^{-6}$, the inner tolerance $\epsilon_0 = \sqrt[3]{\varepsilon}$, and parameters $\theta_\rho = 0.5$, $\kappa_\rho = 0.5$, and $\kappa_\epsilon = 0.5$. RegIP declares success, and returns a ε-KKT stationary point, as soon as $\epsilon_k \leq \varepsilon$ and $C^k \leq \varepsilon$.Instead, if $\epsilon_k \leq \varepsilon$, $C^k > \varepsilon$ and $\rho_k \leq \rho_{\min} := 10^{-20}$, RegIP stops declaring (local) infeasibility. For Ipopt, we set the tolerance `tol` to ε, remove the other desired thresholds, and disable termination based on acceptable iterates. For Percival, we set absolute and relative tolerances to ε.

We consider the CUTEst problems with their default dimension and select all those with at most 1000 variables and constraints, obtaining a test set with 863 problems. All solvers are provided with the default primal-dual initial point, a tolerance $\varepsilon \in \{10^{-3}, 10^{-5}\}$, a time limit of 300 seconds, and the maximum number of iterations set to 10^9. A solver is deemed to solve a problem instance if it returns with a successful status; it fails otherwise. The source codes for the numerical experiments have been archived on Zenodo at DOI: 10.5281/zenodo.7109904.

Table 1. Comparison on CUTEst problems with n variables and m constraints

Size range $\max\{n,m\}$		Tolerance $\varepsilon = 10^{-3}$ W	T+	T-	L	Tolerance $\varepsilon = 10^{-5}$ W	T+	T-	L
RegIP against Ipopt									
0	10	15	418	28	3	17	416	28	3
11	100	14	139	67	6	10	133	73	10
101	1000	9	131	30	3	10	127	34	2
RegIP against Percival									
0	10	16	417	21	10	20	413	23	8
11	100	14	139	63	10	20	123	68	15
101	1000	48	92	28	5	53	84	33	3

Table 1 summarizes the results, stratified by solver, termination tolerance ε and range of problem size $\max\{n, m\}$. For each combination, we indicate the number of times RegIP wins ("W") or loses ("L"), namely it solves a problem that the other solver fails or viceversa, and the number of ties with success ("T+") or failure ("T-"). The results show that RegIP succeeds on more problems than the other solvers, consistently for both low and high accuracy, indicating that the underlying regularized IP approach can form the basis for reliable and scalable solvers.

6. CONCLUSION

This paper has presented a regularized interior point approach to solving constrained nonlinear optimization problems. Operating as an outer regularization layer, a quadratic proximal penalty provides robustness whilst consuming minimal computation effort once embedded into existent interior point codes as a principled inertia correction strategy. Furthermore, regularizing the equality constraints allows to safely adopt more efficient linear algebra routines, while waiving the need for an infeasibility detection mechanism within the subsolver. Preliminary numerical results indicate that a close integration of proximal regularization within interior point schemes is key to provide efficient and robust solvers.

ACKNOWLEDGEMENTS

I gratefully acknowledge the support of Ryan Loxton and the Centre for Optimisation and Decision Science for giving me the opportunity to visit Curtin University. I would also like to thank Hoa T. Bui for her friendly hospitality and lively discussions during this time in Perth.

REFERENCES

Altman, A. and Gondzio, J. (1999). Regularized symmetric indefinite systems in interior point methods for linear and quadratic optimization. *Optimization Methods and Software*, 11(1–4), 275–302. doi:10.1080/10556789908805754.

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. doi:10.1137/060654797.

Armand, P. and Tran, N.N. (2019). Rapid infeasibility detection in a mixed logarithmic barrier-augmented Lagrangian method for nonlinear optimization. *Optimization Methods and Software*, 34(5), 991–1013. doi: 10.1080/10556788.2018.1528250.

Bertolazzi, E., Biral, F., and Da Lio, M. (2007). Real-time motion planning for multibody systems. *Multibody System Dynamics*, 17(2), 119–139. doi:10.1007/s11044-007-9037-7.

Birgin, E.G. and Martínez, J.M. (2014). *Practical Augmented Lagrangian Methods for Constrained Optimization*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Birgin, E.G., Bueno, L.F., and Martínez, J.M. (2016). Sequential equality-constrained optimization for nonlinear programming. *Computational Optimization and Applications*, 65(3), 699–721. doi:10.1007/s10589-016-9849-6.

Cipolla, S. and Gondzio, J. (2022). Proximal stabilized interior point methods for quadratic programming and low-frequency-updates preconditioning techniques. doi: 10.48550/arxiv.2205.01775.

Conn, A.R., Gould, N.I.M., and Toint, P.L. (1991). A globally convergent augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM Journal on Numerical Analysis*, 28(2), 545–572. doi:10.1137/0728030.

Curtis, F.E. (2012). A penalty-interior-point algorithm for nonlinear constrained optimization. *Mathematical Programming Computation*, 4(2), 181–209. doi:10.1007/s12532-012-0041-4.

De Marchi, A. (2021). Augmented Lagrangian methods as dynamical systems for constrained optimization. In *60th IEEE Conference on Decision and Control (CDC)*, 6533–6538. IEEE, Austin, TX. doi:10.1109/CDC45484.2021.9683199.

De Marchi, A. (2022). On a primal-dual Newton proximal method for convex quadratic programs. *Computational Optimization and Applications*. doi:10.1007/s10589-021-00342-y.

De Marchi, A. and Themelis, A. (2022). An interior proximal gradient method for nonconvex optimization. doi:10.48550/arxiv.2208.00799.

Dehghani, M., Lambe, A., and Orban, D. (2020). A regularized interior-point method for constrained linear least squares. *INFOR: Information Systems and Operational Research*, 58(2), 202–224. doi:10.1080/03155986.2018.1559428.

Diehl, M., Ferreau, H.J., and Haverbeke, N. (2009). *Efficient numerical methods for nonlinear MPC and moving horizon estimation*, 391–417. Springer. doi:10.1007/978-3-642-01094-1_32.

dos Santos, E.A. and Siqueira, A.S. (2020). Percival.jl: an augmented Lagrangian method. doi:10.5281/zenodo.3969045. URL https://github.com/JuliaSmoothOptimizers/Percival.jl.

Fiacco, A.V. and McCormick, G.P. (1968). *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*. Wiley, New York.

Friedlander, M.P. and Orban, D. (2012). A primal-dual regularized interior-point method for convex quadratic programs. *Mathematical Programming Computation*, 4(1), 71–107. doi:10.1007/s12532-012-0035-2.

Gondzio, J. (2012). Interior point methods 25 years later. *European Journal of Operational Research*, 218(3), 587–601. doi:10.1016/j.ejor.2011.09.017.

Gould, N.I.M., Orban, D., and Toint, P.L. (2015). CUTEst: a constrained and unconstrained testing environment with safe threads for mathematical optimization. *Computational Optimization and Applications*, 60(3), 545–557. doi:10.1007/s10589-014-9687-3.

Liao-McPherson, D. and Kolmanovsky, I. (2020). FBstab: A proximally stabilized semismooth algorithm for convex quadratic programming. *Automatica*, 113, 108801. doi:10.1016/j.automatica.2019.108801.

Lin, C.J. and Moré, J.J. (1999). Newton's method for large bound-constrained optimization problems. *SIAM Journal on Optimization*, 9(4), 1100–1127. doi:10.1137/S1052623498345075.

Ma, D., Judd, K.L., Orban, D., and Saunders, M.A. (2018). Stabilized optimization via an NCL algorithm. In M. Al-Baali, L. Grandinetti, and A. Purnama (eds.), *Numerical Analysis and Optimization*, 173–191. Springer. doi:10.1007/978-3-319-90026-1_8.

Orban, D. and Siqueira, A.S. (2019). JuliaSmoothOptimizers: Infrastructure and solvers for continuous optimization in Julia. doi:10.5281/zenodo.2655082. URL https://juliasmoothoptimizers.github.io.

Orban, D. and Siqueira, A.S. (2020). A regularization method for constrained nonlinear least squares. *Computational Optimization and Applications*, 76(3), 961–989. doi:10.1007/s10589-020-00201-2.

Potschka, A. and Bock, H.G. (2021). A sequential homotopy method for mathematical programming problems. *Mathematical Programming*, 187(1), 459–486. doi:10.1007/s10107-020-01488-z.

Robinson, D.P. (2007). *Primal-Dual Methods for Nonlinear Optimization*. Ph.D. thesis, University of California, San Diego.

Rockafellar, R.T. (1974). Augmented Lagrange multiplier functions and duality in nonconvex programming. *SIAM Journal on Control*, 12(2), 268–285. doi:10.1137/0312021.

Rockafellar, R.T. (1976). Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5), 877–898. doi:10.1137/0314056.

Saraf, N. and Bemporad, A. (2022). An efficient bounded-variable nonlinear least-squares algorithm for embedded MPC. *Automatica*, 141, 110293. doi:10.1016/j.automatica.2022.110293.

Sopasakis, P., Fresk, E., and Patrinos, P. (2020). OpEn: Code generation for embedded nonconvex optimization. *IFAC-PapersOnLine*, 53(2), 6548–6554. doi:10.1016/j.ifacol.2020.12.071. 21st IFAC World Congress.

Vanderbei, R.J. (1995). Symmetric quasidefinite matrices. *SIAM Journal on Optimization*, 5(1), 100–113. doi: 10.1137/0805005.

Vanderbei, R.J. and Shanno, D.F. (1999). An interior-point algorithm for nonconvex nonlinear programming. *Computational Optimization and Applications*, 13(1), 231–252. doi:10.1023/A:1008677427361.

Wächter, A. and Biegler, L.T. (2006). On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming*, 106(1), 25–57. doi:10.1007/s10107-004-0559-y.