Math. Program., Ser. A (2017) 161:73–134
DOI 10.1007/s10107-016-1003-9

FULL LENGTH PAPER

An interior-point trust-funnel algorithm for nonlinear optimization

Frank E. Curtis[1] · **Nicholas I. M. Gould**[2] · **Daniel P. Robinson**[3] · **Philippe L. Toint**[4]

Received: 20 December 2013 / Accepted: 8 March 2016 / Published online: 7 April 2016

Abstract We present an interior-point trust-funnel algorithm for solving large-scale nonlinear optimization problems. The method is based on an approach proposed by Gould and Toint (Math Prog 122(1):155–196, 2010) that focused on solving equality constrained problems. Our method is similar in that it achieves global convergence guarantees by combining a trust-region methodology with a funnel mechanism, but has the additional capability of being able to solve problems with both equality and inequality constraints. The prominent features of our algorithm are that (i) the subproblems that define each search direction may be solved with matrix-free methods so

Frank E. Curtis: This author was supported by U.S. Department of Energy Grant DE–SC0010615 and U.S. National Science Foundation Grant DMS–1016291.
Nicholas I. M. Gould: This author was supported by the EPSRC Grant EP/I013067/1.
Daniel P. Robinson: This author was supported by U.S. National Science Foundation Grant DMS–1217153.

✉ Daniel P. Robinson
daniel.p.robinson@gmail.com

Frank E. Curtis
frank.e.curtis@gmail.com

Nicholas I. M. Gould
nick.gould@stfc.ac.uk

Philippe L. Toint
philippe.toint@fundp.ac.be

1 Department of Industrial and Systems Engineering, Lehigh University, Bethlehem, PA, USA

2 Numerical Analysis Group, Rutherford Appleton Laboratory, Chilton, Oxfordshire, UK

3 Department of Applied Mathematics and Statistics, Johns Hopkins University, Baltimore, MD, USA

4 Department of Mathematics and NaXys, University of Namur, Namur, Belgium

that derivative matrices need not be formed or factorized so long as matrix-vector products with them can be performed; (ii) the subproblems may be solved approximately in all iterations; (iii) in certain situations, the computed search directions represent inexact sequential quadratic optimization steps, which may be desirable for fast local convergence; (iv) criticality measures for feasibility and optimality aid in determining whether only a subset of computations need to be performed during a given iteration; and (v) no merit function or filter is needed to ensure global convergence.

Keywords Nonlinear optimization · Constrained optimization · Large-scale optimization · Barrier-SQP methods · Trust-region methods · Funnel mechanism

Mathematics Subject Classification 49J52 · 49M37 · 65F22 · 65K05 · 90C26 · 90C30 · 90C55

1 Introduction

We introduce a method for solving optimization problems of the form

$$\underset{x\in\mathbb{R}^N}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad c(x) \leq 0, \tag{NP}$$

where $f : \mathbb{R}^N \to \mathbb{R}$ and $c : \mathbb{R}^N \to \mathbb{R}^M$ are twice continuously differentiable. (Our method can also be applied when equality constraints are present, but, for simplicity in our discussion, these are suppressed in our algorithm development and analysis; see Sect. 6 for further discussion.) Our algorithm is designed to solve large-scale instances of (NP). In particular, it is designed to be matrix-free in the sense that an implementation of it only requires matrix-vector products with the constraint Jacobian, its transpose, symmetric approximations of the Hessian of the Lagrangian, and corresponding preconditioners. Consequently, iterative methods may be used to approximately solve each subproblem arising in the algorithm.

The method we propose utilizes components of both interior-point (IP) and sequential quadratic optimization (commonly known as SQP) methods. Algorithms of this type are often referred to as barrier-SQP methods. The interior-point aspects of our algorithm allow us to avoid the combinatorial explosion that may occur within, say, an active-set approach. The efficiency of interior-point methods for solving linear and convex quadratic optimization problems has been well-established [1,7,12,13,17,24,28,30,31]. Extending these methods for solving nonlinear problems has been the subject of research for decades [3,4,6,14,32–36] and numerical evidence illustrates strong performance. We follow an approach similar to Byrd et al. [3,4] and solve a sequence of barrier subproblems for decreasing values of the barrier parameter. This means that we must solve a sequence of equality constrained subproblems, and these may be solved efficiently with an SQP-based method. It is well known that traditional SQP methods are very efficient for solving small- to medium-sized optimization problems [8,9,15,16], while more recently proposed SQP methods utilize exact second derivatives and are, in theory, capable of solving large problems [19–21,29]. Preliminary results when solving small- to medium-sized problems are promising, but their effectiveness on large problems has not yet been confirmed. There have, however, been several SQP strategies that have proved capable of solving large equality constrained problems [2,23,27].

In this paper, we use the trust-funnel approach originally described in [23], and then corrected in [22], as the basis for solving a sequence of equality constrained barrier subproblems that arise in an interior-point framework. We note, however, that a naïve implementation of the SQP method described in [22,23] within an interior-point paradigm may result in a method for which the establishment of convergence guarantees is elusive. This is a consequence of the fact that interior-point methods—as their name suggests—require the algorithm iterates to remain in the strict interior of the feasible region associated with the inequality constraints, while the method in [22,23] does not innately possess the mechanisms necessary to avoid the boundary of the feasible region in this context. In this paper, we describe modifications of this trust-funnel method that are appropriate for our interior-point setting. These modifications include imposing explicit constraints in the trust-region subproblems to ensure that the iterates remain in the strict interior of the feasible region, and the incorporation of scaled trust-region constraints and optimality measures. Scalings of these types have been used previously [3,6].

The paper is organized as follows. In Sect. 2, to motivate our main ideas, we outline a preliminary trust-funnel algorithm for solving the barrier subproblem in an interior-point approach. This method, which requires the exact solution of subproblems in each iteration, forms the basis for our main trust-funnel algorithm, presented in Sect. 3, which involves various enhancements vis-à-vis the method in Sect. 2. In Sect. 4, we prove that our main trust-funnel algorithm will terminate finitely with arbitrarily small positive tolerances on appropriate criticality measures. In Sect. 5, we consider convergence of the barrier subproblem solutions for a decreasing sequence of the barrier parameter. Finally, conclusions are provided in Sect. 6.

1.1 Notation

The gradient and Hessian of f at x are written as $g(x)$ and $\nabla_{xx} f(x)$ respectively. The $M \times N$ matrix $J(x)$ represents the Jacobian of the constraint function c evaluated at x, with its jth row being $\nabla c_j(x)^T$. The matrix $\nabla_{xx} c_j(x)$ is the Hessian of c_j evaluated at x. We let e denote the vector of all ones and I denote the identity matrix, both of whose dimensions are determined by the context in which they are used. Given a vector $s \in \mathbb{R}^M$, $[s]_j$ is the jth element of s and $S := \operatorname{diag}([s]_1, [s]_2, \ldots, [s]_M)$. A forcing function $\omega : [0, \infty) \to [0, \infty)$ is defined as any continuous and strictly increasing function that satisfies $\omega(0) = 0$. For a real symmetric matrix P, we write $P \succ 0$ to indicate that P is positive definite. Finally, given two scalar sequences $\{a_j\}$ and $\{b_j\}$, we write $a_j = \mathcal{O}(b_j)$ to indicate that there exists a constant $c > 0$ such that $a_j \le c b_j$ for all j.

1.2 NLP and barrier-SQP preliminaries

We make the following assumption throughout the paper.

Assumption 1.1 The functions f and c are twice continuously differentiable.

In fact, the global convergence guarantees that we establish for our algorithm hold even if f and c are only once continuously differentiable and (uniformly bounded) Hessian approximations are employed. However, for simplicity in our discussion and in order to provide commentary on algorithmic choices that should be made to achieve fast local convergence, we make Assumption 1.1.

Problem (NP) is not solved directly by our algorithm. Rather, we introduce a vector of slack variables $s \in \mathbb{R}^M$ and solve the equivalent optimization problem

$$\underset{x\in\mathbb{R}^N, s\in\mathbb{R}^M}{\text{minimize}} \quad f(x) \quad \text{subject to} \quad c(x,s) := c(x) + s = 0, \quad s \geq 0. \tag{NPs}$$

The following definition gives first-order stationarity conditions for (NPs) [25,26].

Definition 1.1 *(First-order KKT point for* (NPs)*)* The vector triple (x, s, y) is a first-order KKT point for problem (NPs) if it satisfies

$$g(x) + J(x)^T y = 0, \;\; c(x,s) = 0, \;\; Sy = 0, \;\; \text{and } (s, y) \geq 0.$$

To solve (NPs), we (approximately) solve the barrier subproblem

$$\underset{x\in\mathbb{R}^N, s\in\mathbb{R}^M}{\text{minimize}} \quad f(x,s) \quad \text{subject to} \quad c(x,s) = 0, \quad s > 0 \tag{BSP}$$

for decreasing values of the barrier parameter $\mu > 0$, where we define

$$f(x,s) := f(x) - \mu \sum_{i=1}^{M} \ln([s]_i). \tag{1.1}$$

Given a Lagrange multiplier vector y for the constraint $c(x,s) = 0$, the Lagrangian associated with (BSP) and its gradient with respect to (x, s) are

$$\mathcal{L}(x,s,y) := f(x,s) + c(x,s)^T y \quad \text{and} \quad \nabla_{(x,s)}\mathcal{L}(x,s,y) := \nabla f(x,s) + J(x,s)^T y,$$

where $J(x,s) := \nabla c(x,s)^T = \begin{pmatrix} J(x) & I \end{pmatrix}$ is the Jacobian of $c(x,s)$ with respect to (x, s). A primal-dual point (x, s, y) is a first-order KKT point of the barrier subproblem if it satisfies $\nabla_{(x,s)}\mathcal{L}(x,s,y) = 0$, $c(x,s) = 0$ and $(s, y) > 0$. Multiplying the second block of the first equation by S leads to the following equivalent definition.

Definition 1.2 (*First-order KKT point for* (BSP)) The vector triple (x, s, y) is a first-order KKT-point for the barrier subproblem (BSP) if it satisfies

$$g(x) + J(x)^T y = 0, \;\; c(x,s) = 0, \;\; Sy = \mu e, \;\; \text{and } (s, y) > 0.$$

A comparison of Definitions 1.1 and 1.2 suggests that, as $\mu \to 0$, KKT points of the barrier subproblem become increasingly accurate KKT points of problem (NPs).

Our trust-funnel strategy generates a sequence $\{(x_k, s_k, y_k)\}$ of primal, slack, and dual variables. As is typical of interior-point methods, we require $s_0 > 0$ and ensure $s_k > 0$ for all k via explicit constraints imposed on all search direction calculations, and ensure that $c(x_k, s_k) \geq 0$ holds at the beginning of iteration k by incorporating the *slack reset* procedure (for all $i \in \{1, \ldots, M\}$)

$$[s_k]_i \leftarrow \begin{cases} [s_k]_i & \text{if } \; [c(x_k, s_k)]_i \geq 0, \\ -[c(x_k)]_i & \text{otherwise.} \end{cases} \tag{1.2}$$

Defining the measure of constraint violation

$$v(x, s) := \|c(x, s)\|_2, \tag{1.3}$$

it follows that if s_k^{prior} is the value of s_k *prior* to the slack reset, then

$$v_k := v(x_k, s_k) \leq v(x_k, s_k^{\text{prior}}), \quad s_k^{\text{prior}} \leq s_k, \quad \text{and} \quad f(x_k, s_k) \leq f(x_k, s_k^{\text{prior}}); \tag{1.4}$$

i.e., the barrier function and constraint violation decrease due to (1.2).

For reference, we now describe the step computation of a conventional SQP method for solving the barrier subproblem (BSP). Given a kth iterate (x_k, s_k, y_k), the trial step in such a method is defined as the solution (when it exists) of

$$\begin{aligned} &\underset{d=(d^x, d^s)}{\text{minimize}} \;\; f(x_k, s_k) + \nabla f(x_k, s_k)^T d + \tfrac{1}{2} d^T \nabla_{(x,s)(x,s)} \mathcal{L}(x_k, s_k, y_k) d \\ &\text{subject to} \;\; c(x_k, s_k) + J(x_k, s_k) d = 0. \end{aligned}$$

It may be verified that a solution $d = (d^x, d^s)$ of this subproblem satisfies

$$\begin{pmatrix} \nabla_{xx} \mathcal{L}(x_k, s_k, y_k) & J(x_k)^T & 0 \\ J(x_k) & 0 & I \\ 0 & S_k & \mu S_k^{-1} \end{pmatrix} \begin{pmatrix} d^x \\ y \\ d^s \end{pmatrix} = - \begin{pmatrix} g(x_k) \\ c(x_k, s_k) \\ -\mu e \end{pmatrix}, \tag{1.5}$$

where y is an estimate of an optimal Lagrange multiplier vector for the constraint $c(x_k, s_k) + J(x_k, s_k)d = 0$. The SQP step generated in this fashion is often called a *primal* step since the dual vector y_k does not appear in (1.5) other than in the Hessian $\nabla_{xx}\mathcal{L}$. We can instead compute a *primal-dual* step by applying Newton's Method to the conditions in Definition 1.2, which leads to

$$\begin{pmatrix} \nabla_{xx} \mathcal{L}(x_k, s_k, y_k) & J(x_k)^T & 0 \\ J(x_k) & 0 & I \\ 0 & S_k & Y_k \end{pmatrix} \begin{pmatrix} d^x \\ y \\ d^s \end{pmatrix} = - \begin{pmatrix} g(x_k) \\ c(x_k, s_k) \\ -\mu e \end{pmatrix}. \tag{1.6}$$

This system is identical to (1.5), except that the (3, 3)-block now contains dual information. It is easily verified that a solution of (1.6) is a KKT point for

$$\begin{aligned} \underset{d=(d^x,d^s)}{\text{minimize}} \quad & f(x_k, s_k) + \nabla f(x_k, s_k)^T d + \tfrac{1}{2} d^T G_k d \\ \text{subject to} \quad & c(x_k, s_k) + J(x_k, s_k) d = 0, \end{aligned}$$

where

$$G_k := \begin{pmatrix} \nabla_{xx}\mathcal{L}(x_k, s_k, y_k) & 0 \\ 0 & Y_k S_k^{-1} \end{pmatrix}. \tag{1.7}$$

In contrast to the conventional SQP trial step computation described in the previous paragraph, our trust-funnel algorithm employs a *step decomposition* approach. In particular, given (x_k, s_k), a trial step $d_k := (d_k^x, d_k^s)$ is computed as the sum of a "normal" step $n_k := (n_k^x, n_k^s)$ and a "tangential" step $t_k := (t_k^x, t_k^s)$, i.e.,

$$d_k = \begin{pmatrix} d_k^x \\ d_k^s \end{pmatrix} = \begin{pmatrix} n_k^x \\ n_k^s \end{pmatrix} + \begin{pmatrix} t_k^x \\ t_k^s \end{pmatrix} = n_k + t_k.$$

The normal step n_k is computed to minimize a Gauss-Newton model of v at (x_k, s_k); thus, it has the purpose of reducing linearized infeasibility. The tangential step t_k is intended to reduce the barrier function (1.1) and is calculated as an minimizer of a quadratic model of the barrier function within an appropriate subspace that does not undo the improvement in reducing linearized infeasibility achieved by n_k. Once $d_k = n_k + t_k$ is computed, an attempt to decrease the constraint violation and/or barrier function is made, where the decision of which to consider is based on quantities that reflect the overall merit of the constituent steps. A detailed explanation of these aspects is given for a preliminary algorithm in Sect. 2 and for our complete algorithm in Sect. 3.

2 A preliminary trust-funnel algorithm for the barrier subproblem

In this section, we present a preliminary trust-funnel algorithm for solving the barrier subproblem (BSP) for a fixed value of the barrier parameter $\mu > 0$. As μ is fixed for a particular instance of (BSP), the dependence on μ of quantities in this section is ignored. However, these dependencies—in particular, with respect to criticality tolerances that are employed in the algorithm—will be a central focus in Sect. 5 when we address the "outer" algorithm for solving problem (NPs).

The algorithm in this section is presented merely to motivate the features of our main algorithm in Sect. 3. Indeed, there are various aspects of the algorithm in this section that may result in computational inefficiencies; most notably, it involves the (exact) solution of a sequence of subproblems during every iteration. By contrast, our main algorithm involves features that aid in avoiding certain computations when they are deemed unnecessary, and it allows for the inexact solution of subproblems. Still, the presentation of the algorithm in this section should aid the reader in understanding the overall strategy of our main algorithm.

2.1 Funnel mechanism

The signifying feature of a funnel method is a sequence, which we call $\{v_k^{\max}\}$, of positive and monotonically decreasing scalars that guide the iterates toward constraint satisfaction. In particular, in our approach, we ensure that

$$s_k > 0, \quad c(x_k, s_k) \geq 0, \quad v_k \leq v_k^{\max}, \quad \text{and} \quad v_{k+1}^{\max} \leq v_k^{\max} \quad \text{for all} \quad k. \tag{2.1}$$

The set of points permitted by the gradually narrowing region defined by $v(x, s) \leq v_k^{\max}$ is the *funnel* [22,23], and the elements of $\{v_k^{\max}\}$ are the funnel radii.

2.2 Step computations

Each iteration of our preliminary algorithm involves the sequential solution of three subproblems: the first to compute a normal step toward linearized constraint satisfaction, the second to compute a new Lagrange multiplier estimate, and the third to compute a tangential step toward optimality. The purpose of this section is to define the quantities and subproblems involved in these computations.

The normal step is designed to predict a reduction in constraint violation. To achieve this goal, consider the step $n_k := (n_k^x, n_k^s)$ as a solution of

$$\underset{n=(n^x,n^s)}{\text{minimize}} \;\; m_k^v(n) \quad \text{subject to} \quad \|P_k^{-1}n\|_2 \leq \delta_k^v, \quad s_k + n^s \geq \kappa_{\text{fbn}} s_k, \tag{2.2}$$

where we define the linearized constraint violation measure and scaling matrix

$$m_k^v(n) := \|c(x_k, s_k) + J(x_k, s_k)n\|_2 \quad \text{and} \quad P_k := \begin{pmatrix} I & 0 \\ 0 & S_k \end{pmatrix} \tag{2.3}$$

along with the fraction-to-the-boundary (e.g., see [32, § 2.2]) constant $\kappa_{\text{fbn}} \in (0, 1)$ and trust region radius $\delta_k^v > 0$. Our introduction of the scaling matrix P_k can be motivated in multiple ways. On the one hand, in terms of defining the trust region constraint in (2.2), it can be motivated as a means of keeping the iterates sufficiently within the nonnegative orthant; e.g., it aids in restricting $[n_k^s]_j$ to be relatively small when $[s_k]_j$ is close to zero [3]. More importantly, however, its introduction can be motivated by the constraint violation minimization problem

$$\underset{x \in \mathbb{R}^N, s \in \mathbb{R}^M}{\text{minimize}} \;\; \tfrac{1}{2} v(x, s)^2 \quad \text{subject to} \quad s \geq 0, \tag{2.4}$$

for which we have the first-order KKT conditions

$$\min\{s, c(x, s)\} = 0 \quad \text{and} \quad J(x)^T c(x, s) = 0. \tag{2.5}$$

A point (x, s) with $s \geq 0$ and $c(x, s) \geq 0$ [recall (2.1)] satisfies (2.5) as long as

$$0 = P_k J(x_k, s_k)^T c(x_k, s_k) = (J(x_k)^T c(x_k, s_k), S_k c(x_k, s_k)). \tag{2.6}$$

With the normal step n_k in hand, our preliminary algorithm next computes a new Lagrange multiplier estimate. For this purpose, we let y_k be the solution of

$$\begin{aligned} &\underset{y\in\mathbb{R}^M}{\text{minimize}}\ m_k^{\mathcal{L}}(y), \\ &\text{where}\quad m_k^{\mathcal{L}}(y) := \tfrac{1}{2}\left\| P_k\left(\nabla f(x_k, s_k) + \hat{G}_k n_k + J(x_k, s_k)^T y\right)\right\|_2^2, \end{aligned} \tag{2.7}$$

where $\hat{G}_k$ has the same form as in (1.7), but with y_k replaced by y_{k-1}. This subproblem can be motivated by observing that its objective function is a valid criticality measure for minimizing the barrier function; recall the first-order KKT conditions for (BSP) and see Sect. 3.2. The role of y_k is two-fold: it is used in the formulation of the Hessian in the tangential subproblem and in checking stationarity conditions for termination of the algorithm.

After the new Lagrange multiplier estimate has been computed, we define—now using the Hessian matrix G_k in (1.7) associated with the conventional SQP subproblem—the tangential subproblem objective function

$$m_k^f(d) := f(x_k, s_k) + \nabla f(x_k, s_k)^T d + \tfrac{1}{2} d^T G_k d. \tag{2.8}$$

Our tangential step is then defined as a solution of the subproblem

$$\begin{aligned} &\underset{t=(t^x,t^s)}{\text{minimize}}\ \ m_k^f(n_k + t) \\ &\text{subject to}\ \ J(x_k, s_k)t = 0, \\ &\qquad \|P_k^{-1}(n_k + t)\|_2 \le \min\{\kappa_{\text{vf}}\delta_k^v, \delta_k^f\},\quad s_k + n_k^s + t^s \ge \kappa_{\text{fbt}}(s_k + n_k^s), \end{aligned} \tag{2.9}$$

where $\kappa_{\text{vf}} > 0$ and $\kappa_{\text{fbt}} \in (0, 1)$ are constants and $\delta_k^f > 0$ is a trust region radius.

2.3 Iteration types and step acceptance

With the normal and tangential steps computed, we must decide how to set the next iterate (x_{k+1}, s_{k+1}), pair of trust region radii δ_{k+1}^v and δ_{k+1}^f, and funnel radius $v_{k+1}^{\max}$. In our approach, these choices depend on first gauging whether progress in reducing the barrier function, the constraint violation, or perhaps neither, is most likely to occur. Specifically, we use the calculated steps to characterize the iteration as a *y-iteration*, *f-iteration* or *v-iteration* in the spirit of [9–11]. The new iterate, trust region radii, and funnel radius are then set based on whether the progress predicted within a given iteration type is realized at the trial point

$$(x_k^+, s_k^+) := (x_k, s_k) + d_k.$$

A y-iteration is any iteration satisfying the following definition.

Definition 2.1 (*y-iteration*) The kth iteration is a y-iteration if $d_k = 0$.

Note that a y-iteration will occur when n_k and t_k are both equal to zero, so that the only outcome of the iteration is a new Lagrange multiplier estimate. Therefore, in such an iteration, we leave the values of the iterate, trust-region radii, and funnel radius unchanged. For our preliminary algorithm, the kth iteration can be a y-iteration only if (x_k, s_k, y_k) is a first-order KKT point for the barrier subproblem; however, in our main trust-funnel algorithm in Sect. 3, y-iterations may occur more frequently when inexact subproblem solutions are allowed and encouraged.

The primary goal of an f-iteration is to reduce the barrier function. In this context, we are interested in the predicted change in the barrier function by the normal step and tangential step as given, respectively, by

$$\Delta m_k^{f,n} := m_k^f(0) - m_k^f(n_k) \quad \text{and} \quad \Delta m_k^{f,t} := m_k^f(n_k) - m_k^f(n_k + t_k).$$

To judge the potential for the full step d_k to decrease the barrier function, we test whether the following inequality holds:

$$\Delta m_k^{f,d} := \Delta m_k^{f,n} + \Delta m_k^{f,t} \geq \kappa_\delta \Delta m_k^{f,t} \ \text{ for some } \ \kappa_\delta \in (0, 1). \tag{2.10}$$

Satisfaction of (2.10) indicates that the decrease in the barrier function predicted by d_k is at least a fraction of that predicted by the tangential step t_k. Based on this observation and the idea of using $v_k \leq v_k^{\max}$ for all k to guide the algorithm toward constraint satisfaction, the following definition is natural.

Definition 2.2 (*f-iteration*) The kth iteration is an f-iteration if $t_k \neq 0$, the inequality (2.10) holds, and

$$v(x_k^+, s_k^+) \leq v_k^{\max}. \tag{2.11}$$

As for conventional trust-region methods, the updates applied at the end of an f-iteration are based on the quantity

$$\rho_k^f := \frac{f(x_k, s_k) - f(x_k^+, s_k^+)}{\Delta m_k^{f,d}}, \tag{2.12}$$

which measures the ratio of actual-to-predicted decrease in the barrier function. In short, if the kth iteration is an f-iteration and $\rho_k^f \geq \eta_1$ for some prescribed constant $\eta_1 \in (0, 1)$, then the trial point is accepted as the new iterate, the funnel radius is left unchanged, and the trust-region radii are potentially increased.

Finally, when the conditions defining a y- and/or f-iteration are not satisfied, the iteration type defaults to that of a v-iteration.

Definition 2.3 (*v-iteration*) The kth iteration is a v-iteration if it is not a y- or an f-iteration, i.e., if $d_k \neq 0$ and either $t_k = 0$, the inequality (2.10) does not hold, or the inequality (2.11) does not hold.

Though perhaps not readily apparent from this definition, the main achievement of a v-iteration is a predicted reduction in constraint violation. (This fact will be clear in the analysis of our main algorithm). Analogous to f-iterations, our updating strategy for v-iterations depends on the quantity

$$\rho_k^v := \frac{v_k - v(x_k^+, s_k^+)}{\Delta m_k^{v,d}} \tag{2.13}$$

that measures the ratio of actual-to-predicted decrease in the constraint violation. It also depends, however, on the predicted change in the constraint violation for the normal and full trial steps, for which we define

$$\Delta m_k^{v,n} := m_k^v(0) - m_k^v(n_k) \quad \text{and} \quad \Delta m_k^{v,d} := m_k^v(0) - m_k^v(d_k). \tag{2.14}$$

Specifically, if the kth iteration is a v-iteration, $\rho_k^v \geq \eta_1$,

$$n_k \neq 0, \quad \text{and} \quad \Delta m_k^{v,d} \geq \kappa_{\text{cd}} \Delta m_k^{v,n} \quad \text{for some} \quad \kappa_{\text{cd}} \in (0, 1), \tag{2.15}$$

then the trial point is accepted as the new iterate, the normal step trust region radius may be increased, and the funnel radius is reduced. (Briefly, the second condition in (2.15), (2.13), and the fact that $\Delta m_k^{v,n}$ is nonnegative due to the normal step computation together imply that $v(x_k^+, s_k^+) < v(x_k, s_k)$.)

2.4 A preliminary trust-funnel algorithm

We are now prepared to state our preliminary algorithm, stated as Algorithm 1 on page 10. It should be noted that while Algorithm 1 outlines the main computational steps in our main approach (see Sect. 3), we do not claim that it is well-defined and/or globally convergent. Indeed, for simplicity, we have stated the algorithm without termination conditions or algorithmic features that would be necessary to ensure that it is well-posed. We have also not given concrete updates for various quantities (e.g., specific trust-region radii updates), since this would distract the reader from understanding the core ideas. Finally, we claim that Algorithm 1 possesses various inefficiencies. For example, despite the fact that the algorithm calls for the computation of a normal step in every iteration, this computation could be wasteful if a given iterate is (nearly) stationary for the measure of infeasibility and significant progress could be made simply by computing a new multiplier estimate and tangential step. These types of situations motivate the various algorithmic features and opportunities for exploiting inexact solutions that are introduced along with the description of our main algorithm in the following section.

3 A trust-funnel algorithm for the barrier subproblem

In this section, we present our main trust-funnel algorithm, which is designed to improve upon the preliminary algorithm of Sect. 2 in two key ways. First, we introduce conditions under which one can exploit inexact solutions of the subproblems defining

Algorithm 1 Preliminary trust-funnel algorithm for the barrier subproblem (BSP)

Input: (x_0, s_0, μ) with $(s_0, \mu) > 0$.
Choose $\{\delta_0^v, \delta_0^f, \kappa_{vf}\} \in (0, \infty)$ and $\{\eta_1, \kappa_\delta, \kappa_{fbn}, \kappa_{fbt}, \kappa_{cd}\} \subset (0, 1)$.
Perform a slack reset to s_0 as given by (1.2).
Set $v_0^{\max} \geq v(x_0, s_0)$.
for $k = 0, 1, \ldots$ **do**
  Compute a normal step n_k that solves (2.2).
  Compute a multiplier vector y_k that solves (2.7).
  Compute a tangential step t_k that solves (2.9).
  Set the trial step $d_k \leftarrow n_k + t_k$ and trial iterate $(x_k^+, s_k^+) \leftarrow (x_k, s_k) + d_k$.
  if $d_k = 0$ **then** [y-iteration]
    Set $(x_{k+1}, s_{k+1}) \leftarrow (x_k, s_k)$, $\delta_{k+1}^v \leftarrow \delta_k^v$, $\delta_{k+1}^f \leftarrow \delta_k^f$, and $v_{k+1}^{\max} \leftarrow v_k^{\max}$.
  else if $t_k \neq 0$ and both (2.10) and (2.11) hold **then** [f-iteration]
    if $\rho_k^f \geq \eta_1$ **then**
      Set $(x_{k+1}, s_{k+1}) \leftarrow (x_k^+, s_k^+)$, $\delta_{k+1}^v \geq \delta_k^v$, $\delta_{k+1}^f \geq \delta_k^f$, and $v_{k+1}^{\max} \leftarrow v_k^{\max}$.
    else
      Set $(x_{k+1}, s_{k+1}) \leftarrow (x_k, s_k)$, $\delta_{k+1}^v \leftarrow \delta_k^v$, $\delta_{k+1}^f \in (0, \delta_k^f)$, and $v_{k+1}^{\max} \leftarrow v_k^{\max}$.
  else [v-iteration]
    if $\rho_k^v \geq \eta_1$ and (2.15) holds **then**
      Set $(x_{k+1}, s_{k+1}) \leftarrow (x_k^+, s_k^+)$, $\delta_{k+1}^v \geq \delta_k^v$, $\delta_{k+1}^f \leftarrow \delta_k^f$, and $v_{k+1}^{\max} \in (0, v_k^{\max})$.
    else
      Set $(x_{k+1}, s_{k+1}) \leftarrow (x_k, s_k)$, $\delta_{k+1}^v \in (0, \delta_k^v)$, $\delta_{k+1}^f \leftarrow \delta_k^f$, and $v_{k+1}^{\max} \leftarrow v_k^{\max}$.
  Perform a slack reset to s_{k+1} as given by (1.2).

the normal step, Lagrange multiplier estimate, and tangential step. This is important since, in large-scale settings, it is often preferable to employ iterative solvers, and the opportunity of accepting inexact solutions allows for early termination of such solvers. Second, to further reduce computational costs, we establish conditions under which one can completely avoid computation of the normal step, Lagrange multiplier estimate, and/or tangential step during certain iterations. The core strategy of the algorithm in this section follows that of Algorithm 1 described in Sect. 2, but, in order to ensure global convergence of our algorithm (which allows much computational flexibility), intricate sets of conditions and safeguards are necessary. These are the main topics of discussion in this section.

3.1 An inexact normal step

We begin our description of a technique for computing an inexact normal step by introducing the "v-criticality" measures [recall (2.6)] given by

$$\pi_k^v := \pi^v(x_k, s_k) := \|P_k J(x_k, s_k)^T c(x_k, s_k)\|_2 \quad \text{and} \tag{3.1a}$$

$$\chi_k^v := \chi^v(x_k, s_k) := \begin{cases} \pi_k^v / v_k & \text{if } v_k > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{3.1b}$$

We use these measures to determine when a normal step must be computed. In particular, we only require a normal step to be computed when either the v-criticality measure π_k^v is large relative to an "f-criticality" measure π_{k-1}^f (defined in (3.14) and associated with minimizing the barrier function), or when v_k is large relative to $v_k^{\max}$.

Specifically, for some $\kappa_{\text{vv}} \in (0, 1)$ and forcing function ω_n, we require the computation of a normal step if either

$$\pi_k^v > \omega_n(\pi_{k-1}^f) \quad \text{or} \quad v_k \geq \kappa_{\text{vv}} v_k^{\max}. \tag{3.2}$$

(If (3.2) does not hold, but $\pi_k^v > 0$, then one may still consider computing a normal step since the fact that $\pi_k^v > 0$ implies that the computation would be well-defined. However, in such cases, a normal step is not necessary for our convergence analysis.) When a normal step is not computed, we set $n_k \leftarrow 0$.

If a normal step $n_k := (n_k^x, n_k^s)$ is computed, then it is computed as an approximate solution to (2.2), meaning that it should be feasible for (2.2) and yield a decrease in m_k^v no less than that achieved along a scaled steepest descent direction for m_k^v. The scaled steepest descent direction that we employ in this setting is derived in the following manner. Performing the change of variables $n^{\text{P}} := P_k^{-1} n$ so that the trust-region constraint becomes $\|n^{\text{P}}\|_2 \leq \delta_k^v$, the transformed problem for minimizing m_k^v has the steepest descent direction $-P_k J(x_k, s_k)^T c(x_k, s_k)$. Returning to the original space gives the scaled steepest descent direction $-P_k^2 J(x_k, s_k)^T c(x_k, s_k)$. For (2.2), we define the Cauchy step $n_k^{\text{C}} = (n_k^{\text{C}x}, n_k^{\text{C}s})$ as the minimizer of the objective of (2.2) in this scaled steepest descent direction, i.e.,

$$n_k^{\text{C}} := n_k^{\text{C}}(\alpha_{\text{N}}^{\text{C}}), \quad \text{where} \quad n_k^{\text{C}}(\alpha) := \begin{pmatrix} n_k^{\text{C}x}(\alpha) \\ n_k^{\text{C}s}(\alpha) \end{pmatrix} := -\alpha P_k^2 J(x_k, s_k)^T c(x_k, s_k) \tag{3.3}$$

and $\alpha_{\text{N}}^{\text{C}}$ is the solution to

$$\underset{\alpha \geq 0}{\text{minimize}} \;\; m_k^v(n_k^{\text{C}}(\alpha)) \quad \text{subject to} \quad \|P_k^{-1} n_k^{\text{C}}(\alpha)\|_2 \leq \delta_k^v, \quad s_k + n_k^{\text{C}s}(\alpha) \geq \kappa_{\text{fbn}} s_k. \tag{3.4}$$

We show in Lemma 3.5 that the decrease in m_k^v obtained by n_k^{C} is positive. Overall, when (3.2) holds, we require a normal step satisfying the constraints of (2.2), i.e.,

$$\|P_k^{-1} n_k\|_2 \leq \delta_k^v, \quad s_k + n_k^s \geq \kappa_{\text{fbn}} s_k, \tag{3.5}$$

along with [recall (2.14)]

$$\Delta m_k^{v,n} \geq m_k^v(0) - m_k^v(n_k^{\text{C}}) \tag{3.6}$$

and

$$n_k \;\; \text{belonging to the range space of} \;\; P_k^2 J(x_k, s_k)^T. \tag{3.7}$$

It is worthwhile to note that many steps satisfy (3.5)–(3.7) with the simplest being n_k^{C}. The condition (3.7) is automatically guaranteed by Krylov-type methods for minimizing $m_k^v(n)$. For future reference, we also define

$$\alpha_N^* := \arg\min_{\alpha \geq 0} \;\; m_k^v(n_k^{\text{C}}(\alpha)) \;\; \text{and} \;\; n_k^* := n_k^{\text{C}}(\alpha_N^*) \tag{3.8}$$

as the minimizer of the feasibility model along the scaled steepest descent direction (ignoring a trust-region constraint). Note that α_N^* is unique whenever $\pi_k^v > 0$.

3.2 Inexact Lagrange multipliers and tangential steps

In contrast to the preliminary algorithm in Sect. 2—which involved the sequential computation of a Lagrange multiplier and tangential step—the conditions that we enforce for an inexact Lagrange multiplier and a *Cauchy step* for the tangential subproblem are intertwined in our main algorithm. Hence, in this subsection, we consider together the computation of new Lagrange multipliers and the tangential step. (It is important to note that the Lagrange multiplier computation can still be performed independently before the tangential step computation; all that is needed in the multiplier computation is, for each multiplier estimate, information about a corresponding Cauchy step for the tangential subproblem, which can be computed at modest computational cost. To clarify this issue, we provide in Sect. 3.2.3 a summary discussion of our multiplier and tangential step computation.)

We remark that for technical reasons in our global convergence analysis, we require a small change to our definition of the matrix G_k [recall (1.7)] appearing in the barrier function model (2.8). Specifically, we now define

$$G_k := \begin{pmatrix} \nabla_{xx}\mathcal{L}(x_k, s_k, y_k^{\text{B}}) & 0 \\ 0 & D_k \end{pmatrix} \tag{3.9}$$

with y_k^{B} being a (bounded) multiplier vector satisfying, for all $i \in \{1, 2, \ldots, M\}$,

$$[y_k^{\text{B}}]_i > 0 \quad \text{and} \quad \|y_k^{\text{B}}\|_2 \le \kappa_{\text{y}} \quad \text{for some scalar} \quad \kappa_{\text{y}} > 0 \tag{3.10}$$

and D_k being a positive definite (p.d.) diagonal matrix satisfying

$$\|D_k\|_2 \le \kappa_{\text{D}} \quad \text{for some scalar} \quad \kappa_{\text{D}} > 0. \tag{3.11}$$

The key aspect of this definition is to ensure boundedness of the components of G_k, which means that, in fact, one may use an approximate Hessian of the Lagrangian as long as the sequence $\{G_k\}$ is uniformly bounded.

Overall, as is typical in a step decomposition approach, our goal is to compute a tangential step t_k lying (approximately) in the null space of the constraint Jacobian $J(x_k, s_k)$ that satisfies $m_k^f(n_k + t_k) \le m_k^f(n_k)$ while not undoing the predicted gain in linearized feasibility provided by the normal step n_k. On one hand, this latter requirement suggests that improvement in the barrier function should be sought within the trust-region $\{d : \|P_k^{-1}d\|_2 \le \delta_k^v\}$, since it is only within this region that the linearized constraint model is believed to be trustworthy. On the other hand, as a separate consideration we assume that the barrier function model m_k^f may only be trusted within $\{d : \|P_k^{-1}d\|_2 \le \delta_k^f\}$. Overall, to allow flexibility in our algorithm, we simply use as a necessary condition for computing a new Lagrange multiplier estimate and (potentially) a tangential step the inequality

$$\|P_k^{-1} n_k\|_2 \le \kappa_{\rm B} \min\{\kappa_{\rm vf}\delta_k^v, \delta_k^f\} \text{ with } \kappa_{\rm B} \in (0,1) \text{ and } \kappa_{\rm vf} > 0. \tag{3.12}$$

If (3.12) does not hold, then we set $y_k \leftarrow y_{k-1}$ and $t_k \leftarrow 0$.

Overall, the main idea of the strategy in the preceding paragraph is that, if (3.12) does not hold, then (i) improvement toward feasibility may be expected from the normal step alone and (ii) the computation of a tangential step—and hence new Lagrange multipliers for computing a productive tangential step—is unnecessary to ensure convergence. Observe that if one chooses $\kappa_{\rm B}\kappa_{\rm vf} \in (0, 1)$, then (3.12) states that new multipliers and a tangential step need not be computed if the normal step lies on its trust region boundary. We claim that one may still consider computing new multipliers and a tangential step in such a case. However, in order to analyze an algorithm that minimizes per-iteration costs as much as possible, we employ (3.12) as described. Also note that if $\kappa_{\rm B}\kappa_{\rm vf} \ge 1$, then, by (3.5), the inequality (3.12) reduces to $\|P_k^{-1} n_k\|_2 \le \kappa_{\rm B}\delta_k^f$, which suggests that new multipliers and a tangential step need not be computed when the normal step lies outside the region in which the barrier function model is trustworthy.

When (3.12) is satisfied, we first compute a new Lagrange multiplier estimate as an approximate solution of (2.7). For determining whether such a solution is acceptable, we consider first the properties of the vector

$$r_k := r_k(y_k) := P_k^2(\nabla m_k^f(n_k) + J(x_k, s_k)^T y_k), \tag{3.13}$$

with which we define the related "f-criticality" measures

$$\pi_k^f := \pi_k^f(y_k) := \|P_k(\nabla m_k^f(n_k) + J(x_k, s_k)^T y_k)\|_2 \quad \text{and} \tag{3.14a}$$

$$\chi_k^f := \chi_k^f(y_k) := \frac{\nabla m_k^f(n_k)^T r_k(y_k)}{\pi_k^f(y_k)} \tag{3.14b}$$

associated with minimizing f. (As in the discussion leading to (3.3) for the normal subproblem, the vector r_k can be motivated as a means of defining a Cauchy point for the tangential subproblem; see (3.17) and (3.21) later.) We determine that subproblem (2.7) has been solved accurately enough as long as y_k, r_k, π_k^f, and χ_k^f at least satisfy one (if not more) of the following three sets of conditions:

$$\pi_k^f \le \epsilon_\pi \quad \text{and} \quad v_k \le \epsilon_v; \tag{3.15a}$$

$$\pi_k^f \le \omega_t(\pi_k^v); \quad \text{or} \tag{3.15b}$$

$$\chi_k^f \ge \kappa_\chi \pi_k^f. \tag{3.15c}$$

Here, $\{\epsilon_\pi, \epsilon_v\} > 0$ and $\kappa_\chi \in (0, 1)$ are constants and ω_t is a forcing function. We require that the functions ω_n and ω_t [see (3.2) and (3.15b)] satisfy

$$\omega_t(\omega_n(\tau)) \le \kappa_\omega \tau \quad \text{for all} \quad \tau \ge 0 \text{ and for some } \kappa_\omega \in (0, 1). \tag{3.16}$$

With respect to the conditions in (3.15), a few remarks are in order. First, we remark that one can show (see Lemma 3.8) that one can always satisfy one of the three sets of conditions in (3.15), and thus this requirement for y_k (and the related quantities r_k, π_k^f, and χ_k^f) is well-posed. One can also see that if (3.15a) is satisfied, then (x_k, s_k, y_k) is an approximate first-order KKT point for the barrier subproblem for the tolerances $\{\epsilon_\pi, \epsilon_v\} > 0$. (If this condition holds, then, as seen in the formal statement of it at the end of this section, our main algorithm will terminate.) However, if (3.15a) is not satisfied, but (3.15b) holds, then the f-criticality measure π_k^f is insubstantial compared to the v-criticality measure π_k^v. In this case, the computation of a tangential step is skipped, i.e., we simply set $t_k \leftarrow 0$. Otherwise, when (3.15a) and (3.15b) do not hold (and necessarily (3.15c) holds), we decide that we must compute a tangential step. In this case, it follows from the Definition (3.14), the condition (3.15c) and the fact that $\pi_k^f > 0$ (since otherwise (3.15b) would have held) that r_k is a direction of strict ascent for $m_k^f(\cdot)$ at n_k. This property allows us to compute a tangential step t_k satisfying one of two sets of conditions as outlined in the following two subsections. Our choice of which set of conditions to satisfy depends on whether a normal step is computed. Specifically, if $n_k \neq 0$, then we require the computation of what we call a relaxed SQP tangential step. Otherwise, if $n_k = 0$, then we are still free to attempt to compute a relaxed SQP tangential step, but we may instead compute what we call a very relaxed SQP tangential step. In such a case, this latter option may be preferable as it involves a weaker restriction on linearized infeasibility of the step.

3.2.1 A relaxed SQP tangential step

Given a constant κ_{tg} small enough such that $\kappa_{\text{cd}} \in (0, 1 - \kappa_{\text{tg}}] \subset (0, 1)$ [recall that κ_{cd} was defined in (2.15)], a relaxed SQP tangential step is defined as follows.

Definition 3.1 (*Relaxed SQP tangential step*) Define the Cauchy point

$$t_k^{\text{C}} := t_k^{\text{C}}(\alpha_{\text{T}}^{\text{C}}), \quad \text{where} \quad t_k^{\text{C}}(\alpha) := \begin{pmatrix} t_k^{\text{C}x}(\alpha) \\ t_k^{\text{C}s}(\alpha) \end{pmatrix} := -\alpha \begin{pmatrix} r_k^x \\ r_k^s \end{pmatrix} = -\alpha r_k \tag{3.17}$$

and $\alpha_{\text{T}}^{\text{C}}$ is the minimizer of

$$\begin{aligned} &\underset{\alpha \geq 0}{\text{minimize}} \quad m_k^f\big(n_k + t_k^{\text{C}}(\alpha)\big) \\ &\text{subject to} \quad \|P_k^{-1}\big(n_k + t_k^{\text{C}}(\alpha)\big)\|_2 \leq \min\{\kappa_{\text{vf}}\delta_k^v, \delta_k^f\}, \\ &\qquad\qquad\quad s_k + n_k^s + t_k^{\text{C}s}(\alpha) \geq \kappa_{\text{fbt}}(s_k + n_k^s). \end{aligned} \tag{3.18}$$

Then, t_k is a relaxed SQP tangential step if

$$\Delta m_k^{f,t} \geq m_k^f(n_k) - m_k^f(n_k + t_k^{\text{C}}), \tag{3.19a}$$

$$s_k + n_k^s + t_k^s \geq \kappa_{\text{fbt}}(s_k + n_k^s), \tag{3.19b}$$

$$\|P_k^{-1}(n_k + t_k)\|_2 \le \min\{\kappa_{\text{vf}}\delta_k^v, \delta_k^f\}, \quad \text{and} \tag{3.19c}$$

$$m_k^v(n_k + t_k) \le \kappa_{\text{tg}} m_k^v(0) + (1 - \kappa_{\text{tg}}) m_k^v(n_k). \tag{3.19d}$$

Condition (3.19a) ensures that the model of the barrier function is decreased at least as much as by the Cauchy point t_k^{C}, (3.19b) is a fraction-to-the-boundary constraint, (3.19c) is a trust-region constraint, and (3.19d) is a relaxation of the traditional SQP constraint that $c(x_k, s_k) + J(x_k, s_k)(n_k + t_k) = 0$ that ensures that linearized constraint infeasibility is sufficiently reduced.

If a relaxed SQP tangential step satisfying (3.19) is computed, then we must evaluate its usefulness in the sense that we must ensure that a relatively large tangential step results in a sufficient decrease in the model m_k^f of the barrier function. With this in mind, we check whether the conditions

$$\|P_k^{-1} t_k\|_2 > \kappa_{\text{tn}} \|P_k^{-1} n_k\|_2 \quad \text{for some} \quad \kappa_{\text{tn}} > 1 \tag{3.20}$$

and (2.10) are satisfied. The inequality (2.10) indicates that the predicted decrease in the barrier function obtained from the tangential step is substantial compared to any potential increase resulting from the normal step. If the step t_k satisfies (3.20) but violates (2.10), it does not serve its role and we reset it to zero.

3.2.2 A very relaxed SQP tangential step

Condition (3.19) may be too restrictive in certain cases. Specifically, if $v_k = 0$, then the algorithm will set $n_k \leftarrow 0$, from which it follows that (3.19d) requires t_k to be in the null space of $J(x_k, s_k)$. This is an unreasonable requirement in matrix-free settings; indeed (3.19d) may be unreasonable in any situation when the normal step computation is skipped and $n_k \leftarrow 0$. Thus, to avoid such a requirement, we allow for the computation of an alternative tangential step. Given the constant $\kappa_{\text{fbt}} \in (0, 1)$ employed in (3.19b), a constant $\kappa_{\text{v}} \in (1, \infty)$, and a constant $\kappa_{\text{tt}} \in (\kappa_{\text{vv}}, 1)$ (with $\kappa_{\text{vv}} \in (0, 1)$ defined for (3.2)), the salient feature of our alternative is that it involves a relaxed condition on the linearized infeasibility of the step. We emphasize that we are only allowed to compute a tangential step of this type when $n_k = 0$, though we incorporate n_k into the conditions in the following definition so that one may more easily compare them to the conditions in Definition 3.1.

Definition 3.2 (*Very relaxed SQP tangential step*) Define the Cauchy point

$$t_k^{\text{C}} = t_k^{\text{C}}(\alpha_{\text{T}}^{\text{C}}), \quad \text{where} \quad t_k^{\text{C}}(\alpha) := \begin{pmatrix} t_k^{\text{C}x}(\alpha) \\ t_k^{\text{C}s}(\alpha) \end{pmatrix} := -\alpha \begin{pmatrix} r_k^x \\ r_k^s \end{pmatrix} = -\alpha r_k \tag{3.21}$$

and $\alpha_{\text{T}}^{\text{C}}$ is the minimizer of

$$\begin{aligned} \underset{\alpha \ge 0}{\text{minimize}} \quad & m_k^f\left(n_k + t_k^{\text{C}}(\alpha)\right) \\ \text{subject to} \quad & \|P_k^{-1}\left(n_k + t_k^{\text{C}}(\alpha)\right)\|_2 \le \min\{\kappa_{\text{vf}}\delta_k^v, \delta_k^f, \kappa_{\text{v}} v_k^{\max}\}, \\ & s_k + n_k^s + t_k^{\text{C}s}(\alpha) \ge \kappa_{\text{fbt}}(s_k + n_k^s). \end{aligned} \tag{3.22}$$

Then, t_k is a very relaxed SQP tangential step if

$$\Delta m_k^{f,t} \geq m_k^f(n_k) - m_k^f(n_k + t_k^{\mathrm{C}}), \tag{3.23a}$$

$$s_k + n_k^s + t_k^s \geq \kappa_{\mathrm{fbt}}(s_k + n_k^s), \tag{3.23b}$$

$$\|P_k^{-1}(n_k + t_k)\|_2 \leq \min\{\kappa_{\mathrm{vf}}\delta_k^v, \delta_k^f, \kappa_{\mathrm{v}} v_k^{\max}\}, \quad \text{and} \tag{3.23c}$$

$$m_k^v(n_k + t_k) \leq \kappa_{\mathrm{tt}} v_k^{\max}. \tag{3.23d}$$

Conditions (3.23a)–(3.23c) play the same role as conditions (3.19a)–(3.19c). However, since the Cauchy point defined by (3.21)–(3.22) involves a potentially smaller trust-region radius than that defined in (3.18), the bound imposed in (3.23a) may be different from that imposed in (3.19a), and this difference in the trust-region radii is matched in (3.23c) [see (3.19c)]. The name "very relaxed SQP tangential step" has been chosen because of condition (3.23d), which merely requires that the predicted constraint violation be sufficiently less than a fraction of the upper bound $v_k^{\max}$ rather than a fraction of the current violation [see (3.19d)]. In fact, the potentially smaller trust-region radii in (3.22) and (3.23c) (as compared to those in (3.18) and (3.19c)) have been chosen to compensate for this relaxation.

3.2.3 Summary of inexact Lagrange multiplier and tangential step computation

Overall, the Lagrange multiplier and tangential step computation may proceed as follows. First, an iterative solver may be applied to the least-squares subproblem (2.7) until an approximate solution y_k and the corresponding r_k, π_k^f, and χ_k^f satisfy at least one of (3.15a), (3.15b), or (3.15c). If (3.15a) or (3.15b) is satisfied, then the algorithm may proceed with y_k as the new multiplier estimate. Otherwise, if only (3.15c) holds, then one should check whether the Cauchy step defined by (3.17)–(3.18) satisfies (3.19) or (if $n_k = 0$) the Cauchy step defined by (3.21)–(3.22) satisfies (3.23). (In fact, by construction of the Cauchy steps, one need only check (3.19d) in the former case and (3.23d) in the latter case since all other conditions in (3.19) and (3.23) are guaranteed to hold by definition of the corresponding Cauchy steps.) If either Cauchy step satisfies its corresponding set of conditions (with $n_k = 0$ required in the latter case), then the algorithm may proceed with y_k as the new multiplier estimate. Otherwise, the iterative solver for (2.7) should be continued until the above strategy yields an acceptable new multiplier y_k. Once a new multiplier estimate is obtained in this manner, the algorithm may proceed to compute a tangential step satisfying (3.19) or (if $n_k = 0$) (3.23). This latter computation is well-defined as the strategy for computing y_k has at least guaranteed that a corresponding Cauchy point satisfies the required conditions. (Indeed, under reasonable assumptions on the iterative solver for (2.7), this entire strategy for computing y_k and t_k is well-posed; see Lemma 3.8.

3.3 Iteration types, step acceptance, and updating strategies

Our inexact method uses the same iteration types as our preliminary algorithm in Sect. 2. In this section, we give the precise updates that we use for the iterates, the trust-region radii, and the funnel radius for the three types of iterations.

First, consider y-iterations as in Definition 2.1, which occur when n_k and t_k are both zero, but could also (presumably) occur if $n_k = -t_k$ and some components are nonzero. (In fact, this latter case is ruled out by Lemma 3.3(vi).) During a y-iteration, we perform—as in Algorithm 1—the updates

$$(x_{k+1}, s_{k+1}) \leftarrow (x_k, s_k), \quad \delta^f_{k+1} \leftarrow \delta^f_k, \quad \delta^v_{k+1} \leftarrow \delta^v_k, \quad \text{and} \quad v^{\max}_{k+1} \leftarrow v^{\max}_k. \tag{3.24}$$

As previously mentioned, since a y-iteration is defined by a zero primal step, the only computation of interest is that of a new vector of Lagrange multiplier estimates. Therefore, the updates in (3.24) leave the trust-region radii and bound on the maximum allowed infeasibility unchanged for the subsequent iteration.

Second, consider f-iterations as in Definition 2.2, which have the primary purpose of reducing the barrier function [recall (2.10)] while ensuring that the constraint violation remains within the funnel radius [recall (2.11)]. If the kth iteration is an f-iteration and $\rho^f_k \geq \eta_1$ [recall (2.12)], then we set

$$(x_{k+1}, s_{k+1}) \leftarrow (x^+_k, s^+_k) \tag{3.25}$$

$$[s_{k+1}]_i \leftarrow \begin{cases} [s_{k+1}]_i & \text{if} \quad [c(x_{k+1}, s_{k+1})]_i \geq 0, \\ -[c(x_{k+1})]_i & \text{otherwise,} \end{cases} \tag{3.26}$$

$$\delta^f_{k+1} \in \begin{cases} [\delta^f_k, \infty) & \text{if} \quad \rho^f_k \geq \eta_2, \\ [\gamma_2 \delta^f_k, \delta^f_k] & \text{otherwise,} \end{cases} \tag{3.27}$$

$$\delta^v_{k+1} \in [\delta^v_k, \infty). \tag{3.28}$$

Otherwise (i.e., if $\rho^f_k < \eta_1$), we set

$$(x_{k+1}, s_{k+1}) \leftarrow (x_k, s_k), \quad \delta^f_{k+1} \in [\gamma_1 \delta^f_k, \gamma_2 \delta^f_k], \quad \text{and} \quad \delta^v_{k+1} \leftarrow \delta^v_k. \tag{3.29}$$

In both cases, we set

$$v^{\max}_{k+1} \leftarrow v^{\max}_k. \tag{3.30}$$

In (3.25)–(3.30), the constants should be chosen to satisfy $0 < \eta_1 \leq \eta_2 < 1$ and $0 < \gamma_1 \leq \gamma_2 < 1$. Overall, we accept the trial point (x^+_k, s^+_k) if the achieved decrease in the barrier function is comparable to the predicted decrease (and reject it otherwise), update δ^f_{k+1} using a typical trust-region updating strategy, possibly increase the normal step trust-region radius, and leave the funnel radius unchanged.

For technical reasons, after a f-iteration in which the trial point is accepted, we reset the size of the normal step trust region radius during the next iteration in which a normal step is computed. Specifically, if a normal step is computed during iteration k and the last successful iteration was an f-iteration, we enforce

$$\delta_k^v \geq \kappa_{\rm n} \| P_k^{-1} n_k^* \|_2 \quad \text{for some} \quad \kappa_{\rm n} > 0, \tag{3.31}$$

where n_k^* is given by (3.8). Besides being needed for our convergence analysis, this safeguard is practical in that a (sequence of) f-iteration(s) with $\rho_k^f \geq \eta_1$ may make inaccurate the information on the adequacy of the model $m_k^v(\cdot)$ and trust region radius δ_k^v gathered during previous iterations.

Third, consider v-iterations as in Definition 2.3, which have as their main goal an improvement toward feasibility. If $\rho_k^v \geq \eta_1$ and (2.15) holds, then we set

$$(x_{k+1}, s_{k+1}) \leftarrow (x_k^+, s_k^+) \tag{3.32}$$

$$[s_{k+1}]_i \leftarrow \begin{cases} [s_{k+1}]_i & \text{if} \quad [c(x_{k+1}, s_{k+1})]_i \geq 0, \\ -[c(x_{k+1})]_i & \text{otherwise,} \end{cases} \tag{3.33}$$

$$\delta_{k+1}^v \in \begin{cases} [\delta_k^v, \infty) & \text{if} \quad \rho_k^v \geq \eta_2, \\ \delta_k^v & \text{otherwise,} \end{cases} \tag{3.34}$$

$$v_{k+1}^{\max} \leftarrow \max\{\kappa_{\rm t1} v_k^{\max}, v(x_{k+1}, s_{k+1}) + \kappa_{\rm t2}\big(v_k - v(x_{k+1}, s_{k+1})\big)\}. \tag{3.35}$$

Otherwise (i.e., if $\rho_k^v < \eta_1$ or (2.15) does not hold), we set

$$(x_{k+1}, s_{k+1}) \leftarrow (x_k, s_k), \quad \delta_{k+1}^v \in [\gamma_1 \delta_k^v, \gamma_2 \delta_k^v], \quad \text{and} \quad v_{k+1}^{\max} \leftarrow v_k^{\max}. \tag{3.36}$$

In both cases, we set

$$\delta_{k+1}^f \leftarrow \delta_k^f. \tag{3.37}$$

In (3.32)–(3.37), the constants should be chosen to satisfy $\{\kappa_{\rm t1}, \kappa_{\rm t2}\} \subset (0, 1)$. In this manner, the trial point is accepted if the normal step is nonzero and the improvement in linearized feasibility is comparable to its predicted value, which is itself comparable to the improvement yielded by the normal step.

3.4 A trust-funnel algorithm

We are now prepared to state our trust-funnel method for solving (BSP). For convenience, we define sets that classify each iteration, as well as the computations performed in them. The first group of sets distinguishes between iteration types:

$$\mathcal{Y} := \{k \in \mathbb{N} : d_k = 0\}; \ \mathcal{F} := \{k \in \mathbb{N} : t_k \neq 0 \text{ and (2.10)-(2.11) hold}\};$$
$$\mathcal{V} := \mathbb{N} \backslash (\mathcal{Y} \cup \mathcal{F}).$$

(Lemma 3.3 below shows that these sets are mutually exclusive and exhaustive.) The second group distinguishes iterations for which the normal and/or tangential steps satisfy various conditions, and whether the tangential step was reset to zero:

$$\mathcal{N} := \{k \in \mathbb{N} : n_k \text{ was computed to satisfy (3.5)-(3.7)}\};$$
$$\mathcal{T} := \{k \in \mathbb{N} : t_k \text{ was computed to satisfy either (3.19) or (3.23)}\};$$

$$\mathcal{T}_{\mathcal{D}} := \{k \in \mathcal{T} : \text{the computed } t_k \text{ satisfied (3.19)}\};$$
$$\mathcal{T}_0 := \{k \in \mathcal{T}_{\mathcal{D}} : \text{the computed } t_k \text{satisfied (3.19) and (3.20), but not (2.10)}\};$$

(Note that t_k is reset to zero for $k \in \mathcal{T}_0$.) Furthermore, the set of iterations for which d_k satisfies the linearized constraint contraction condition (3.19d) plays an important role in our analysis; thus, along with the sets above, we define

$$\mathcal{D} := \{k \in \mathbb{N} : \text{the step } d_k = n_k + t_k \text{ satisfies (3.19d)}\}.$$

Our last group of sets distinguishes iterations that produce a change in the primal space. In particular, if $\rho_k^f \geq \eta_1$ holds during an f-iteration, or if (2.15) holds and $\rho_k^v \geq \eta_1$ during a v-iteration, then iteration k is called *successful*. The following sets capture the types of successful iterations:

$$\mathcal{S}_f := \{k \in \mathcal{F} : \rho_k^f \geq \eta_1\}; \quad \mathcal{S}_v := \{k \in \mathcal{V} : \text{(2.15) holds and } \rho_k^v \geq \eta_1\}; \quad \mathcal{S} := \mathcal{S}_f \cup \mathcal{S}_v.$$

Finally, for convenience when referring to the trust-region radius for the tangential subproblem (see (3.19c) and (3.23c)), we define $\delta_{-1}^t := 1$ and, for $k \geq 0$,

$$\delta_k^t := \begin{cases} \delta_{k-1}^t & \text{if } \ k \notin \mathcal{T}, \\ \min\{\kappa_{\text{vf}} \delta_k^v, \delta_k^f\} & \text{if } \ k \in \mathcal{T} \cap \mathcal{T}_{\mathcal{D}}, \\ \min\{\kappa_{\text{vf}} \delta_k^v, \delta_k^f, \kappa_{\text{v}} v_k^{\max}\} & \text{if } \ k \in \mathcal{T} \backslash \mathcal{T}_{\mathcal{D}}. \end{cases} \tag{3.38}$$

We formally state our trust-funnel method as Algorithm 2 on page 19, and provide an informal flow diagram in the "Appendix" on page 56.

As a guide for the reader with respect to the salient properties of the various types of iterations we have defined, we provide the following lemma regarding basic facts that may be deduced from the design of our algorithm. Unless stated otherwise, reference to the tangential step t_k corresponds to the value used in Step 37 of Algorithm 2, i.e., the value after the possible reset in Step 31. For the purposes of this lemma, we assume that if the algorithm does not terminate during iteration k, then all steps of the algorithm during the iteration are well-defined. We prove this fact formally in the next subsection.

Lemma 3.3 *If Algorithm 2 does not terminate during the kth iteration, then:*

(i) *If* $k \in \mathcal{N}$, *then* $\chi_k^v > 0$, $\pi_k^v > 0$, $m_k^v(0) - m_k^v(n_k^{\text{C}}) > 0$, $\Delta m_k^{v,n} > 0$, *and* $n_k \neq 0$.
(ii) *If* $n_k \neq 0$, *then* $k \in \mathcal{N}$.
(iii) *If* $k \in \mathcal{T}$, *then* $\chi_k^f \geq \kappa_\chi \pi_k^f > 0$ *and* $m_k^f(n_k) - m_k^f(n_k + t_k^{\text{C}}) > 0$.
(iv) *If* $k \in \mathcal{T} \backslash \mathcal{T}_0$, *then* $t_k \neq 0$ *and* $\Delta m_k^{f,t} > 0$. *If* $k \in \mathcal{T}_0$, *then* $t_k = 0$ *and* (3.12) *holds.*
(v) *If* $t_k \neq 0$, *then* $k \in \mathcal{T} \backslash \mathcal{T}_0$.
(vi) $k \in \mathcal{Y}$ *if and only if* $n_k = t_k = 0$.
(vii) *If* $k \in \mathcal{Y}$, *then* $k \in \mathcal{D}$ *and* $\pi_k^f \leq \kappa_\omega \pi_{k-1}^f$ *with* $\kappa_\omega \in (0, 1)$ *defined as in* (3.16).
(viii) *If* $k \notin \mathcal{D}$, *then* $k \in \mathcal{T} \backslash \mathcal{T}_{\mathcal{D}}$ *and* (3.23) *holds.*

Algorithm 2 Trust-funnel algorithm for the barrier subproblem (BSP)

Input: $(x_0, s_0, y_{-1}, \mu, \epsilon_\pi, \epsilon_v)$ with $(s_0, y_{-1}, \mu, \epsilon_\pi, \epsilon_v) > 0$.
Choose $\{\delta_0^f, \delta_0^v, \kappa_{\rm vf}, \kappa_{\rm ca}, \kappa_{\rm y}, \kappa_{\rm D}, \kappa_{\rm n}\} \subset (0, \infty)$, $\{\kappa_{\rm cr}, \kappa_{\rm tn}, \kappa_{\rm v}\} \subset (1, \infty)$, $0 < \eta_1 \le \eta_2 < 1$, $0 < \gamma_1 \le \gamma_2 < 1$, $\{\kappa_{\rm tt}, \kappa_\delta, \kappa_{\rm tg}, \kappa_\omega, \kappa_\chi, \kappa_{\rm B}, \kappa_{\rm vv}, \kappa_{\rm fbn}, \kappa_{\rm fbt}, \kappa_{\rm t1}, \kappa_{\rm t2}\} \subset (0, 1)$, and $\kappa_{\rm cd} \in (0, 1 - \kappa_{\rm tg}]$.
Perform a slack reset to s_0 as given by (1.2).
Set $v_0^{\max} \leftarrow \max\{\kappa_{\rm ca}, \kappa_{\rm cr} v(x_0, s_0)\}$ and $\pi_{-1}^f \leftarrow 0$.
Set $\mathcal{S}_f$`-flag` $\leftarrow$ `false`.
for $k = 0, 1, \ldots$ **do**
  Compute v_k from (2.1) and π_k^v and χ_k^v from (3.1).
  if $v_k > 0$ and $\chi_k^v = 0$ **then**
    Return the infeasible stationary point (x_k, s_k).
  Normal Step Computation
  if (3.2) holds, or at least $\pi_k^v > 0$ **then** $[k \in \mathcal{N}]$
    if $k \ge 1$ and $\mathcal{S}_f$`-flag` $=$ `true` **then**
      Compute n_k^* satisfying (3.8).
      Set $\delta_k^v \leftarrow \max\{\delta_k^v, \kappa_{\rm n}\|P_k^{-1} n_k^*\|_2\}$ and $\mathcal{S}_f$`-flag` $\leftarrow$ `false`.
    Compute n_k satisfying (3.5)–(3.7).
  else
    Set $n_k \leftarrow 0$.
  Lagrange Multiplier and Tangential Step Computation
  Choose $y_k^{\rm B}$ and p.d. diagonal D_k satisfying (3.10)–(3.11), then set G_k by (3.9).
  if (3.12) holds **then**
    Compute y_k, r_k, π_k^f, and χ_k^f from (2.7) and (3.13)–(3.14) by the strategy in §3.2.3.
    if (3.15a) holds **then**
      Return the (approximate) first-order KKT point (x_k, s_k, y_k).
    else if (3.15b) holds **then**
      Set $t_k \leftarrow 0$.
    else $[k \in \mathcal{T}]$
      if $k \in \mathcal{N}$ **then**
        Compute t_k so that (3.19) is satisfied.
      else
        Compute t_k so that either (3.19) or (3.23) is satisfied.
      if (3.19) holds **then** $[k \in \mathcal{T}_\mathcal{D}]$
        if (3.20) is satisfied but (2.10) fails **then** $[k \in \mathcal{T}_0]$
          Set $t_k \leftarrow 0$.
  else
    Set $y_k \leftarrow y_{k-1}$ and $t_k \leftarrow 0$, then set r_k, π_k^f, and χ_k^f by (3.13)–(3.14).
    if (3.15a) holds **then**
      Return the (approximate) first-order KKT point (x_k, s_k, y_k).
  Iteration Type and Step Acceptance Determination
  (**if** (3.19d) holds **then** add k to the set $\mathcal{D}$.) $[k \in \mathcal{D}]$
  Set the trial step $d_k \leftarrow n_k + t_k$ and trial iterate $(x_k^+, s_k^+) \leftarrow (x_k, s_k) + d_k$.
  if $d_k = 0$ **then** $[k \in \mathcal{Y}]$
    Perform the y-iteration updates given by (3.24).
  else if $t_k \ne 0$ and both (2.10) and (2.11) hold **then** $[k \in \mathcal{F}]$
    if $\rho_k^f \ge \eta_1$ **then** $[k \in \mathcal{S}_f]$
      Perform the successful f-iteration updates given by (3.25)–(3.28) and (3.30).
      Set $\mathcal{S}_f$`-flag` $\leftarrow$ `true`.
    else
      Perform the unsuccessful f-iteration updates given by (3.29) and (3.30).
  else $[k \in \mathcal{V}]$
    if $\rho_k^v \ge \eta_1$ and (2.15) holds **then** $[k \in \mathcal{S}_v]$
      Perform the successful v-iteration updates given by (3.32)–(3.35) and (3.37).
    else
      Perform the unsuccessful v-iteration updates given by (3.36) and (3.37).

(ix) *If $k \in \mathcal{D}$, then the inequality in (2.15) holds.*
(x) $\mathcal{T}_\mathcal{D} \subseteq \mathcal{D}$.
(xi) *If $k \in \mathcal{T} \backslash \mathcal{T}_\mathcal{D}$, then $n_k = 0$ and $k \notin \mathcal{N}$.*

Proof To prove part (i), let $k \in \mathcal{N}$, in which case we have that the conditions in Step 10 held true. This could occur only if $\pi_k^v > 0$, or if in (3.2) we had $\pi_k^v > \omega_n(\pi_{k-1}^f) \geq 0$ or $v_k \geq \kappa_{vv} v_k^{\max}$. Thus, to prove that $k \in \mathcal{N}$ implies $\pi_k^v > 0$, all that remains is to investigate the case when $v_k \geq \kappa_{vv} v_k^{\max}$. Since $v_k^{\max} > 0$ by construction, this inequality implies $v_k > 0$. If $\pi_k^v = 0$ (which, since $v_k > 0$, implies $\chi_k^v = 0$), then the algorithm would have terminated in Step 9. Thus, we may again conclude that $\pi_k^v > 0$, which establishes this strict inequality for all $k \in \mathcal{N}$. In turn, by (3.1) and the fact that $v_k > 0$ when $\pi_k^v > 0$, we must have $\chi_k^v > 0$ for all $k \in \mathcal{N}$. Now, since $\pi_k^v > 0$, it follows that $-P_k^2 J(x_k, s_k)^T c(x_k, s_k)$ is a direction of strict decrease for m_k^v at $n = 0$, from which it follows by (3.3) that $m_k^v(0) - m_k^v(n_k^c) > 0$. In turn, (3.6) implies the remainder of part (i).

Part (ii) follows since if $n_k \neq 0$, then the conditions in Step 10 must have held (or else the algorithm would have set $n_k \leftarrow 0$), in which case $k \in \mathcal{N}$.

Next, we prove part (iii). If $k \in \mathcal{T}$, then it follows from Steps 19–28 of the algorithm that after the computation of y_k (and all dependent quantities) both (3.15a) and (3.15b) did not hold (implying that $\pi_k^f > 0$), but (3.15c) did. Combining (3.15c) and the fact that $\pi_k^f > 0$ yields $\nabla m_k^f(n_k)^T r_k \geq \kappa_\chi (\pi_k^f)^2 > 0$ (as desired), which implies that r_k is a direction of strict ascent for m_k^f at n_k. Combining this fact with (3.17)–(3.18) and (3.21)–(3.22) yields $m_k^f(n_k) - m_k^f(n_k + t_k^c) > 0$.

Building on the proof of part (iii), we next prove part (iv). If we have $k \in \mathcal{T} \backslash \mathcal{T}_0$, we may combine $m_k^f(n_k) - m_k^f(n_k + t_k^c) > 0$ with (3.19a)/(3.23a) to conclude that $t_k \neq 0$ and $\Delta m_k^{f,t} > 0$, as desired. (Since $k \notin \mathcal{T}_0$, this tangential step was not reset to zero, so we have maintained $t_k \neq 0$ in Step 37.) If $k \in \mathcal{T}_0$, it follows from Steps 18–31 that (3.12) holds, but that the algorithm reset $t_k \leftarrow 0$.

To prove part (v), we first note that if $t_k \neq 0$, then a tangential step was computed and thus $k \in \mathcal{T}$. Moreover, since $t_k \neq 0$, we know that $k \notin \mathcal{T}_0$, which means $k \in \mathcal{T} \backslash \mathcal{T}_0$, as desired.

We now prove part (vi). If $n_k = t_k = 0$, then $d_k = 0$ and we have $k \in \mathcal{Y}$ by the definition of $\mathcal{Y}$; this proves one direction. For the other direction, in order to derive a contradiction, suppose that $k \in \mathcal{Y}$ (so that $d_k = n_k + t_k = 0$), but that $n_k \neq 0$ and/or $t_k \neq 0$. Indeed, since $n_k + t_k = 0$, we must have $n_k \neq 0$ and $t_k \neq 0$. It then follows from parts (ii) and (v) that $k \in \mathcal{Y} \cap \mathcal{N} \cap (\mathcal{T} \backslash \mathcal{T}_0)$. Consequently, from part (i) we have that $m_k^v(0) > m_k^v(n_k)$. This fact and the equation $n_k + t_k = 0$ imply that (3.19d) must not be satisfied. However, according to Steps 25–26 of the algorithm, since $k \in \mathcal{N}$ we compute t_k to satisfy (3.19), a contradiction.

To prove part (vii), suppose $k \in \mathcal{Y}$. It follows from part (vi) that $n_k = t_k = 0$ so that (3.19d) holds (which means $k \in \mathcal{D}$, as desired), and then from part (i) that $k \notin \mathcal{N}$. Hence, from Step 10 of the algorithm, it follows that (3.2) must be violated. Moreover, since $n_k = 0$, we also know that (3.12) holds and thus an oblique projected gradient r_k was computed (as stipulated in Step 19) to satisfy at least one of (3.15a), (3.15b) and (3.15c). In fact, under the conditions of this lemma, it follows that (3.15a) must not have held, so we know that either (3.15b) or (3.15c) is satisfied as a result of this calculation. Suppose that (3.15c) holds so that the algorithm would have proceeded to compute a tangential step and $k \in \mathcal{T}$. If $k \notin \mathcal{T}_0$, then it would follow from part (iv) that $t_k \neq 0$, which by part (vi) contradicts the fact that $k \in \mathcal{Y}$. Thus, we must have $k \in \mathcal{T}_0$, i.e., we reset $t_k \leftarrow 0$ because the computed tangential step satisfied (3.20), but not (2.10). This is a contradiction because (2.10) would have been satisfied trivially since $n_k = 0$. Thus (3.15c) must not hold, which implies that (3.15b) must hold. Since we have shown that (3.15b) holds and (3.2) does not hold, we conclude that $\pi_k^f \le \omega_t(\pi_k^v) \le \omega_t(\omega_n(\pi_{k-1}^f)) \le \kappa_\omega \pi_{k-1}^f$, where we have used the monotonicity of ω_t and (3.16).

To establish part (viii), let $k \notin \mathcal{D}$. It follows from part (vii) that $k \notin \mathcal{Y}$. Now, suppose that $t_k = 0$. Combining this with the fact that $k \notin \mathcal{Y}$ implies from part (vi) that $n_k \neq 0$, which may then be combined with part (ii) to deduce that $k \in \mathcal{N}$. This fact along with part (i) and the fact that $t_k = 0$ implies that $m_k^v(n_k + t_k) \le \kappa_{\text{tg}} m_k^v(0) + (1 - \kappa_{\text{tg}}) m_k^v(n_k)$ (see (3.19d)), and hence $k \in \mathcal{D}$, which is a contradiction. Therefore, we must have $t_k \neq 0$, which from part (v) implies that $k \in \mathcal{T} \backslash \mathcal{T}_0$ and that the computed tangential step was not reset to zero. Thus, t_k satisfies either (3.19) or (3.23). In fact, since $k \notin \mathcal{D}$ so that (3.19d) is not satisfied, we conclude that $k \notin \mathcal{T}_{\mathcal{D}}$ and (3.23) must be satisfied.

To prove part (ix), suppose $k \in \mathcal{D}$ so that (3.19d) holds. It follows that

$$\begin{aligned} \Delta m_k^{v,d} &= m_k^v(0) - m_k^v(d_k) \ge m_k^v(0) - \kappa_{\text{tg}} m_k^v(0) - (1 - \kappa_{\text{tg}}) m_k^v(n_k) \\ &= (1 - \kappa_{\text{tg}})\big(m_k^v(0) - m_k^v(n_k)\big) = (1 - \kappa_{\text{tg}}) \Delta m_k^{v,n}, \end{aligned} \tag{3.39}$$

which, since $\kappa_{\text{cd}} \in (0, 1 - \kappa_{\text{tg}}]$, means that the inequality in (2.15) holds, as desired.

To prove (x), let $k \in \mathcal{T}_{\mathcal{D}}$. It follows that a relaxed SQP tangential step t_k was computed to satisfy (3.19). Thus, if t_k is not reset to zero, we know that (3.19d) holds. However, if t_k was reset to zero, then (3.19d) holds trivially when $n_k = 0$ and from parts (i) and (ii) when $n_k \neq 0$. We have shown in all cases that (3.19d) holds, and therefore $k \in \mathcal{D}$.

Finally, to prove part (xi), let $k \in \mathcal{T} \backslash \mathcal{T}_{\mathcal{D}}$. By Steps 25–31, it follows that (3.23) holds and $k \notin \mathcal{N}$ for all $k \in \mathcal{T} \backslash \mathcal{T}_{\mathcal{D}}$. It then follows from part (ii) that $n_k = 0$. □

3.5 Well-posedness

The purpose of this section is to prove that Algorithm 2 is well-posed in the sense that if iteration k is reached, then, in a reasonable implementation of the algorithm, all computations within iteration k will terminate finitely.

Our first result shows important consequences of the slack reset procedure.

Lemma 3.4 *The slack reset* (3.26) *and* (3.33) *in Steps 42 and 48 yields s_k such that (x_k, s_k) satisfies $s_k > 0$ and $c(x_k, s_k) \ge 0$.*

Proof The fact that $s_k > 0$ follows from the choice $s_0 > 0$, the fact that the slack reset (3.26) and (3.33) only possibly increases the slack variables (as shown in (1.4)), and the fact that the fraction-to-the-boundary rules in (3.5) and (3.19b)/(3.23b) hold when normal and tangential steps are computed.

We now prove that $c(x_k, s_k) \ge 0$ holds. Prior to the slack reset performed in Steps 42 and 48, if $[c(x_k, s_k)]_i \ge 0$, then (3.26) and (3.33) leave $[s_k]_i$ unchanged so that $[c(x_k, s_k)]_i \ge 0$ still holds. Otherwise, if $[c(x_k, s_k)]_i < 0$, then after the slack reset (3.26)/(3.33) we have that $[c(x_k) + s_k]_i = 0$, which completes the proof. □

We now show that the Cauchy step for the normal step problem is well-posed.

Lemma 3.5 *If $k \in \mathcal{N}$, then n_k^{C} defined by (3.3)–(3.4) is computed and satisfies*

$$m_k^v(0) - m_k^v(n_k^{\text{C}}) \geq \kappa_k^{\text{cn}} \chi_k^v \min\left\{\pi_k^v, \delta_k^v, 1 - \kappa_{\text{fbn}}\right\} > 0, \tag{3.40}$$

where

$$\kappa_k^{\text{cn}} := \frac{1}{1 + \|J(x_k, s_k)P_k\|_2^2} \in (0, 1]. \tag{3.41}$$

Proof Since $k \in \mathcal{N}$, we may observe from Lemma 3.3(i) that $\pi_k^v > 0$ and $\chi_k^v > 0$, and hence $v_k > 0$. We now show that $n_k^{\text{C}}(\alpha)$ [recall (3.3)] is feasible for (3.4) when

$$k \in \mathcal{N} \quad \text{and} \quad 0 \leq \alpha \leq \frac{1}{\pi_k^v} \min\{\delta_k^v, (1 - \kappa_{\text{fbn}})\} =: \alpha_{\text{B}}. \tag{3.42}$$

Consider any $\alpha \in [0, \alpha_{\text{B}}]$. It follows from the definitions of $n_k^{\text{C}}(\alpha)$ and π_k^v that

$$\|P_k^{-1} n_k^{\text{C}}(\alpha)\|_2 = \|\alpha P_k J(x_k, s_k)^T c(x_k, s_k)\|_2 = \alpha \pi_k^v \leq \delta_k^v.$$

It also follows from the definition of $n_k^{\text{C}s}(\alpha)$ and Lemma 3.4 that

$$\begin{aligned}
[-n_k^{\text{C}s}(\alpha)]_i &= \alpha [S_k]_{ii}^2 [c(x_k, s_k)]_i \leq \alpha [s_k]_i \|P_k J(x_k, s_k)^T c(x_k, s_k)\|_2 \\
&= \alpha \pi_k^v [s_k]_i \leq (1 - \kappa_{\text{fbn}})[s_k]_i \quad \text{for} \quad i = 1, 2, \ldots M,
\end{aligned}$$

so $s_k + n_k^{\text{C}s}(\alpha) \geq \kappa_{\text{fbn}} s_k$. Overall, $n_k^{\text{C}}(\alpha)$ is feasible for (2.2) for all $\alpha \in [0, \alpha_{\text{B}}]$.

Now, observe that $\alpha_{\text{N}}^{\text{C}}$ [recall (3.4)] yields $m_k^v(n_k^{\text{C}}) = m_k^v(n_k^{\text{C}}(\alpha_{\text{N}}^{\text{C}})) \leq m_k^v(n_k^{\text{C}}(\alpha))$ for all $\alpha \in [0, \alpha_{\text{B}}]$. It then follows from [3, Lemma 1] with the quantities

$$\text{“t”} := \alpha_{\text{B}}, \quad \text{“a”} := 2\|J(x_k, s_k)P_k^2 J(x_k, s_k)^T c(x_k, s_k)\|_2^2, \quad \text{“b”} := 2(\pi_k^v)^2 > 0,$$

the fact that

$$\text{“a”} \leq 2\|J(x_k, s_k)P_k\|_2^2 \|P_k J(x_k, s_k)^T c(x_k, s_k)\|_2^2 = 2\|J(x_k, s_k)P_k\|_2^2 (\pi_k^v)^2$$

and the definition of π_k^v that

$$\begin{aligned}
&\left(m_k^v(0)\right)^2 - \left(m_k^v(n_k^{\text{C}})\right)^2 \\
&\quad \geq \text{“b”} \min\left\{\frac{\text{“b”}}{\text{“a”}}, \text{“t”}\right\} \\
&\quad \geq 2(\pi_k^v)^2 \min\left\{\frac{1}{\|J(x_k, s_k)P_k\|_2^2}, \frac{\delta_k^v}{\pi_k^v}, \frac{1 - \kappa_{\text{fbn}}}{\pi_k^v}\right\} \\
&\quad \geq 2\pi_k^v \min\left\{\frac{\pi_k^v}{1 + \|J(x_k, s_k)P_k\|_2^2}, \delta_k^v, 1 - \kappa_{\text{fbn}}\right\} \\
&\quad = 2 v_k \chi_k^v \min\left\{\frac{\pi_k^v}{1 + \|J(x_k, s_k)P_k\|_2^2}, \delta_k^v, 1 - \kappa_{\text{fbn}}\right\} > 0.
\end{aligned} \tag{3.43}$$

Hence, $m_k^v(n_k^{\mathrm{C}}) < m_k^v(0)$, and therefore

$$m_k^v(0) - m_k^v(n_k^{\mathrm{C}}) = \frac{(m_k^v(0))^2 - (m_k^v(n_k^{\mathrm{C}}))^2}{m_k^v(0) + m_k^v(n_k^{\mathrm{C}})} \\ \geq \frac{(m_k^v(0))^2 - (m_k^v(n_k^{\mathrm{C}}))^2}{2m_k^v(0)} = \frac{(m_k^v(0))^2 - (m_k^v(n_k^{\mathrm{C}}))^2}{2v_k}. \quad (3.44)$$

Inequality (3.40) follows from (3.44), (3.43), and $1 + \|J(x_k, s_k)P_k\|_2^2 \geq 1$. □

Since we impose the bound (3.31) on the trust-region radius for the normal step problem on certain iterations, we derive a lower bound on its right-hand side.

Lemma 3.6 *If $k \in \mathcal{N}$, then, with n_k^* defined in* (3.8) *and κ_k^{cn} defined in* (3.41)*,*

$$\|P_k^{-1} n_k^*\|_2 \geq \kappa_k^{\mathrm{cn}} \pi_k^v.$$

Proof Let $w_k := P_k J(x_k, s_k)^T c(x_k, s_k)$. By (3.8) and since m_k is convex with unconstrained minimizer corresponding to a nonnegative α, it follows that $n_k^* = n_k^{\mathrm{C}}(\alpha_N^*))$ is the unconstrained minimizer of $[m_k^v(n_k^C(\alpha))]^2$, from which it follows that

$$P_k^{-1} n_k^* = -\frac{\|w_k\|_2^2}{\|J(x_k, s_k)P_k w_k\|_2^2} P_k J(x_k, s_k)^T c(x_k, s_k).$$

This, along with $\|J(x_k, s_k)P_k w_k\|_2 \leq \|J(x_k, s_k)P_k\|_2 \|w_k\|_2$ and (3.1), yields

$$\|P_k^{-1} n_k^*\|_2 = \frac{\|w_k\|_2^2}{\|J(x_k, s_k)P_k w_k\|_2^2} \pi_k^v \geq \frac{\pi_k^v}{\|J(x_k, s_k)P_k\|_2^2}.$$

The desired bound then follows from (3.41). □

Next, we establish the remaining claims made in (2.1). (We remark that certain bounds established in the proof of Lemma 3.7 are refined in Lemma 4.12.)

Lemma 3.7 *The slack reset* (3.26) *and* (3.33) *in Steps 42 and 48 yields s_{k+1} such that (x_{k+1}, s_{k+1}) satisfies $v_{k+1} \leq v_{k+1}^{\max}$ and, at the end of iteration $k+1$, $v_{k+2}^{\max} \leq v_{k+1}^{\max}$.*

Proof Our proof is by induction. We have $v_0 \leq v_0^{\max}$ by the initialization of $v_0^{\max}$. Now suppose that $v_i \leq v_i^{\max}$ for $i \in \{0, \ldots, k\}$ for some $k \geq 1$. The slack reset in Steps 42 and 48 cannot increase the constraint violation [recall (1.4)], which implies, for $k \in \mathcal{Y} \cup \mathcal{F}$, the inequality $v_{k+1} \leq v_{k+1}^{\max}$. Hence, it remains to consider $k \in \mathcal{V}$. If $\rho_k^v < \eta_1$ or (2.15) does not hold, then the step is rejected, so $v_{k+1} \leq v_{k+1}^{\max}$ as a consequence of (3.36). Otherwise, (2.15) states that $n_k \neq 0$, from which Lemma 3.3 implies $k \in \mathcal{N}$, and thus Lemma 3.5 and (3.6) imply that $\Delta m_k^{v,n} > 0$. It then follows from the fact that $\rho_k^v \geq \eta_1$, (2.13), and (3.32) that $v_{k+1} < v_k$. Since $\kappa_{\mathrm{t2}} \in (0, 1)$ in (3.35), this implies

$$v_{k+1} < v_{k+1} + \kappa_{\mathrm{t2}}(v_k - v_{k+1}) < v_k \leq v_k^{\max}, \quad (3.45)$$

so (3.35) implies $v_{k+1}^{\max} \leq v_k^{\max}$. Combining (3.35) and (3.45), we have that $v_{k+1}^{\max} \geq v_{k+1} + \kappa_{t2}(v_k - v_{k+1}) > v_{k+1}$. Thus, in all cases, we have $v_{k+1} \leq v_{k+1}^{\max}$.

To establish that $v_{k+2}^{\max} \leq v_{k+1}^{\max}$, note that if $k \notin \mathcal{V}$, then $v_{k+2}^{\max} \leftarrow v_{k+1}^{\max}$, so all that remains is to consider $k \in \mathcal{V}$. Observing (3.35), we see again that $v_{k+2}^{\max} \leftarrow v_{k+1}^{\max}$ if either (2.15) is violated or $\rho_{k+1}^v < \eta_1$. By contrast, if (2.15) holds and $\rho_{k+1}^v \geq \eta_1$, then we must have $n_{k+1} \neq 0$ and from Lemma 3.3(ii) that $k+1 \in \mathcal{N}$. Moreover, it follows from (3.32), (2.13), (2.15), (3.6) and Lemma 3.5 as above that $v_{k+2} < v_{k+1}$. Thus, if the maximum value in (3.35) is the second term, it follows that $v_{k+2}^{\max} < v_{k+1} \leq v_{k+1}^{\max}$. Otherwise, if the maximum value in (3.35) is the first term, then $v_{k+2}^{\max} < v_{k+1}^{\max}$ trivially follows since $\kappa_{t1} \in (0, 1)$. □

We now show that the computations of the least-squares multipliers, y_k,—along with the accompanying quantities r_k, π_k^f, and χ_k^f—are well-defined. For this, we make the following reasonable assumption.

Assumption 3.1 If the iterative solver employed to solve (2.7) runs for an infinite number of iterations, then it produces a bounded sequence $\{y^{(i)}\}_{i=0}^{\infty}$ with

$$\lim_{i \to \infty} \nabla m_k^{\mathcal{L}}(y^{(i)}) = 0. \tag{3.46}$$

We now confirm that our strategy for computing Lagrange multiplier estimates and tangential steps is well-defined. In particular, it shows that the strategy in Sect. 3.2.3 produces a Lagrange multiplier estimate and Cauchy point for a tangential subproblem, and that the Cauchy point is a valid option for the tangential step.

Lemma 3.8 *If* $\{y^{(i)}\}_{i=0}^{\infty}$ *is produced by an iterative solver employed to solve* (2.7) *that satisfies Assumption* 3.1, *then for some (finite) index i the vector* $y_k \leftarrow y^{(i)}$ *yields* r_k, π_k^f, *and* χ_k^f *satisfying* (3.15a), (3.15b), *or* (3.15c), *where, in case only* (3.15c) *is satisfied, we also have that either*

(i) the Cauchy point t_k^{C} *defined by* (3.17)–(3.18) *satisfies (3.19), or*
(ii) the Cauchy point t_k^{C} *defined by* (3.21)–(3.22) *satisfies (3.23).*

Proof If, for any i, either (3.15a) or (3.15b) is satisfied, then the result follows. Thus, without loss of generality, let us assume for the remainder of the proof that both (3.15a) and (3.15b) do not hold for all i.

In order to derive a contradiction, suppose that for all i either (3.15c) does not hold or it does while neither statement (i) nor (ii) holds. This means that the iterative solver employed to solve (2.7) (that satisfies Assumption 3.1) does not terminate finitely, which, in turn, implies the existence of a limit point y^{∞} and an infinite index set $\mathcal{I}$ such that $\{y^{(i)}\}_{i \in \mathcal{I}} \to y^{\infty}$. Moreover, (3.46) implies that

$$0 = \nabla m_k^{\mathcal{L}}(y^{\infty}) = J(x_k, s_k) r_k(y^{\infty}). \tag{3.47}$$

Suppose that $\pi_k^f(y^{\infty}) = 0$. If $v_k \leq \epsilon_v$, then this implies that, for all sufficiently large $i \in \mathcal{I}$, the vector $y_k \leftarrow y^{(i)}$ yields (3.15a), a contradiction. Otherwise, if $v_k > \epsilon_v$, then we must have $\chi_k^v > 0$ or else Algorithm 2 would have terminated in Step 9. Since this fact, the fact that $v_k > \epsilon_v$, and (3.1) imply that $\pi_k^v > 0$, it follows along with $\pi_k^f(y^\infty) = 0$ that, for all sufficiently large $i \in \mathcal{I}$, the vector $y_k \leftarrow y^{(i)}$ yields (3.15b), another contradiction. Since we have reached a contradiction in both of these cases, we must conclude that $\pi_k^f(y^\infty) > 0$. Combining this strict inequality with (3.47) and the fact that

$$\nabla m_k^f(n_k) = P_k^{-2} r_k(y^\infty) - J(x_k, s_k)^T y^\infty$$

shows that

$$\chi_k^f(y^\infty) = \frac{r_k(y^\infty)^T(P_k^{-2} r_k(y^\infty) - J(x_k, s_k)^T y^\infty)}{\pi_k^f(y^\infty)} = \frac{(\pi_k^f(y^\infty))^2}{\pi_k^f(y^\infty)} = \pi_k^f(y^\infty).$$

Since $\kappa_\chi \in (0, 1)$, the outer equations in this sequence show that, for all sufficiently large $i \in \mathcal{I}$, the vector $y_k \leftarrow y^{(i)}$ yields (3.15c).

Now, to complete the proof, we must show that either statement (i) or (ii) must be satisfied for some sufficiently large $i \in \mathcal{I}$. To this end, first observe from (3.47) that $\{r_k(y^{(i)})\}_{i\in\mathcal{I}} \to r_k(y^\infty) \in \text{Null}(J(x_k, s_k))$. We introduce the notation $t_k^{\text{Cr}}(i) := t_k^{\text{C}}$ when t_k^{C} is defined by (3.17)–(3.18) with $r_k = r_k(y^{(i)})$ associated with the relaxed SQP tangential subproblem, and $t_k^{\text{Cv}}(i) := t_k^{\text{C}}$ when t_k^{C} is defined by (3.21)–(3.22) with $r_k = r_k(y^{(i)})$ associated with the very relaxed SQP tangential subproblem. We observe from (3.17) and (3.21), the constraints of (3.18) and (3.22), and the fact that $r_k(y^\infty) \in \text{Null}(J(x_k, s_k))$ that there exist $t_k^{\text{Cr}}(\infty)$ and $t_k^{\text{Cv}}(\infty)$ such that $\{t_k^{\text{Cr}}(i)\}_{i\in\mathcal{I}} \to t_k^{\text{Cr}}(\infty) \in \text{Null}(J(x_k, s_k))$ and $\{t_k^{\text{Cv}}(i)\}_{i\in\mathcal{I}} \to t_k^{\text{Cv}}(\infty) \in \text{Null}(J(x_k, s_k))$. By definition, the Cauchy point $t_k^{\text{Cr}}(i)$ satisfies (3.19a)–(3.19c) for all i. Similarly, the Cauchy point $t_k^{\text{Cv}}(i)$ satisfies (3.23a)–(3.23c) for all i. Thus, we need only show that for some sufficiently large $i \in \mathcal{I}$ either $t_k^{\text{Cr}}(i)$ satisfies (3.19d) or $t_k^{\text{Cv}}(i)$ satisfies (3.23d). Suppose that $n_k \neq 0$, in which case Lemma 3.3(ii) implies that $k \in \mathcal{N}$. It then follows from Lemma 3.3(i) that $m_k^v(n_k) < m_k^v(0)$, and thus the right-hand side of (3.19d) is strictly greater than $m_k^v(n_k)$. Therefore, since $t_k^{\text{Cr}}(\infty) \in \text{Null}(J(x_k, s_k))$, it follows that $t_k^{\text{Cr}}(i)$ satisfies (3.19d) for all sufficiently large $i \in \mathcal{I}$, which is to say that statement (i) holds. Now suppose that $n_k = 0$, in which case Lemma 3.3(i) implies that $k \notin \mathcal{N}$. By virtue of (3.2), this must mean that $v_k < \kappa_{\text{vv}} v_k^{\max}$. It follows from the facts that $n_k = 0$, $v_k < \kappa_{\text{vv}} v_k^{\max}$, $\kappa_{\text{tt}} \in (\kappa_{\text{vv}}, 1)$, and $\{t_k^{\text{Cv}}(i)\}_{i\in\mathcal{I}} \to t_k^{\text{Cv}}(\infty) \in \text{Null}(J(x_k, s_k))$ that $t_k^{\text{Cv}}(i)$ satisfies (3.23d) for all sufficiently large $i \in \mathcal{I}$. We have reached the conclusion that statement (ii) holds. This completes the proof. □

Finally, we give a bound on the decrease in our barrier model provided by the Cauchy step for the tangential subproblem.

Lemma 3.9 *If $k \in \mathcal{T}$, then t_k^{C} defined by* (3.17)–(3.18) *or* (3.21)–(3.22) *is computed and satisfies*

$$m_k^f(n_k) - m_k^f(n_k + t_k^{\text{C}}) \geq \kappa_k^{\text{ct}} \pi_k^f \min\left\{\pi_k^f, (1 - \kappa_{\text{B}})\delta_k^t, (1 - \kappa_{\text{fbt}})\kappa_{\text{fbn}}\right\} > 0,$$

where

$$\kappa_k^{\text{ct}} := \frac{\kappa_\chi^2}{2(1+\|P_k G_k P_k\|_2)} \in (0, 1/2).$$

Proof We first consider $k \in \mathcal{T}_{\mathcal{D}}$, i.e., when the Cauchy step t_k^{C} is computed from (3.17)–(3.18) with the trust region radius $\delta_k^t = \min\{\kappa_{\text{vf}}\delta_k^v, \delta_k^f\}$ (see (3.38)). It follows from Lemma 3.3(iii) that $\chi_k^f \geq \kappa_\chi \pi_k^f > 0$ so, by (3.14), $\nabla m_k^f(n_k)^T r_k \geq \kappa_\chi (\pi_k^f)^2 > 0$. We now show that $t_k^{\text{C}}(\alpha)$ [recall (3.17)] is feasible for (3.18) when

$$k \in \mathcal{T}_{\mathcal{D}} \quad \text{and} \quad 0 \leq \alpha \leq (\pi_k^f)^{-1} \min\left\{(1-\kappa_{\text{B}})\delta_k^t, (1-\kappa_{\text{fbt}})\kappa_{\text{fbn}}\right\} =: \alpha_{\text{B}}.$$

Indeed, consider any $\alpha \in [0, \alpha_{\text{B}}]$. The definitions of $t_k^{\text{C}}(\alpha)$, r_k, and α_{B} imply

$$\|P_k^{-1} t_k^{\text{C}}(\alpha)\|_2 = \|P_k^{-1}\alpha r_k\|_2 = \alpha\|P_k^{-1} r_k\|_2 = \alpha\pi_k^f \leq (1-\kappa_{\text{B}})\delta_k^t. \tag{3.48}$$

Using the triangle inequality, (3.12) (which must hold since $k \in \mathcal{T}_{\mathcal{D}} \subseteq \mathcal{T}$), (3.38), and (3.48), we then have

$$\begin{aligned}\|P_k^{-1}\big(n_k + t_k^{\text{C}}(\alpha)\big)\|_2 &\leq \|P_k^{-1} n_k\|_2 + \|P_k^{-1} t_k^{\text{C}}(\alpha)\|_2 \\ &\leq \kappa_{\text{B}}\delta_k^t + (1-\kappa_{\text{B}})\delta_k^t \leq \delta_k^t = \min\{\kappa_{\text{vf}}\delta_k^v, \delta_k^f\},\end{aligned}$$

which shows that $t_k^{\text{C}}(\alpha)$ satisfies the first constraint in problem (3.18). To show that $t_k^{\text{C}s}(\alpha)$ also satisfies the second constraint in problem (3.18), first observe that if $[t_k^{\text{C}s}(\alpha)]_i = [-\alpha r_k^s]_i \geq 0$, then $[s_k + n_k^s + t_k^{\text{C}s}(\alpha)]_i \geq [s_k + n_k^s]_i \geq \kappa_{\text{fbt}}[s_k + n_k^s]_i \geq 0$ since $\kappa_{\text{fbt}} \in (0, 1)$. Thus, it suffices to consider i such that $[r_k^s]_i > 0$. It follows from the definitions of α_{B} and π_k^f, (3.13), $[r_k^s]_i > 0$, Lemma 3.4, and (3.5) that

$$\begin{aligned}\alpha \leq \alpha_{\text{B}} \leq \frac{(1-\kappa_{\text{fbt}})\kappa_{\text{fbn}}}{\pi_k^f} &\leq \frac{(1-\kappa_{\text{fbt}})\kappa_{\text{fbn}}}{\|S_k^{-1} r_k^s\|_2} \\ \leq \frac{(1-\kappa_{\text{fbt}})\kappa_{\text{fbn}}}{|[r_k^s]_i/[S_k]_{ii}|} = \frac{(1-\kappa_{\text{fbt}})\kappa_{\text{fbn}}[s_k]_i}{[r_k^s]_i} &\leq \frac{(1-\kappa_{\text{fbt}})[s_k + n_k^s]_i}{[r_k^s]_i}.\end{aligned}$$

Using the definition of $t_k^{\text{C}s}(\alpha)$ and the previous inequality leads to

$$[-t_k^{\text{C}s}(\alpha)]_i = \alpha[r_k^s]_i \leq (1-\kappa_{\text{fbt}})[s_k + n_k^s]_i$$

from which we may conclude that $[s_k + n_k^s + t_k^{\text{C}s}(\alpha)]_i \geq \kappa_{\text{fbt}}[s_k + n_k^s]_i$ for all $i \in \{1, \ldots, M\}$. This proves that $t_k^{\text{C}s}(\alpha)$ satisfies the constraints of problem (3.18), and completes the proof that $t_k^{\text{C}}(\alpha)$ is feasible for problem (3.18) for all $\alpha \in [0, \alpha_{\text{B}}]$.

We now observe that the minimizer $\alpha_{\text{T}}^{\text{C}}$ of (3.18) yields $m_k^f(n_k + t_k^{\text{C}}) \equiv m_k^f(n_k + t_k^{\text{C}}(\alpha_{\text{T}}^{\text{C}})) \leq m_k^f(n_k + t_k^{\text{C}}(\alpha))$ for all $\alpha \in [0, \alpha_{\text{B}}]$. We also have from the Cauchy-Schwarz and standard norm inequalities that

$$|r_k^T G_k r_k| = \left|\left(\nabla m_k^f(n_k) + J(x_k, s_k)^T y_k\right)^T P_k^2 G_k P_k^2 \left(\nabla m_k^f(n_k) + J(x_k, s_k)^T y_k\right)^T\right|$$
$$\leq (\pi_k^f)^2 \|P_k G_k P_k\|_2.$$

It then follows from [3, Lemma 1] with the quantities

$$\text{“t”} := \alpha_{\text{B}}, \quad \text{“a”} := |r_k^T G_k r_k|, \quad \text{“b”} := \nabla m_k^f(n_k)^T r_k > 0,$$

(the strict inequality was shown earlier in this proof) that

$$\begin{aligned}
& m_k^f(n_k) - m_k^f(n_k + t_k^{\text{C}}) \\
& \geq \frac{\text{“b”}}{2} \min\left\{\frac{\text{“b”}}{\text{“a”}}, \text{“t”}\right\} \\
& \geq \frac{\nabla m_k^f(n_k)^T r_k}{2} \min\left\{\frac{\nabla m_k^f(n_k)^T r_k}{(\pi_k^f)^2 \|P_k G_k P_k\|_2}, \frac{(1-\kappa_{\text{B}})\delta_k^t}{\pi_k^f}, \frac{(1-\kappa_{\text{fbt}})\kappa_{\text{fbn}}}{\pi_k^f}\right\} \\
& \geq \frac{\nabla m_k^f(n_k)^T r_k}{2\pi_k^f} \min\left\{\frac{\nabla m_k^f(n_k)^T r_k}{\pi_k^f (1 + \|P_k G_k P_k\|_2)}, (1-\kappa_{\text{B}})\delta_k^t, (1-\kappa_{\text{fbt}})\kappa_{\text{fbn}}\right\} \\
& = \frac{\chi_k^f}{2} \min\left\{\frac{\chi_k^f}{(1 + \|P_k G_k P_k\|_2)}, (1-\kappa_{\text{B}})\delta_k^t, (1-\kappa_{\text{fbt}})\kappa_{\text{fbn}}\right\} \\
& \geq \frac{\kappa_\chi^2 \pi_k^f}{2(1 + \|P_k G_k P_k\|_2)} \min\left\{\pi_k^f, (1-\kappa_{\text{B}})\delta_k^t, (1-\kappa_{\text{fbt}})\kappa_{\text{fbn}}\right\},
\end{aligned}$$

where we have used $1 + \|P_k G_k P_k\|_2 \geq 1$ and $\chi_k^f \geq \kappa_\chi \pi_k^f$ with $\kappa_\chi \in (0, 1)$.

The proof for $k \in \mathcal{T} \backslash \mathcal{T}_{\mathcal{D}}$ is similar, but uses $\delta_k^t = \min\{\kappa_{\text{vf}}, \delta_k^v, \delta_k^f, \kappa_{\text{v}} v_k^{\max}\}$, (3.21) instead of (3.17), (3.22) instead of (3.18), and (by Lemma 3.3(xi)) the fact that $n_k = 0$ for $k \in \mathcal{T} \backslash \mathcal{T}_{\mathcal{D}}$. □

4 Convergence of the trust-funnel algorithm for the barrier subproblem

The following assumption is assumed to hold for the remainder of the paper.

Assumption 4.1 The sequence of iterates $\{x_k\}$ is contained in a compact set.

The following is an immediate consequence of Assumptions 1.1 and 4.1.

Lemma 4.1 *There exists an upper bound* $\kappa_{\text{H}} \geq 1$ *for* $\|g(x_k)\|_2$, $\|c(x_k)\|_2$, $\|J(x_k)\|_2$, $\|\nabla_{xx} f(x_k)\|_2$, *and* $\|\nabla_{xx} c_i(x_k)\|_2$ *for all* k *and* $i \in \{1, \ldots, M\}$.

We now prove that important sequences related to our method are bounded.

Lemma 4.2 *There exists a upper bound* $\kappa_{\text{ub}} \geq \kappa_{\text{H}}$ *for* v_k, $\|s_k\|_2$, $\|J(x_k, s_k)^T c(x_k, s_k)\|_2$, π_k^v, $\|P_k J(x_k, s_k)^T\|_2$, χ_k^v, $\|P_k G_k P_k\|_2$, *and* $\|P_k \nabla f(x_k, s_k)\|_2$ *for all* k.

Proof The result is clearly true if the algorithm terminates finitely. Otherwise, it follows from Lemma 3.7 that $v_k \leq v_k^{\max} \leq v_0^{\max}$ for all k, which proves that $\{v_k\}$ can be bounded as claimed. Combining this with the triangle inequality yields

$$\|s_k\|_2 - \|c(x_k)\|_2 \leq \|c(x_k) + s_k\|_2 = \|c(x_k, s_k)\|_2 \leq v_0^{\max} \text{ for all } k.$$

We may deduce from this bound and Lemma 4.1 that $\{\|s_k\|_2\}$ can be bounded as claimed. It then follows from the triangle inequality that

$$\|J(x_k, s_k)^T c(x_k, s_k)\|_2 \leq \left\| \begin{pmatrix} J(x_k)^T c(x_k, s_k) \\ 0 \end{pmatrix} \right\|_2 + \left\| \begin{pmatrix} 0 \\ c(x_k, s_k) \end{pmatrix} \right\|_2,$$

which may then be combined with the Cauchy-Schwarz inequality, Lemma 4.1, and the boundedness of $\{v_k\}$ (already proved) to conclude that $\{\|J(x_k, s_k)^T c(x_k, s_k)\|_2\}$ can be bounded as claimed. The boundedness of $\{\pi_k^v\}$ follows from that of $\{\|s_k\|_2\}$ and $\{\|J(x_k, s_k)^T c(x_k, s_k)\|_2\}$. It then follows from the boundedness of $\{\|s_k\|_2\}$ and, by Lemma 4.1, that of $\{\|J(x_k)\|_2\}$ that $\{\|P_k J(x_k, s_k)^T\|_2\}$ can be bounded as claimed. This, along with the Cauchy-Schwarz inequality, implies that $\{\chi_k^v\}$ can be bounded as claimed. The boundedness of $\|P_k G_k P_k\|_2$ follows from the boundedness of $\{\|s_k\|_2\}$, (3.9), (3.10), Assumptions 1.1 and 4.1, and (3.11). Finally, it follows from Lemma 4.1 and the fact that $P_k \nabla f(x_k, s_k) = (g(x_k), -\mu e)$ that $\{\|P_k \nabla f(x_k, s_k)\|_2\}$ can be bounded as claimed. □

Using Lemma 4.2, we may now improve the Cauchy decrease bounds provided in Lemmas 3.5 and 3.9, as well as the result of Lemma 3.6 by making the leading constants independent of the iteration number.

Lemma 4.3 *For all k, the following hold:*

(i) *If* $k \in \mathcal{N}$, *then* n_k^{C} *defined by* (3.3)–(3.4) *is computed and*

$$m_k^v(0) - m_k^v(n_k^{\mathrm{C}}) \geq \kappa_{\mathrm{cn}} \chi_k^v \min\{\pi_k^v, \delta_k^v, 1 - \kappa_{\mathrm{fbn}}\} > 0$$

for some constant $\kappa_{\mathrm{cn}} \in (0, 1]$ *independent of* k.

(ii) *If* $k \in \mathcal{T}$, *then* t_k^{C} *defined by* (3.17)–(3.18) *or* (3.21)–(3.22) *is computed and*

$$m_k^f(n_k) - m_k^f(n_k + t_k^{\mathrm{C}}) \geq \kappa_{\mathrm{ct}} \pi_k^f \min\{\pi_k^f, (1 - \kappa_{\mathrm{B}})\delta_k^t, (1 - \kappa_{\mathrm{fbt}})\kappa_{\mathrm{fbn}}\} > 0$$

for some constant $\kappa_{\mathrm{ct}} \in (0, 1/2]$ *independent of* k.

(iii) *If* $k \in \mathcal{N}$, *then, with* n_k^* *defined in* (3.8) *and* $\kappa_{\mathrm{cn}} \in (0, 1]$ *from part (i),*

$$\|P_k^{-1} n_k^*\|_2 \geq \kappa_{\mathrm{cn}} \pi_k^v.$$

Proof The results follow from Lemmas 3.5, 3.9 and 3.6 along with Lemma 4.2. □

The next lemma bounds the size of the trial step in different scenarios.

Lemma 4.4 *If Algorithm 2 does not terminate during iteration k, then*

$$\|P_k^{-1}d_k\|_2 \begin{cases} = \|P_k^{-1}n_k\|_2 \le \delta_k^v & \text{if } k \notin \mathcal{T}, \\ = \|P_k^{-1}n_k\|_2 \le \min\{\kappa_{\text{vf}}\delta_k^v, \delta_k^v, \delta_k^f\} & \text{if } k \in \mathcal{T}_0, \\ \le \delta_k^t & \text{if } k \in \mathcal{T}\backslash\mathcal{T}_0. \end{cases}$$

In particular, for all k, we have $\|P_k^{-1}d_k\|_2 \le \max\{\kappa_{\text{vf}}\delta_k^v, \delta_k^v\}$.

Proof Let $k \notin \mathcal{T}$, from which Lemma 3.3(v) implies $t_k = 0$ and $d_k = n_k$. If $n_k = 0$, then the result holds trivially. Conversely, if $n_k \neq 0$, then Lemma 3.3(ii) implies that $k \in \mathcal{N}$ and the result follows from (3.5). Now let $k \in \mathcal{T}$, for which we have three cases. First, if $k \in \mathcal{T}_0$, then it follows from Lemma 3.3(iv) that $t_k = 0$ and (3.12) holds. Combining this with $t_k = 0$, (3.5), and the fact that $\kappa_{\text{B}} \in (0, 1)$ shows that

$$\|P_k^{-1}d_k\|_2 = \|P_k^{-1}n_k\|_2 \le \min\{\kappa_{\text{B}} \min\{\kappa_{\text{vf}}\delta_k^v, \delta_k^f\}, \delta_k^v\} \le \min\{\kappa_{\text{vf}}\delta_k^v, \delta_k^v, \delta_k^f\}.$$

Second, if $k \in \mathcal{T}_{\mathcal{D}}\backslash\mathcal{T}_0$, then the result follows from (3.19c) and (3.38). Third, if $k \in \mathcal{T}\backslash\mathcal{T}_{\mathcal{D}}$, then the result follows from (3.23c) and (3.38). □

We now bound the differences between the problem functions and their models.

Lemma 4.5 *The following hold:*

(i) *There exists a constant* $\kappa_{\text{G}} > 0$ *independent of k such that*

$$|f(x_k + d_k^x, s_k + d_k^s) - m_k^f(d_k)| \le \kappa_{\text{G}}\|P_k^{-1}d_k\|_2^2 \quad \textit{for all} \quad k. \tag{4.1}$$

(ii) *There exists a constant* $\kappa_{\text{C}} > 0$ *independent of k such that*

$$|v(x_k + d_k^x, s_k + d_k^s) - m_k^v(d_k)| \le \kappa_{\text{C}}\|P_k^{-1}d_k\|_2^2 \quad \textit{for all} \quad k. \tag{4.2}$$

Proof We first prove part (i). By the triangle inequality, we have

$$\begin{aligned} &|f(x_k + d_k^x, s_k + d_k^s) - m_k^f(d_k)| \\ &\quad\le |f(x_k + d_k^x) - f(x_k) - \nabla f(x_k)^T d_k^x - \tfrac{1}{2}d_k^{xT}\nabla_{xx}\mathcal{L}(x_k, y_k^{\text{B}})d_k^x| \\ &\qquad + \left|-\mu\sum_{i=1}^M \ln([s_k + d_k^s]_i) + \mu\sum_{i=1}^M \ln([s_k]_i) + \mu e^T S_k^{-1}d_k^s - \tfrac{1}{2}d_k^{sT}D_k d_k^s\right|. \end{aligned} \tag{4.3}$$

Under Assumptions 1.1 and 4.1, and by (3.10), there exists $\kappa_{\text{G1}} > 0$ such that

$$|f(x_k + d_k^x) - f(x_k) - \nabla f(x_k)^T d_k^x - \tfrac{1}{2}d_k^{xT}\nabla_{xx}\mathcal{L}(x_k, y_k^{\text{B}})d_k^x| \le \kappa_{\text{G1}}\|d_k^x\|_2^2. \tag{4.4}$$

Moreover, note that for each $i \in \{1, \ldots, M\}$, we have by (3.5) and (3.19b)/(3.23b) that $[s_k]_i + [d_k^s]_i \ge \kappa_{\text{fbt}}\kappa_{\text{fbn}}[s_k]_i > 0$ for all k regardless of whether a tangential step t_k

was computed. The Mean Value Theorem yields $\ln([s_k]_i + [d_k^s]_i) - \ln[s_k]_i = [d_k^s]_i/\xi_i$, where ξ_i lies between $[s_k]_i$ and $[s_k]_i + [d_k^s]_i$. Hence

$$\left|\ln([s_k]_i + [d_k^s]_i) - \ln[s_k]_i - \frac{[d_k^s]_i}{[s_k]_i}\right| \leq \sup_{\xi \in [[s_k]_i, [s_k]_i + [d_k^s]_i]} \left|\frac{[d_k^s]_i}{\xi} - \frac{[d_k^s]_i}{[s_k]_i}\right|$$
$$= \frac{[s_k]_i}{[s_k]_i + [d_k^s]_i}\left(\frac{[d_k^s]_i}{[s_k]_i}\right)^2 \leq \frac{1}{\kappa_{\text{fbt}}\kappa_{\text{fbn}}}\left(\frac{[d_k^s]_i}{[s_k]_i}\right)^2,$$

where in the middle equation we have used the fact that the sup occurs at $\xi = [s_k]_i + [d_k^s]_i$. Hence, by (3.11) and Lemma 4.2, we have that

$$\left|-\mu \sum_{i=1}^{M} \ln([s_k + d_k^s]_i) + \mu \sum_{i=1}^{M} \ln([s_k]_i) + \mu e^T S_k^{-1} d_k^s - \tfrac{1}{2} d_k^{sT} D_k d_k^s\right|$$
$$\leq \frac{1}{\kappa_{\text{fbt}}\kappa_{\text{fbn}}} d_k^{sT}(\mu S_k^{-2}) d_k^s + \tfrac{1}{2}|d_k^{sT} D_k d_k^s| \leq \kappa_{\text{G2}} \|S_k^{-1} d_k^s\|_2^2, \tag{4.5}$$

where $\kappa_{\text{G2}} = \mu/\kappa_{\text{fbt}}\kappa_{\text{fbn}} + \frac{1}{2}\kappa_{\text{ub}}^2 \kappa_{\text{D}} > 0$. The result now follows from (4.3)– (4.5) and Lemma 4.4 with $\kappa_{\text{G}} := \kappa_{\text{G1}} + \kappa_{\text{G2}}$.

We now prove part (ii). By Lemma 4.1, Taylor's expansion theorem yields

$$c(x_k + d_k^x, s_k + d_k^s) = c(x_k, s_k) + J(x_k, s_k) d_k + w_k \quad \text{where} \quad [w_k]_i = \tfrac{1}{2} d_k^{xT} \nabla_{xx} c_i(\xi_{ik}) d_k^x$$

for some scalars $\xi_{ik} \in [x_k, x_k + d_k^x]$. As a consequence, we obtain with the triangle inequality that there exists a constant $\kappa_{\text{C}} > 0$ so that

$$|v(x_k + d_k^x, s_k + d_k^s) - m_k^v(d_k)| = |\|c(x_k + d_k^x, s_k + d_k^s)\|_2 - \|c(x_k, s_k) + J(x_k, s_k) d_k\|_2|$$
$$\leq \|w_k\|_2 \leq \kappa_{\text{C}} \|d_k^x\|_2^2 \leq \kappa_{\text{C}} \|P_k^{-1} d_k\|_2^2,$$

where we have used Lemma 4.1 and the Cauchy-Schwarz inequality. □

We now prove an important fact about v-iterations; namely, if $k \in \mathcal{V}$ and one of the trust region radii or funnel radius is sufficiently small, then $k \in \mathcal{D}$.

Lemma 4.6 *If $k \in \mathcal{V}$ and*

$$\min\{\kappa_{\text{vf}} \delta_k^v, \delta_k^f, \kappa_{\text{v}} v_k^{\max}\} \leq \frac{(1 - \kappa_{\text{tt}})}{\kappa_{\text{C}} \kappa_{\text{v}}} =: \kappa_{\mathcal{V}}, \tag{4.6}$$

then $k \in \mathcal{D}$.

Proof For a proof by contradiction, suppose that (4.6) holds while $k \in \mathcal{V} \backslash \mathcal{D}$. We show that all of the conditions of an f-iteration are satisfied, implying that $k \in \mathcal{F}$, contradicting the supposition that $k \in \mathcal{V}$.

Since $k \notin \mathcal{D}$, we have from Lemma 3.3(viii) that $k \in \mathcal{T} \backslash \mathcal{T}_{\mathcal{D}}$ and (3.23) holds. Then, since $\mathcal{T}_0 \subseteq \mathcal{T}_{\mathcal{D}}$, it follows that $k \in \mathcal{T} \backslash \mathcal{T}_0$, so by Lemma 3.3(iv) we have $t_k \neq 0$.

Moreover, $k \in \mathcal{T} \backslash \mathcal{T}_0$ implies by Lemma 4.4 that $\|P_k^{-1} d_k\|_2 \leq \delta_k^t$, which along with the fact that $k \in \mathcal{T} \backslash \mathcal{T}_{\mathcal{D}}$ and (3.38) implies

$$\|P_k^{-1} d_k\|_2 \leq \min\{\kappa_{\text{vf}} \delta_k^v, \delta_k^f, \kappa_{\text{v}} v_k^{\text{max}}\} \leq \kappa_{\text{v}} v_k^{\text{max}}. \tag{4.7}$$

Thus, with (4.2), the triangle inequality, (3.23d), (4.7), and (4.6), we have

$$\begin{aligned} v(x_k + d_k^x, s_k + d_k^s) &\leq \kappa_{\text{tt}} v_k^{\text{max}} + \kappa_{\text{C}} \|P_k^{-1} d_k\|_2^2 \\ &\leq \kappa_{\text{tt}} v_k^{\text{max}} + \kappa_{\text{C}} \kappa_{\text{v}} v_k^{\text{max}} \min\{\kappa_{\text{vf}} \delta_k^v, \delta_k^f, \kappa_{\text{v}} v_k^{\text{max}}\} \leq v_k^{\text{max}}, \end{aligned}$$

so (2.11) holds. We also have from Lemma 3.3(xi) that $n_k = 0$, so (2.10) holds. Thus, all of the conditions of an f-iteration are satisfied, so the result follows. □

The preceding lemmas have the following useful consequence.

Lemma 4.7 *There exists a constant* $\kappa_{\text{n}\Delta 2} \in (0, 1]$ *such that if* $k \in \mathcal{V}$ *and*

$$\min\{\kappa_{\text{vf}} \delta_k^v, \delta_k^f\} \leq \min\{1, \kappa_{\mathcal{V}}, \kappa_{\text{n}\Delta 2} \pi_k^v\}, \tag{4.8}$$

then $k \in \mathcal{N} \cap \mathcal{D}$.

Proof We first note that, by Lemma 4.2, we have $\chi_k^v \leq \kappa_{\text{ub}}$ for all k. Then, with

$$\kappa_{\text{n}\Delta 2} := \min\left\{1, \frac{\kappa_{\text{v}}}{\kappa_{\text{ub}}}\right\} \in (0, 1], \tag{4.9}$$

we have with Lemma 3.7 that

$$\kappa_{\text{n}\Delta 2} \pi_k^v = \kappa_{\text{n}\Delta 2} \chi_k^v v_k \leq \kappa_{\text{n}\Delta 2} \kappa_{\text{ub}} v_k \leq \kappa_{\text{v}} v_k \leq \kappa_{\text{v}} v_k^{\text{max}}. \tag{4.10}$$

Let $k \in \mathcal{V}$ and (4.8) hold. Then, along with (4.10) we have that

$$\min\{\kappa_{\text{vf}} \delta_k^v, \delta_k^f, \kappa_{\text{v}} v_k^{\text{max}}\} = \min\{\kappa_{\text{vf}} \delta_k^v, \delta_k^f\} \leq \kappa_{\mathcal{V}}.$$

Then, by Lemma 4.6, we have $k \in \mathcal{D}$ (as desired), so $k \in \mathcal{V} \cap \mathcal{D}$. Now, in order to derive a contradiction to the claim that $k \in \mathcal{N}$, suppose that $k \in (\mathcal{V} \cap \mathcal{D}) \backslash \mathcal{N}$. Since $k \notin \mathcal{N}$, we have from Lemma 3.3(ii) that $n_k = 0$, so (2.10) holds. Then, since $k \in \mathcal{V}$, we must have $t_k \neq 0$ (since otherwise Lemma 3.3(vi) would imply that $k \in \mathcal{Y}$, which is a contradiction). Thus, we have that $k \in \mathcal{T} \backslash \mathcal{T}_0$. At the same time, $k \notin \mathcal{N}$ implies that (3.2) does not hold, so $v_k < \kappa_{\text{vv}} v_k^{\text{max}} < \kappa_{\text{tt}} v_k^{\text{max}}$. This bound, (4.2), the triangle inequality, (3.19d), the fact that $n_k = 0$, Lemma 3.7, the fact that $k \in \mathcal{T} \backslash \mathcal{T}_0$, Lemma 4.4, (3.38), (4.10) and (4.8) imply

$$\begin{aligned} v(x_k + d_k^x, s_k + d_k^s) &\le |m_k^v(d_k)| + \kappa_C \|P_k^{-1} d_k\|_2^2 \\ &\le |m_k^v(0)| + \kappa_C \|P_k^{-1} d_k\|_2^2 \\ &< \kappa_{tt} v_k^{\max} + \kappa_C (\min\{\kappa_{vf}\delta_k^v, \delta_k^f\})^2 \\ &\le \kappa_{tt} v_k^{\max} + \kappa_C \kappa_v v_k^{\max} \min\{\kappa_{vf}\delta_k^v, \delta_k^f\}, \end{aligned}$$

which, when combined with (4.8) and (4.6), yields

$$v(x_k + d_k^x, s_k + d_k^s) \le \kappa_{tt} v_k^{\max} + (1 - \kappa_{tt}) v_k^{\max} = v_k^{\max}$$

so that (2.11) holds. Combining this with the facts that $t_k \neq 0$ and (2.10) hold shows that $k \in \mathcal{F}$, which is a contradiction. Thus, we conclude that $k \in \mathcal{N}$. □

We now prove that, in certain situations, a sufficiently small trust region radius is guaranteed to lead to a successful iteration.

Lemma 4.8 *The following hold:*

(i) *If $k \in \mathcal{F}$ and*

$$\delta_k^t \le \min\left\{\frac{(1-\kappa_{fbt})\kappa_{fbn}}{1-\kappa_B}, \frac{\pi_k^f}{1-\kappa_B}, \frac{\kappa_\delta \kappa_{ct}(1-\kappa_B)(1-\eta_2)\pi_k^f}{\kappa_G}\right\} =: \min\{\kappa_{\Delta f1}, \kappa_{\Delta f2}\pi_k^f\}$$

then $\rho_k^f \ge \eta_2$, $k \in \mathcal{S}_f$, and $\delta_{k+1}^f \ge \delta_k^f$.

(ii) *If $k \in \mathcal{V}$ and*

$$\begin{aligned} \delta_k^v &\le \min\left\{\frac{1}{\kappa_{vf}}, \frac{\kappa_{\mathcal{V}}}{\kappa_{vf}}, \frac{\kappa_{n\Delta 2}\pi_k^v}{\max\{\kappa_{vf}, 1\}}, 1-\kappa_{fbn}, \frac{\kappa_{cd}\kappa_{cn}\chi_k^v(1-\eta_2)}{\kappa_C(\max\{\kappa_{vf}, 1\})^2}\right\} \\ &=: \min\{\kappa_{\Delta c1}, \kappa_{\Delta c2}\pi_k^v, \kappa_{\Delta c3}\chi_k^v\}, \end{aligned}$$

then $k \in \mathcal{N} \cap \mathcal{D} \cap \mathcal{S}_v$, $\rho_k^v \ge \eta_2$, and $\delta_{k+1}^v \ge \delta_k^v$.

Proof For part (i), the proof that $\rho_k^f \ge \eta_2$, which implies that $k \in \mathcal{S}_f$, is the same as for [5, Theorem 6.4.2] and uses (2.12), (2.10) (which holds since $k \in \mathcal{F}$), (3.19a)/(3.23a), Lemma 4.3(ii), the assumed upper bound on δ_k^t, (4.1), $t_k \neq 0$, and Lemma 4.4. The fact that $\delta_{k+1}^f \ge \delta_k^f$ then follows from (3.27) and (3.29).

To prove part (ii), we first observe from the assumed upper bound on δ_k^v that $\pi_k^v > 0$ and $\chi_k^v > 0$ since $\delta_k^v > 0$ by construction in the algorithm. Moreover, the assumed upper bound on δ_k^v and Lemma 4.7 imply that $k \in \mathcal{N} \cap \mathcal{D}$. We now conclude from Lemma 3.3(ix) that (2.15) holds. Thus, using (4.2), Lemma 4.4, (2.15), (3.6), and Lemma 4.3(i), we have

$$\begin{aligned} |\rho_k^v - 1| &= \left|\frac{v(x_k + d_k^x, s_k + d_k^s) - m_k^v(d_k)}{\Delta m_k^{v,d}}\right| \\ &\le \left|\frac{\kappa_C(\max\{\kappa_{vf}, 1\}\delta_k^v)^2}{\kappa_{cd}\Delta m_k^{v,n}}\right| \le \frac{\kappa_C(\max\{\kappa_{vf}, 1\}\delta_k^v)^2}{\kappa_{cd}\kappa_{cn}\chi_k^v \min\{\pi_k^v, \delta_k^v, 1-\kappa_{fbn}\}}. \end{aligned}$$

In fact, we have from the assumed upper bound on δ_k^v and $\kappa_{n\Delta 2} \in (0,1]$ that $\delta_k^v = \min\{\pi_k^v, \delta_k^v, 1-\kappa_{fbn}\}$, so that

$$|\rho_k^v - 1| \le \frac{\kappa_C(\max\{\kappa_{vf}, 1\})^2 \delta_k^v}{\kappa_{cd}\kappa_{cn}\chi_k^v} \le 1 - \eta_2.$$

Thus, $\rho_k^v \ge \eta_2 \ge \eta_1$, which means that $k \in \mathcal{S}_v$ and, by (3.34), that $\delta_{k+1}^v \ge \delta_k^v$. □

We now give a lower bound on the trust-region radii when the criticality measures π_k^f and $\min\{v_k, \chi_k^v\}$ are bounded away from zero on f- or v-iterations.

Lemma 4.9 *If there exists a constant* $\epsilon_f > 0$ *such that*

$$\pi_k^f \ge \epsilon_f \ \textit{for all} \ \ k \in \mathcal{F}, \tag{4.11}$$

then, for some constant $\epsilon_{\mathcal{F}} > 0$*, we have*

$$\delta_k^f \ge \epsilon_{\mathcal{F}} \ \textit{for all} \ \ k. \tag{4.12}$$

Proof The statement follows from Lemma 4.8(i), (3.38), $\mathcal{F} \subseteq \mathcal{T} \backslash \mathcal{T}_0$, and the fact that $\delta_{k+1}^f \leftarrow \delta_k^f$ for $k \notin \mathcal{F}$. □

Lemma 4.10 *If there exists a constant* $\epsilon_\theta > 0$ *such that*

$$\min\{v_k, \chi_k^v\} \ge \epsilon_\theta \textit{ for all } k \in \mathcal{V}, \tag{4.13}$$

then

$$\delta_k^v \ge \gamma_1 \min\left\{\delta_0^v, \kappa_{\Delta c1}, \kappa_{\Delta c2}\epsilon_\theta^2, \kappa_{\Delta c3}\epsilon_\theta\right\} =: \epsilon_{\mathcal{C}} \ \textit{for all} \ \ k. \tag{4.14}$$

Proof With $\gamma_1 \in (0,1)$ defined for (3.29), we prove by induction that, for all k,

$$\delta_k^v \ge \gamma_1 \min\left\{\delta_0^v, \kappa_{\Delta c1}, \kappa_{\Delta c2}\left[\min_{j\in\{0,\ldots,k\}\cap\mathcal{V}} \pi_j^v\right], \kappa_{\Delta c3}\left[\min_{j\in\{0,\ldots,k\}\cap\mathcal{V}} \chi_j^v\right]\right\}. \tag{4.15}$$

This inequality holds trivially for $k = 0$, so supposing that it holds for iteration k, we prove that it holds for iteration $k+1$. Observe that δ_k^v cannot be decreased if Step 13 is reached; hence, we may ignore this safeguard throughout this proof.

First, suppose that $k \in \mathcal{Y} \cup (\mathcal{F}\backslash\mathcal{S}_f)$. Since $\delta_{k+1}^v \leftarrow \delta_k^v$ and $(x_{k+1}, s_{k+1}) \leftarrow (x_k, s_k)$ for such iterations, we conclude that (4.15) holds at iteration $k+1$. Second, if $k \in \mathcal{S}_f \cup \mathcal{S}_v$, then the fact that $\delta_{k+1}^v \ge \delta_k^v$ ensures that (4.15) holds at iteration $k+1$. Finally, suppose that $k \in \mathcal{V}\backslash\mathcal{S}_v$. In this case, Lemma 4.8(ii) implies that $\delta_k^v > \min\{\kappa_{\Delta c1}, \kappa_{\Delta c2}\pi_k^v, \kappa_{\Delta c3}\chi_k^v\}$. This may then be combined with (3.36) to deduce that $\delta_{k+1}^v \ge \gamma_1 \min\{\kappa_{\Delta c1}, \kappa_{\Delta c2}\pi_k^v, \kappa_{\Delta c3}\chi_k^v\}$ so that (4.15) holds at iteration $k+1$. The bound (4.14) then follows from (4.15), (4.13), (3.1), Lemma 4.2, and the observation that δ_k^v is not decreased for $k \in \mathcal{Y} \cup \mathcal{F}$. □

We now give our first main result, which states that if there are finitely many successful iterations, then Algorithm 2 terminates finitely.

Theorem 4.11 *If $|\mathcal{S}| < \infty$, then Algorithm 2 terminates finitely.*

Proof To derive a contradiction, suppose that Algorithm 2 does not terminate finitely. It then follows from the fact that $|\mathcal{S}| < \infty$, (3.24), (3.29), (3.30), and (3.36) that for some $x_* \in \mathbb{R}^N$, $s_* \in \mathbb{R}^M$, and $\{v_*, v_*^{\max}, \pi_*^v, \chi_*^v\} \subset \mathbb{R}$ there exists an integer k_s such that, for all $k \geq k_s$, Step 13 is not reached,

$$(x_k, s_k, v_k, v_k^{\max}, \pi_k^v, \chi_k^v) = (x_*, s_*, v_*, v_*^{\max}, \pi_*^v, \chi_*^v), \quad \text{and} \quad k \notin \mathcal{S}. \tag{4.16}$$

Also, $v_*^{\max} > 0$ while the fact that $|\mathcal{S}| < \infty$ and Lemma 3.4 ensure that $s_* > 0$.

First, we prove that $|\mathcal{V}| < \infty$. In order to derive a contradiction, suppose that $|\mathcal{V}| = \infty$. Then, by (4.16) (in particular, the fact that $k \notin \mathcal{S}$ for $k \geq k_s$), it follows that (3.36) sets $\delta_{k+1}^v \leq \gamma_2 \delta_k^v$ for all $k \in \mathcal{V}$ with $k \geq k_s$. Combining this with the fact that (3.24) and (3.29) set $\delta_{k+1}^v \leftarrow \delta_k^v$ for all $k \in \mathcal{Y} \cup \mathcal{F}$ with $k \geq k_s$, it follows that $\{\delta_k^v\} \to 0$. We also have from Lemma 4.8(ii) and the facts that $|\mathcal{V}| = \infty$ and $|\mathcal{S}| < \infty$ that we must have $0 = \lim_{k \in \mathcal{V}} \min\{\pi_k^v, \chi_k^v\} = \lim_{k \in \mathcal{V}} \min\{\chi_k^v v_k, \chi_k^v\} = \min\{\chi_*^v v_*, \chi_*^v\}$. If $v_* > 0$, then this implies that $\chi_*^v = 0$. However, this implies that for $k = k_s$ the algorithm would terminate finitely in Step 9, which contradicts the supposition of the proof. Thus, we must have that $v_* = 0$. Since $v_* = 0$, it follows from the conditions of Step 10 that $n_k = 0$ for all $k \geq k_s$. This implies that (3.12) will be satisfied for all $k \geq k_s$, which in turn implies by Step 18 of the algorithm that y_k, r_k, π_k^f, and χ_k^f will be computed to satisfy (3.15a), (3.15b), or (3.15c). If (3.15a) were to hold, then the algorithm would terminate finitely, which is a contradiction. Thus, we know that (3.15a) does not hold for all $k \geq k_s$, which combined with the fact that $v_* = 0$ implies that $\pi_k^f > \epsilon_\pi > 0$ for all $k \geq k_s$. It follows from this fact, Lemma 4.8(i), (3.38), and $\{\delta_k^v\} \to 0$ that if $|\mathcal{F}| = \infty$ (recall $\mathcal{F} \subseteq \mathcal{T}$), then we would have $\{\delta_k^t\}_{k \in \mathcal{F}} \to 0$ and an infinite number of successful f-iterations. However, since this violates the fact that $|\mathcal{S}| < \infty$, it follows at this point that we must have $|\mathcal{F}| < \infty$. Next, it follows from the facts that $v_* = 0$ and $\{\delta_k^v\} \to 0$, the last conclusion in Lemma 4.4, and (4.16) (specifically, that $v_*^{\max} > 0$) that (2.11) will be satisfied for all sufficiently large k. We may also deduce from the fact that $n_k = 0$ for all $k \geq k_s$ that (2.10) holds for all $k \geq k_s$. Since we have shown that $|\mathcal{F}| < \infty$ and that both (2.10) and (2.11) hold for sufficiently large k, we may conclude that $t_k = 0$ for all sufficiently large k. Therefore, since we have shown that $n_k = t_k = 0$ for all sufficiently large k, we have from Lemma 3.3(vi) that $k \in \mathcal{Y}$ for all sufficiently large k, which combined with Lemma 3.3(vii) implies that $\{\pi_k^f\} \to 0$. However, this contradicts our earlier conclusion that $\pi_k^f \geq \epsilon_\pi > 0$ for all $k \geq k_s$. Overall, we have contradicted the supposition that $|\mathcal{V}| = \infty$.

Next, suppose that $|\mathcal{F}| < \infty$. Combining this with the fact that $|\mathcal{V}| < \infty$ ensures that $k \in \mathcal{Y}$ for all sufficiently large k. It follows from this fact and Lemma 3.3(vii) that $\{\pi_k^f\} \to 0$, and that y_k, r_k, π_k^f, and χ_k^f will be computed to satisfy (3.15a), (3.15b), or (3.15c) for all sufficiently large k. During the computation of these quantities, (3.15a) can never be satisfied, since in that case the algorithm would terminate finitely, which contradicts the supposition of the proof. Hence, since (3.15a) is never satisfied and

$\{\pi_k^f\} \to 0$, we may deduce that $v_* > \epsilon_v > 0$. It then follows that $\chi_*^v > 0$ (and from (3.1) that $\pi_*^v > 0$), or else for $k = k_s$ the algorithm would terminate in Step 9, which is a contradiction. Thus, $\min\{\chi_*^v, \pi_*^v, v_*\} > 0$, which with (4.16), the fact that $\{\pi_k^f\} \to 0$, and (3.2) implies that $k \in \mathcal{N}$ for all sufficiently large k. Thus, by Lemma 3.3(i), we have $n_k \neq 0$, which by Lemma 3.3(vi) contradicts our earlier conclusion that $k \in \mathcal{Y}$. Overall, we have proven that we cannot have $|\mathcal{F}| < \infty$, so we must have $|\mathcal{F}| = \infty$.

Since $|\mathcal{F}| = \infty$, $|\mathcal{V}| < \infty$, and $|\mathcal{S}| < \infty$, we know from (3.24) and (3.29) that $\{\delta_k^f\} \to 0$, which when combined with (3.38), the fact that $\mathcal{F} \subseteq \mathcal{T} \backslash \mathcal{T}_0$, and Lemma 4.8(i) implies that $\{\pi_k^f\}_{k \in \mathcal{F}} \to 0$. Since (3.15a), (3.15b), or (3.15c) holds for $k \in \mathcal{F} \subseteq \mathcal{T} \backslash \mathcal{T}_0$, and since the algorithm does not terminate finitely, we know that (3.15a) must not hold for all $k \in \mathcal{F}$. Combining this with the fact that $\{\pi_k^f\}_{k \in \mathcal{F}} \to 0$ implies that $v_k > \epsilon_v$ for all sufficiently large $k \in \mathcal{F}$. Hence, since $|\mathcal{F}| = \infty$, it follows from (4.16) that $v_* > \epsilon_v > 0$. We then must conclude that $\min\{v_*, \chi_*^v\} > 0$, or else for $k = k_s$ the algorithm would terminate finitely in Step 9, which is a contradiction. Also, from $\chi_*^v > 0$ and (3.1), it follows that $\pi_*^v > 0$. Since $\{\pi_k^f\}_{k \in \mathcal{F}} \to 0$, it follows that (3.15b) will be satisfied for all sufficiently large $k \in \mathcal{F}$, which implies that $t_k = 0$ and thus $k \notin \mathcal{F}$, which once again is a contradiction.

Overall, in all cases, we have reached contradictions of our supposition that Algorithm 2 does not terminate finitely, so the result is proved. □

We now bound the constraint violation following a successful v-iteration.

Lemma 4.12 *There are constants $\{\kappa_{v\pi 1}, \kappa_{v\pi 2}, \kappa_{v\pi 3}\} \subset (0, \infty)$ so that if $k \in \mathcal{S}_v$, then*

$$v_{k+1} \leq v_k - \chi_k^v \min\{\kappa_{v\pi 1}, \kappa_{v\pi 2}\pi_k^v, \kappa_{v\pi 3}\delta_k^v\}, \quad \text{and} \tag{4.17a}$$

$$v_{k+1}^{\max} \leq \max\{\kappa_{\text{t1}} v_k^{\max}, v_k - (1 - \kappa_{\text{t2}})\chi_k^v \min\{\kappa_{v\pi 1}, \kappa_{v\pi 2}\pi_k^v, \kappa_{v\pi 3}\delta_k^v\}\}, \tag{4.17b}$$

while (3.20) does not hold.

Proof Let $k \in \mathcal{S}_v$, which by the definition of $\mathcal{S}_v$ means that (2.15) holds. In particular, we have $n_k \neq 0$. Combining this fact with Lemma 3.3(ii) means that $k \in \mathcal{S}_v \cap \mathcal{N}$. It follows from this fact, (3.32), (2.13), (2.15), (3.6), and Lemma 4.3(i) that

$$v_{k+1} \leq v_k - \eta_1 \Delta m_k^{v,d} \leq v_k - \eta_1 \kappa_{\text{cd}} \Delta m_k^{v,n} \leq v_k - \eta_1 \kappa_{\text{cd}} \kappa_{\text{cn}} \chi_k^v \min\{\pi_k^v, \delta_k^v, 1 - \kappa_{\text{fbn}}\};$$

i.e., there exist $\{\kappa_{v\pi 1}, \kappa_{v\pi 2}, \kappa_{v\pi 3}\} \subset (0, \infty)$ such that (4.17a) holds. Combining this with (3.35) yields (4.17b). Note that (4.17a) and Lemma 3.7 imply (2.11) holds.

We now prove that (3.20) does not hold. To reach a contradiction, suppose that (3.20) holds, which immediately implies that $t_k \neq 0$. Lemma 3.3(iv) then implies that $k \in \mathcal{T} \backslash \mathcal{T}_0$, which combined with the fact that (3.20) is assumed to hold shows that (2.10) holds. Thus all the conditions of an f-iteration are satisfied so that $k \in \mathcal{F}$, which, since $\mathcal{V} \cap \mathcal{F} = \emptyset$, contradicts the fact that $k \in \mathcal{S}_v \subseteq \mathcal{V}$. □

We now show that, if there are infinitely many iterations, then the v-criticality measure $\min\{v_k, \chi_k^v\}$ converges to zero, at least along a subsequence of iterates.

Lemma 4.13 *If Algorithm 2 does not terminate finitely, then*

$$0 = \begin{cases} \liminf_{k \in \mathcal{S}_v} \min\{v_k, \chi_k^v\} & \text{if} \quad |\mathcal{S}_v| = \infty, \\ \liminf_{k \in \mathcal{S}_f} \min\{v_k, \chi_k^v\} & \text{if} \quad |\mathcal{S}_v| < \infty. \end{cases} \tag{4.18}$$

Proof We proceed by considering the two cases distinguished in (4.18).

Case 1: Suppose that $|\mathcal{S}_v| = \infty$. We first recall that, with Lemma 3.7, we have that $\{v_k^{\max}\}$ is monotonically decreasing and bounded below by zero. We now proceed by considering the consequences of the update (3.35), which is applied for all $k \in \mathcal{S}_v$. Since $|\mathcal{S}_v| = \infty$, if (3.35) sets $v_{k+1}^{\max} \leq \kappa_{\text{t1}} v_k^{\max}$ infinitely often, then $\{v_k^{\max}\} \to 0$, which implies by Lemma 3.7 that $\{v_k\} \to 0$, yielding the desired limit in (4.18). Otherwise, if the update (3.35) sets $v_{k+1}^{\max} > \kappa_{\text{t1}} v_k^{\max}$ for all sufficiently large $k \in \mathcal{S}_v$, then by Lemmas 4.12 and 3.7 we have for sufficiently large $k \in \mathcal{S}_v$ that

$$v_{k+1}^{\max} \leq v_k^{\max} - (1 - \kappa_{\text{t2}}) \chi_k^v \min\{\kappa_{v\pi 1}, \kappa_{v\pi 2} \pi_k^v, \kappa_{v\pi 3} \delta_k^v\}. \tag{4.19}$$

If there is a subsequence of $\mathcal{S}_v$ along which $\{\chi_k^v\}$ converges to zero, then the first limit of (4.18) follows. Let us suppose, therefore, that $\{\chi_k^v\}_{k \in \mathcal{S}_v}$ is bounded away from zero. Then, the fact that $\{v_k^{\max}\}$ is monotonically decreasing and bounded below implies that $\{v_k^{\max} - v_{k+1}^{\max}\} \to 0$, and hence (4.19) gives that

$$\{\min\{\pi_k^v, \delta_k^v\}\}_{k \in \mathcal{S}_v} \to 0. \tag{4.20}$$

We now consider two subcases with the goal of showing that there exists a subsequence of $\{\pi_k^v\}_{k \in \mathcal{S}_v}$ that vanishes. First, suppose that $|\mathcal{S}_f| < \infty$ and let k_0 be the last index in the (ordered) set $\mathcal{S}_f$. Thus, for $k > k_0$, the inclusion $k \in \mathcal{S}$ implies $k \in \mathcal{S}_v$. As a consequence, for $k > k_0$, we have by (3.24) and (3.29) that the normal step trust region radius is only increased when $k \in \mathcal{S}_v$ and only decreased when $k \in \mathcal{V} \backslash \mathcal{S}_v$. (Here, since $|\mathcal{S}_f| < \infty$ and by the procedure for updating $\mathcal{S}_f$-`flag`, we may assume without loss of generality that Step 13 is not reached $k > k_0$). If $|\mathcal{V} \backslash \mathcal{S}_v| < \infty$, then δ_k^v is bounded away from zero due to (3.24), (3.29), and (3.34), from which (4.20) implies $\{\pi_k^v\}_{k \in \mathcal{S}_v} \to 0$. On the other hand, if $|\mathcal{V} \backslash \mathcal{S}_v| = \infty$, then, since $|\mathcal{S}_v| = \infty$, we may define the infinite set $\mathcal{K}_0$ whose elements are the indices of the first successful v-iterations following a set of iterations that includes elements of $\mathcal{V}$ but not $\mathcal{S}_v$. Consider an arbitrary $k \in \mathcal{K}_0 \subseteq \mathcal{S}_v$ with $k \geq k_0$ and define $k_u(k) \in \mathcal{V} \backslash \mathcal{S}_v$ to be the index of the last unsuccessful v-iteration before iteration k. (By convention, let $k_u(k) = k_0$ if $(\mathcal{V} \backslash \mathcal{S}_v) \cap \{k \mid k \geq k_0\} = \emptyset$.) Note that, by construction, δ_k^v is not modified between iterations $k_u(k) + 1$ and k (as these must correspond to y-iterations or unsuccessful f-iterations), which implies that $\delta_{k_u(k)+1}^v = \delta_k^v$. Moreover, the primal and slack variables are not modified between iterations $k_u(k)$ and k and thus $\pi_{k_u(k)}^v = \pi_k^v$ and $\chi_{k_u(k)}^v = \chi_k^v$. These observations, (3.36) and Lemma 4.8(ii) imply that, for $k \in \mathcal{K}_0 \subseteq \mathcal{S}_v$ sufficiently large,

$$\delta_k^v = \delta_{k_u(k)+1}^v \geq \gamma_1 \delta_{k_u(k)}^v \\ \geq \gamma_1 \min\{\kappa_{\Delta c1}, \kappa_{\Delta c2}\pi_{k_u(k)}^v, \kappa_{\Delta c3}\chi_{k_u(k)}^v\} = \gamma_1 \min\{\kappa_{\Delta c1}, \kappa_{\Delta c2}\pi_k^v, \kappa_{\Delta c3}\chi_k^v\}. \tag{4.21}$$

Now, to reach a contradiction to (4.20), suppose that there exists a subsequence $\mathcal{K}_1 \subseteq \mathcal{K}_0$ such that $\{\pi_k^v\}_{k\in\mathcal{K}_1}$ is bounded away from zero. Combining this with (4.21) and the fact that $\{\chi_k^v\}_{k\in\mathcal{S}_v}$ is assumed to be bounded away from zero (which led to (4.20)) shows that $\{\delta_k^v\}_{k\in\mathcal{K}_1}$ is bounded away from zero. This contradicts (4.20) since $\mathcal{K}_1 \subseteq \mathcal{K}_0 \subseteq \mathcal{S}_v$. Thus, we conclude that $\{\pi_k^v\}_{k\in\mathcal{K}_0} \to 0$. As a consequence, we deduce that, in this first subcase where $|\mathcal{S}_f| < \infty$, there always exists an infinite subsequence ($\mathcal{S}_v$ or $\mathcal{K}_0$) of $\mathcal{S}_v$ along which $\{\pi_k^v\}$ converges to zero.

Consider next the subcase where $|\mathcal{S}_f| = \infty$, which means that successful f- and v-iterations interlace infinitely often. In this subcase, letting $\mathcal{K}_1$ denote the infinite set whose elements are the indices of the first successful v-iterations following a set of iterations that includes elements of $\mathcal{S}_f$ but not $\mathcal{S}_v$, we may define for any $k \in \mathcal{K}_1 \subseteq \mathcal{S}_v$ the index $k_p(k)$ representing the last successful f-iteration prior to iteration k. With this definition, it follows that any iteration between $k_p(k) \in \mathcal{S}_f$ and $k \in \mathcal{S}_v$ is either a y-iteration or unsuccessful, from which it follows that

$$(x_{k_p(k)+1}, s_{k_p(k)+1}) = \cdots = (x_k, s_k) \ \text{ and } \ \pi_{k_p(k)+1}^v = \cdots = \pi_k^v.$$

On one hand, if for all sufficiently large $k \in \mathcal{K}_1$ the indices in $\{k_p(k)+1, \ldots, k-1\}$ do not belong to $\mathcal{V}$, then the only possible modification of the normal step trust region radius would be the safeguard (3.31). This and Lemma 4.3(iii) show that

$$\delta_k^v \geq \max\{\delta_k^v, \kappa_{\mathrm{n}}\|P_k^{-1}n_k^*\|\} \geq \kappa_{\mathrm{n}}\kappa_{\mathrm{cn}}\pi_k^v \ \text{ for all sufficiently large } \ k \in \mathcal{K}_1, \tag{4.22}$$

where we have used the fact that $k \in \mathcal{K}_1 \subseteq \mathcal{S}_v$ and Lemma 3.3(ii) implies that $k \in \mathcal{N}$. The inequalities in (4.22) may be followed by the same argument as that following (4.21) to conclude that $\{\pi_k^v\}_{k\in\mathcal{K}_1} \to 0$. On the other hand, if for infinitely many k any element of $\{k_p(k)+1, \ldots, k-1\}$ is an element of $\mathcal{V}\backslash\mathcal{S}_v$, then along this subsequence we may define $k_u(k) \in \mathcal{V}\backslash\mathcal{S}_v$ to be the index of the last unsuccessful v-iteration before iteration k. Then, using the same reasoning as in the first subcase, we may conclude that (4.21) holds. Employing (4.21) and (4.22) and applying similar arguments, we conclude that a subsequence of $\{\pi_k^v\}$ vanishes.

We have obtained from the two above subcases that there exists an infinite subsequence $\mathcal{K} \subseteq \mathcal{S}_v$ with $\{\pi_k^v\}_{k\in\mathcal{K}} \to 0$, regardless of the cardinality of $\mathcal{S}_f$. The fact that $\{\chi_k^v\}_{k\in\mathcal{S}_v}$ is bounded away from zero and (3.1) then imply that $\{v_k\}_{k\in\mathcal{K}} \to 0$, ensuring the desired limit in (4.18).

Case 2: Suppose that $|\mathcal{S}_v| < \infty$. In this case, by the fact that $v_{k+1}^{\max} < v_k^{\max}$ only when $k \in \mathcal{S}_v$, there exists a constant $v_\infty^{\max} > 0$ such that $v_k^{\max} = v_\infty^{\max}$ for all sufficiently large k. By Theorem 4.11, the assumption that Algorithm 2 does not terminate finitely, and $|\mathcal{S}_v| < \infty$, it follows that $|\mathcal{S}_f| = \infty$. Now, to derive a contradiction, suppose that there exists a constant $\varepsilon_{\min} > 0$ such that

$$\min\{v_k, \chi_k^v\} \geq \varepsilon_{\min} \ \text{ for all sufficiently large } k. \tag{4.23}$$

Since $|\mathcal{S}_v| < \infty$, we know from (3.24) for $k \in \mathcal{Y}$, from (2.12), (3.25), and (3.29) for $k \in \mathcal{F}$, from (3.36) for $k \in \mathcal{V}\backslash\mathcal{S}_v$, and the fact that the slack reset only possibly decreases the barrier function that $\{f(x_k, s_k)\}$ is monotonically decreasing. Moreover, it follows from Assumptions 1.1 and 4.1 and Lemma 4.2 that $\{f(x_k, s_k)\}$ is bounded below, so overall we have that $\{f(x_k, s_k)\} \to f_{\text{low}}$ for some $f_{\text{low}} > -\infty$. It follows from this fact, $|\mathcal{S}_f| = \infty$, (2.12), (3.25), (2.10) (which holds for $k \in \mathcal{F}$), (3.19a)/(3.23a), and Lemma 4.3(ii) that $\{\min\{\pi_k^f, \delta_k^t\}\}_{k\in\mathcal{S}_f} \to 0$. Suppose that for some infinite index set $\mathcal{K}_3 \subseteq \mathcal{S}_f$ and scalar $\pi_{\min}^f > 0$ we have $\pi_k^f \geq \pi_{\min}^f$ for all $k \in \mathcal{K}_3$. It follows that $\{\delta_k^t\}_{k\in\mathcal{K}_3} \to 0$. However, from Lemma 4.10 and (4.23), it follows that $\{\delta_k^v\}_{k\in\mathcal{V}}$ is bounded away from zero for all k. Combining this with the facts that $\{\delta_k^t\}_{k\in\mathcal{K}_3} \to 0$ and $v_k^{\max} = v_\infty^{\max} > 0$ for all sufficiently large k implies from (3.38) that $\{\delta_k^f\}_{k\in\mathcal{K}_3} \to 0$. It then follows from Lemma 4.9 that there exists an infinite index set $\mathcal{K}_4 \subseteq \mathcal{F}$ such that $\{\pi_k^f\}_{k\in\mathcal{K}_4} \to 0$. Since $\mathcal{K}_4 \subseteq \mathcal{F} \subseteq \mathcal{T}\backslash\mathcal{T}_0$, we know that (3.15a), (3.15b), or (3.15c) is satisfied for all $k \in \mathcal{K}_4$. However, we also know that (3.15a) cannot be satisfied since Algorithm 2 is assumed not to terminate finitely. It does, however, follow from $\{\pi_k^f\}_{k\in\mathcal{K}_4} \to 0$ and (4.23) that (3.15b) will be satisfied for all sufficiently large $k \in \mathcal{K}_4$ so that $t_k = 0$ for all sufficiently large $k \in \mathcal{K}_4 \subseteq \mathcal{F} \subseteq \mathcal{T}\backslash\mathcal{T}_0$, which is a contradiction. Thus, we conclude that the set $\mathcal{K}_3$ cannot exist, so that $\{\pi_k^f\}_{k\in\mathcal{S}_f} \to 0$. It follows from this fact, (4.23), the definition of χ_k^v given in (3.1), and since the algorithm does not terminate finitely that (3.15b) will be satisfied (and hence $t_k = 0$) for all sufficiently large $k \in \mathcal{S}_f \subseteq \mathcal{F} \subseteq \mathcal{T}\backslash\mathcal{T}_0$, which again is a contradiction. Thus, our supposition that (4.23) held must be incorrect and therefore there is a subsequence $\mathcal{K}_5$ such that $\{\min\{v_k, \chi_k^v\}\}_{k\in\mathcal{K}_5} \to 0$. Moreover, since $|\mathcal{S}_v| < \infty$ and $|\mathcal{S}_f| = \infty$, we conclude that (4.18) holds. □

To proceed further, at $s \in \mathbb{R}^M$, we define the active and inactive sets

$$\mathcal{A}(s) := \{i \in \{1, 2, \ldots, M\} : [s]_i = 0\} \quad \text{and} \quad \mathcal{I}(s) := \{1, 2, \ldots M\}\backslash\mathcal{A}(s) \tag{4.24}$$

and denote these sets at a point $s_* \in \mathbb{R}^M$ by

$$\mathcal{A}_* := \mathcal{A}(s_*) \ \text{ and } \ \mathcal{I}_* := \mathcal{I}(s_*).$$

In addition, recalling that $P_k := \text{diag}(I, S_k)$, we define $\sigma_{\min}(x_k, s_k)$ as the smallest singular value of $(J(x_k)\ S_k)^T = (J(x_k, s_k)P_k)^T$.

Lemma 4.14 *If there exists an infinite index set $\mathcal{K}$ with $\{\min\{v_k, \chi_k^v\}\}_{k\in\mathcal{K}} \to 0$, then, for an arbitrary limit point (x_*, s_*) of $\{(x_k, s_k)\}_{k\in\mathcal{K}}$, it follows that either*

(i) $v(x_*, s_*) = 0$, i.e., (x_*, s_*) is feasible for problem (NPs), or
(ii) $\chi^v(x_*, s_*) = 0$ and x_* is an infeasible point at which the Jacobian of active constraints $J_{\mathcal{A}_*}(x_*)$ has linearly dependent rows.

Proof We consider three cases. First, suppose that $\lim_{k\in\mathcal{K}} v_k = 0$. Then, any limit point (x_*, s_*) of the sequence $\{(x_k, s_k)\}_{k\in\mathcal{K}}$ yields $v(x_*, s_*) = 0$ so that (x_*, s_*) is feasible for problem (NPs), as desired.

Second, suppose that $v_k \geq v_{\min}$ for some $v_{\min} > 0$ and all sufficiently large $k \in \mathcal{K}$. Let (x_*, s_*) be any limit point of the sequence $\{(x_k, s_k)\}_{k\in\mathcal{K}}$. Combining these facts with the slack reset procedure (c.f., (1.2)), it follows that (x_*, s_*) is infeasible for problem (NPs). Moreover, from $v_k \geq v_{\min}$ for all sufficiently large $k \in \mathcal{K}$ and the assumptions of this lemma, it follows that

$$0 = \lim_{k\in\mathcal{K}} \chi_k^v = \lim_{k\in\mathcal{K}} \frac{\|P_k J(x_k, s_k)^T c(x_k, s_k)\|_2}{\|c(x_k, s_k)\|_2} \geq \lim_{k\in\mathcal{K}} \sigma_{\min}(x_k, s_k) = \sigma_{\min}(x_*, s_*).$$

Thus, $(J(x_*)\ S_*) = J(x_*, s*)P_*$ with $P_* := \text{diag}(I, S_*)$ must have a subset of linearly dependent rows. Due to the structure of this matrix, it follows that this subset does not contain row i when $[s_*]_i > 0$; it only contains rows indexed by $\mathcal{A}_*$, and thus $J_{\mathcal{A}_*}(x_*)$ has linearly dependent rows, which proves the result.

Finally, if the first two cases do not occur, we can partition $\mathcal{K}$ into two infinite disjoint index sets, call them $\mathcal{K}_1$ and $\mathcal{K}_2$, such that for some $\varepsilon > 0$ we have

$$\lim_{k\in\mathcal{K}_1} v_k = 0, \quad \lim_{k\in\mathcal{K}_2} \chi_k^v = 0, \quad \text{and} \quad v_k \geq \varepsilon \ \text{ for } \ k \in \mathcal{K}_2. \tag{4.25}$$

Since any limit point associated with $\mathcal{K}$ must be a limit point for $\mathcal{K}_1$ and/or $\mathcal{K}_2$, it suffices to prove the result for an arbitrarily chosen limit point of $\mathcal{K}_1$ and $\mathcal{K}_2$. Any limit point (x_*, s_*) of the sequence $\{(x_k, s_k)\}_{k\in\mathcal{K}_1}$ yields $v(x_*, s_*) = 0$ so that (x_*, s_*) is feasible for problem (NPs), as desired. Next, consider any limit point of the sequence $\{(x_k, s_k)\}_{k\in\mathcal{K}_2}$, call it (x_*, s_*). We may now use the same argument as for the second case (with $\mathcal{K}$ replaced by $\mathcal{K}_2$), to conclude that (x^*, s^*) is infeasible for problem (NPs) and that $J_{\mathcal{A}_*}(x_*)$ has linearly dependent rows. □

We now prove a useful fact about our employed infeasibility measures.

Lemma 4.15 *For any infinite index set $\mathcal{K}$, we have*

$$\lim_{k\in\mathcal{K}} \min\{v_k, \chi_k^v\} = 0 \quad \textit{if and only if} \quad \lim_{k\in\mathcal{K}} \pi_k^v = 0. \tag{4.26}$$

Proof First, if $\lim_{k\in\mathcal{K}} v_k = 0$, then (4.26) follows from Lemma 4.2. Second, if $v_k \geq v_{\min}$ for some $v_{\min} > 0$ and all sufficiently large $k \in \mathcal{K}$, then it follows from (3.1) that $\{\chi_k^v\}_{k\in\mathcal{K}} \to 0$ if and only if $\{\pi_k^v\}_{k\in\mathcal{K}} \to 0$, which again establishes (4.26).

Finally, suppose that the two previous cases do not hold. To prove the “only if” implication, suppose that $\{\min\{v_k, \chi_k^v\}\}_{k\in\mathcal{K}} \to 0$. Then, as in the third case of the proof of Lemma 4.14, we can partition $\mathcal{K}$ into disjoint subsets $\mathcal{K}_1$ and $\mathcal{K}_2$ such that (4.25) holds. By Lemma 4.2, it then follows that $\{\pi_k^v\}_{k\in\mathcal{K}_1} \to 0$, and by (3.1) we must also have $\{\pi_k^v\}_{k\in\mathcal{K}_2} \to 0$. Consequently, $\{\pi_k^v\}_{k\in\mathcal{K}} \to 0$, as desired. Now, to prove the “if” implication, suppose that $\{\pi_k^v\}_{k\in\mathcal{K}} \to 0$ and, to obtain a contradiction, suppose further that there exists a constant $\epsilon > 0$ such that $\mathcal{K}_\epsilon := \{k \in \mathcal{K} : \min\{v_k, \chi_k^v\} \geq \epsilon\}$ is infinite. It then follows from the definition of χ_k^v in (3.1) that the infinite sequence $\{\pi_k^v\}_{k\in\mathcal{K}_\epsilon}$ is bounded away from zero, which is a contradiction. Hence, $\{\min\{v_k, \chi_k^v\}\}_{k\in\mathcal{K}} \to 0$. □

The relevance of having an infinite index set $\mathcal{K}$ such that (4.26) holds is elucidated in the following lemma.

Lemma 4.16 *If there exists an infinite index set* $\mathcal{K}$ *such that* $\{\pi_k^v\}_{k\in\mathcal{K}} \to 0$*, then any limit point* (x_*, s_*) *of* $\{(x_k, s_k)\}_{k\in\mathcal{K}}$ *is a first-order KKT point for* (2.4).

Proof For an arbitrary limit point (x_*, s_*) of $\{(x_k, s_k)\}_{k\in\mathcal{K}}$, it follows from Lemma 3.4 and the supposition $\{\pi_k^v\}_{k\in\mathcal{K}} \to 0$ that

$$s_* \geq 0, \quad c(x_*, s_*) \geq 0, \quad S_* c(x_*, s_*) = 0, \quad \text{and} \quad J(x_*)^T c(x_*, s_*) = 0, \tag{4.27}$$

from which it follows that (2.5) holds at (x_*, s_*). □

We now make the following assumption throughout the rest of the paper.

Assumption 4.2 If there exists an infinite index set $\mathcal{K}$ such that $\{\pi_k^v\}_{k\in\mathcal{K}} \to 0$, then, for an arbitrary limit point (x_*, s_*) of $\{(x_k, s_k)\}_{k\in\mathcal{K}}$, it follows that $\mathcal{A}_* = \emptyset$ or $J_{\mathcal{A}_*}(x_*)$ has full row rank (i.e., $\sigma_{\min}(x_*, s_*) > 0$).

An important consequence of this assumption is the following.

Lemma 4.17 *If there exists an infinite index set* $\mathcal{K}$ *such that* $\{\pi_k^v\}_{k\in\mathcal{K}} \to 0$*, then for an arbitrary limit point* (x_*, s_*) *of* $\{(x_k, s_k)\}_{k\in\mathcal{K}}$*, it follows that* $v(x_*, s_*) = 0$*, i.e.,* (x_*, s_*) *is feasible for problem* (NPs)*. Moreover,* $\{v_k\}_{k\in\mathcal{K}} \to 0$.

Proof Under the conditions of the lemma, we have from Lemma 4.16 that (4.27) holds. In particular, using the definitions in (4.24) and (4.27), we have

$$[s_*]_{\mathcal{I}_*} > 0 \quad \text{and} \quad c_{\mathcal{I}_*}(x_*) < c_{\mathcal{I}_*}(x_*, s_*) = 0; \tag{4.28a}$$

$$[s_*]_{\mathcal{A}_*} = 0 \quad \text{and} \quad c_{\mathcal{A}_*}(x_*) = c_{\mathcal{A}_*}(x_*, s_*) \geq 0. \tag{4.28b}$$

If $\mathcal{A}_* = \emptyset$, then (4.28a) implies $v(x_*, s_*) = 0$. Otherwise, by (4.27) and (4.28a),

$$0 = J(x_*)^T c(x_*, s_*) = J_{\mathcal{A}_*}(x_*)^T c_{\mathcal{A}_*}(x_*, s_*) = J_{\mathcal{A}_*}(x_*)^T c_{\mathcal{A}_*}(x_*).$$

Assumption 4.2 implies that $J_{\mathcal{A}_*}(x_*)$ has full row rank, so the above implies that $0 = c_{\mathcal{A}_*}(x_*) = c_{\mathcal{A}_*}(x_*, s_*)$. Combining this with (4.28a) yields $v(x_*, s_*) = 0$. This fact and Lemmas 4.1 and 4.2 imply that $\{v_k\}_{k\in\mathcal{K}} \to 0$. □

We now prove a crucial bound on the size of the normal step relative to π_k^v.

Lemma 4.18 *Let* $k \in \mathcal{N}$ *and define* $m_k^{v,P}(a) := \|c(x_k, s_k) + J(x_k, s_k)P_k a\|_2$*. If* $\sigma_{\min}(x_k, s_k) > 0$ *and* a_k *is any (nonzero) vector satisfying*

$$m_k^{v,P}(a_k) < m_k^{v,P}(0) \quad \textit{with} \quad a_k \textit{ belonging to the range of } P_k J(x_k, s_k)^T, \tag{4.29}$$

then

$$\|a_k\|_2 \leq \frac{2}{\sigma_{\min}(x_k, s_k)^2} \pi_k^v. \tag{4.30}$$

In particular,

$$\|P_k^{-1} n_k\|_2 \leq \frac{2}{\sigma_{\min}(x_k, s_k)^2} \pi_k^v. \tag{4.31}$$

Proof Let $k \in \mathcal{N}$ and define the quadratic model $\widehat{m}_k^{v,P}(\cdot) := \frac{1}{2}(m_k^{v,P}(\cdot))^2$. Note that

$$\nabla_{xx}\widehat{m}_k^{v,P}(0) = P_k^T J(x_k, s_k)^T J(x_k, s_k) P_k.$$

By definition, $\sigma_{\min}(x_k, s_k)$ is the smallest eigenvalue of this matrix on the range space of $P_k J(x_k, s_k)^T$. Therefore, the second part of (4.29) yields

$$a_k^T \nabla_{xx}\widehat{m}_k^{v,P}(0) a_k \geq \sigma_{\min}(x_k, s_k)^2 \|a_k\|_2^2 > 0. \tag{4.32}$$

Let

$$t_* := \arg\min_{t \geq 0} \widehat{m}_k^{v,P}(t a_k).$$

It then follows from [5, Lemma 6.5.1] (and its proof) and (4.29) that

$$\frac{1}{2} \leq t_* = \frac{|a_k^T \nabla_x \widehat{m}_k^{v,P}(0)|}{a_k^T \nabla_{xx}\widehat{m}_k^{v,P}(0) a_k} \leq \frac{\|a_k\|_2 \, \pi_k^v}{a_k^T \nabla_{xx}\widehat{m}_k^{v,P}(0) a_k} \leq \frac{\pi_k^v}{\sigma_{\min}(x_k, s_k)^2 \|a_k\|_2}, \tag{4.33}$$

where we have used the Cauchy–Schwarz inequality to deduce the second inequality and (4.32) to deduce the third. Rewriting (4.33), we obtain (4.30). The inequality (4.31) then follows by choosing $a_k = P_k^{-1} n_k$, which is allowed by (3.7) and the observation that $m_k^{v,P}(P_k^{-1} n_k) = m_k^v(n_k) < m_k^v(0) = v_k = m_k^{v,P}(0)$. □

We next prove a result illustrating the importance of the sequence $\{\pi_k^f\}$. In particular, the result establishes that π_k^f is a valid criticality measure for (BSP).

Lemma 4.19 *If there exists an infinite index set $\mathcal{K}$ and a point (x_*, s_*) such that*

$$\lim_{k \in \mathcal{K}} \pi_k^v = 0, \quad \lim_{k \in \mathcal{K}} \pi_k^f = 0, \quad \textit{and} \quad \lim_{k \in \mathcal{K}} (x_k, s_k) = (x_*, s_*),$$

then $\{y_k\}_{k \to \mathcal{K}} \to y_*$ where (x_*, s_*, y_*) is a first-order KKT point for problem (BSP).

Proof Under the conditions of the lemma, Assumption 4.2 yields $\sigma_{\min}(x_*, s_*) > 0$, which, by continuity of $\sigma_{\min}$, implies that $\sigma_{\min}(x_k, s_k) \geq \frac{1}{2}\sigma_{\min}(x_*, s_*) > 0$ for sufficiently large k. We now claim that

$$\|P_k^{-1} n_k\|_2 \leq \frac{8}{\sigma_{\min}(x_*, s_*)^2} \pi_k^v \quad \text{for all sufficiently large} \quad k \in \mathcal{K}. \tag{4.34}$$

First, for all $k \in \mathcal{K} \backslash \mathcal{N}$, (4.34) holds since Lemma 3.3(ii) states that $n_k = 0$ for such k. On the other hand, for sufficiently large $k \in \mathcal{K} \cap \mathcal{N}$ such that $\sigma_{\min}(x_k, s_k) \geq \frac{1}{2}\sigma_{\min}(x_*, s_*)$, (4.34) follows as a result of (4.31). Thus, we have established (4.34). It now follows from Lemma 4.2, (4.34), and $\{\pi_k^v\}_{k \in \mathcal{K}} \to 0$ that

$$\lim_{k \in \mathcal{K}} n_k = 0. \tag{4.35}$$

Next, observe that

$$
\begin{aligned}
0 = \lim_{k\in\mathcal{K}} \pi_k^f &= \lim_{k\in\mathcal{K}} \left\| P_k\left(\nabla m_k^f(n_k) + J(x_k,s_k)^T y_k\right)\right\|_2 \\
&= \lim_{k\in\mathcal{K}} \left\| \begin{pmatrix} g(x_k) + \nabla_{xx}\mathcal{L}(x_k, y_k^{\text{B}})n_k^x + J(x_k)^T y_k \\ -\mu e + S_k D_k n_k^s + S_k y_k \end{pmatrix}\right\|_2 \qquad (4.36)\\
&= \lim_{k\in\mathcal{K}} \left\| \begin{pmatrix} g(x_k) + \nabla_{xx}\mathcal{L}(x_k, y_k^{\text{B}})n_k^x + J(x_k)^T y_k \\ [-\mu e + S_k D_k n_k^s + S_k y_k]_{\mathcal{A}_*} \\ [-\mu e + S_k D_k n_k^s + S_k y_k]_{\mathcal{I}_*} \end{pmatrix}\right\|_2 . \qquad (4.37)
\end{aligned}
$$

Using (4.37) (specifically the third row of the matrix inside the norm), the fact that $\{(x_k, s_k)\}_{k\in\mathcal{K}} \to (x_*, s_*)$ where $[s_*]_{\mathcal{I}_*} > 0$, (3.11), Lemma 4.2, and (4.35),

$$\lim_{k\in\mathcal{K}} [y_k]_{\mathcal{I}_*} = [\mu S_*^{-1} e]_{\mathcal{I}_*} =: [y_*]_{\mathcal{I}_*}.$$

It then follows from (4.37) (specifically the first row inside the norm), the fact that $\{(x_k, s_k)\}_{k\in\mathcal{K}} \to (x_*, s_*)$, (3.11), (3.10), Lemma 4.1, (4.35), and the fact that $\{\pi_k^v\}_{k\in\mathcal{K}} \to 0$—and hence the full rank of $J_{\mathcal{A}_*}(x_*)$ stated in Assumption 4.2—that

$$\lim_{k\in\mathcal{K}} [y_k]_{\mathcal{A}_*} = -\left[J_{\mathcal{A}_*}(x_*) J_{\mathcal{A}_*}(x_*)^T\right]^{-1} J_{\mathcal{A}_*}(x_*) \left(g(x_*) + J_{\mathcal{I}_*}(x_*)^T [y_*]_{\mathcal{I}_*}\right) =: [y_*]_{\mathcal{A}_*}.$$

We have shown that the multiplier sequence converges on $\mathcal{K}$, i.e., $\{y_k\}_{k\in\mathcal{K}} \to y_*$ for some $y_* \in \mathbb{R}^M$. Combining this with (4.36), the fact that $\{(x_k, s_k)\}_{k\in\mathcal{K}} \to (x_*, s_*)$, (3.11), (3.10), Lemma 4.1, and (4.35) proves that

$$g(x_*) + J(x_*)^T y_* = 0 \quad \text{and} \quad S_* y_* = \mu e. \qquad (4.38)$$

Now note that (4.38), Lemma 3.4, and the fact that $\mu > 0$ imply that $(s_*, y_*) > 0$. Combining this with (4.38) and the fact that the conditions of the lemma and Lemma 4.17 ensure that $v(x_*, s_*) = 0$, we have that (x_*, y_*, s_*) is a first-order KKT point for problem (BSP) as given by Definition 1.2. □

Lemmas 4.17 and 4.19 prove that, with Assumption 4.2, we obtain a first-order KKT point for problem (BSP) from any convergent subsequence over which $\{\pi_k^v\}$ and $\{\pi_k^f\}$ vanish. To prove that such a subsequence will exist, we make the following assumption henceforth, for which we define

$$\hat{s}_k = \max\{-c(x_k), 0\}. \qquad (4.39)$$

Assumption 4.3 There exist constants $\kappa_{\text{c}} > 0$ and $\kappa_{\text{J}} > 0$ independent of k such that if $v_k \le \kappa_{\text{c}}$, then, with $\hat{s}_k$ defined in (4.39), we have $\sigma_{\min}(x_k, \hat{s}_k) \ge \kappa_{\text{J}}$.

Remark 4.20 Observe that (2.1) and (4.39) imply that $\hat{s}_k \le s_k$, from which it follows that $v_k \ge v(x_k, \hat{s}_k)$ and $\sigma_{\min}(x_k, s_k) \ge \sigma_{\min}(x_k, \hat{s}_k)$. Hence, Assumption 4.3 implies that if $v_k \le \kappa_{\text{c}}$, then $\sigma_{\min}(x_k, s_k) \ge \kappa_{\text{J}}$, from which it follows that $\chi_k^v \ge \kappa_{\text{J}}$.

Our next results require the following projection operator. This operator is used for theoretical purposes only; such projections need not be computed.

Definition 4.21 Let $\text{Proj}_k(d)$ denote the projection of d onto $\text{Range}(P_k J(x_k, s_k)^T)$.

Lemma 4.22 *If* $k \in \mathcal{N}$ *and* $v_k \leq \kappa_{\text{c}}$, *then*

$$\|P_k^{-1} n_k\|_2 \leq \frac{2}{\kappa_{\text{J}}^2} \pi_k^v. \tag{4.40}$$

Moreover, there exist constants $\{\kappa_{\text{R1}}, \kappa_{\text{R2}}\} \subset (0, \infty)$ *so that if, in addition,* $k \in \mathcal{D}$, *then*

$$\|\text{Proj}_k(P_k^{-1} d_k)\|_2 \leq \frac{2}{\kappa_{\text{J}}^2} \pi_k^v \quad \text{and} \quad \Delta m_k^{v,d} \geq \kappa_{\text{J}} \min\{\kappa_{\text{R1}}, \kappa_{\text{R2}} \|\text{Proj}_k(P_k^{-1} d_k)\|_2\}. \tag{4.41}$$

Proof If $k \in \mathcal{N}$ and $v_k \leq \kappa_{\text{c}}$, inequality (4.40) is an immediate consequence of (4.31) and Assumption 4.3. Assume now that, in addition, $k \in \mathcal{D}$ and define $d_k^P := P_k^{-1} d_k$. Then, it follows from the fact that $J(x_k, s_k) P_k \text{Proj}_k(d_k^P) = J(x_k, s_k) P_k d_k^P$, Lemma 3.3(i), (3.19d), and the definition of $m_k^{v,P}$ in Lemma 4.18 that

$$\begin{aligned} m_k^{v,P}(\text{Proj}_k(d_k^P)) &= \|c(x_k, s_k) + J(x_k, s_k) P_k \text{Proj}_k(d_k^P)\|_2 \\ &= \|c(x_k, s_k) + J(x_k, s_k) P_k d_k^P\|_2 \\ &= \|c(x_k, s_k) + J(x_k, s_k) d_k\|_2 < \|c(x_k, s_k)\|_2 = m_k^{v,P}(0). \end{aligned} \tag{4.42}$$

We may then deduce from (4.30) with $a_k = \text{Proj}_k(d_k^P)$ and Assumption 4.3 that

$$\|\text{Proj}_k(P_k^{-1} d_k)\|_2 = \|\text{Proj}_k(d_k^P)\|_2 \leq \frac{2}{\kappa_{\text{J}}^2} \pi_k^v,$$

which proves the first inequality in (4.41). It also follows from Lemma 4.4 and the fact that the projection operator is nonexpansive that

$$\max\{\kappa_{\text{vf}}, 1\} \delta_k^v \geq \|P_k^{-1} d_k\|_2 \geq \|\text{Proj}_k(P_k^{-1} d_k)\|_2.$$

Combining this with $k \in \mathcal{D} \cap \mathcal{N}$, Lemma 3.3(ix), (2.15), (3.6), Lemma 4.3(i), Assumption 4.3, and the first inequality in (4.41), we have

$$\begin{aligned} \Delta m_k^{v,d} &\geq \kappa_{\text{cd}} \Delta m_k^{v,n} \geq \kappa_{\text{cd}} \kappa_{\text{cn}} \chi_k^v \min\{\pi_k^v, \delta_k^v, 1 - \kappa_{\text{fbn}}\} \\ &\geq \kappa_{\text{cd}} \kappa_{\text{cn}} \kappa_{\text{J}} \min\left\{\frac{\kappa_{\text{J}}^2 \|\text{Proj}_k(P_k^{-1} d_k)\|_2}{2}, \frac{\|\text{Proj}_k(P_k^{-1} d_k)\|_2}{\max\{\kappa_{\text{vf}}, 1\}}, 1 - \kappa_{\text{fbn}}\right\}; \end{aligned}$$

i.e., there exists $\{\kappa_{\text{R1}}, \kappa_{\text{R2}}\} \subset (0, \infty)$ for the second inequality in (4.41). □

We now prove that if the number of successful v-iterations is infinite, then, amongst other things, limit points of the sequence of iterates are feasible.

Lemma 4.23 *If* $|\mathcal{S}_v| = \infty$, *then* $\{v_k^{\max}\} \to 0$, $\{v_k\} \to 0$, $\{\pi_k^v\} \to 0$, *and* $\{n_k\} \to 0$.

Proof Since $|\mathcal{S}_v| = \infty$, it must be true that Algorithm 2 does not terminate finitely. This implies that the result of Lemma 4.13 holds true. Moreover, Lemma 3.7 shows that $\{v_k^{\max}\}$ is monotonically decreasing and bounded below by zero. Then, as in the proof of Lemma 4.13, we have that if the update (3.35) sets $v_{k+1}^{\max} \le \kappa_{\text{t1}} v_k^{\max}$ infinitely often, then $\{v_k^{\max}\} \to 0$ and $\{v_k\} \to 0$, from which it follows by Lemma 4.2 that $\{\pi_k^v\} \to 0$. It then follows from these facts and (4.40) that $\{n_k\} \to 0$.

All that remains is to consider when the update (3.35) sets $v_{k+1}^{\max} > \kappa_{\text{t1}} v_k^{\max}$ for all large k. From Lemma 4.13 we have that $\{\min\{v_k, \chi_k^v\}\}_{k\in\mathcal{K}_1} \to 0$ for some infinite $\mathcal{K}_1 \subseteq \mathcal{S}_v$, which in turn by Lemma 4.15 implies that $\{\pi_k^v\}_{k\in\mathcal{K}_1} \to 0$. Then, by Lemma 4.17, $\{v_k\}_{k\in\mathcal{K}_1} \to 0$. We then have from Lemma 4.12 [in particular, (4.17b)] that $\{v_{k+1}^{\max}\}_{k\in\mathcal{K}_1} \to 0$, which means that $\{v_k^{\max}\} \to 0$ and hence $\{v_k\} \to 0$ by Lemma 3.7. Combining this with Assumptions 1.1 and 4.1 and Lemma 4.2, we have $\{\pi_k^v\} \to 0$. It follows from this, the fact that $n_k = 0$ for $k \notin \mathcal{N}$ [see Lemma 3.3(ii)], and (4.40) that $\{n_k\} \to 0$. □

We now provide bounds for a certain type of unsuccessful v-iteration.

Lemma 4.24 *If* $k \in (\mathcal{N} \cap \mathcal{V} \cap \mathcal{D})\backslash\mathcal{S}_v$ *and*

$$v_k \le \min\left\{\kappa_{\text{c}}, \frac{\kappa_{\Delta\text{c1}}}{\kappa_{\Delta\text{c2}}\kappa_{\text{J}}}, \frac{\kappa_{\Delta\text{c3}}}{\kappa_{\Delta\text{c2}}}, \frac{1-\kappa_{\text{fbn}}}{\kappa_{\text{J}}}, \frac{1-\kappa_{\text{fbn}}}{\kappa_{\Delta\text{c2}}\kappa_{\text{J}}}\right\}, \tag{4.43}$$

then, for some constants $\{\kappa_{\text{cld}}, \kappa_{\text{sRn}}\} \subset (0,1)$, *we have*

$$m_k^v(d_k) \le \kappa_{\text{cld}} v_k \quad \textit{and} \quad \|\text{Proj}_k(P_k^{-1} d_k)\|_2 \ge \kappa_{\text{sRn}} \|P_k^{-1} n_k\|_2. \tag{4.44}$$

Proof Consider $k \in (\mathcal{N} \cap \mathcal{V} \cap \mathcal{D})\backslash\mathcal{S}_v$ such that (4.43) holds. It follows from the fact that $k \in \mathcal{N} \cap \mathcal{D}$, Lemma 3.3(ix), the inequality in (2.15), (3.6), Lemma 4.3(i), (4.43), Assumption 4.3, and (3.1) that

$$\begin{aligned} m_k^v(d_k) &\le m_k^v(0) - \kappa_{\text{cd}}\kappa_{\text{cn}}\chi_k^v \min\left\{\pi_k^v, \delta_k^v, 1-\kappa_{\text{fbn}}\right\} \\ &\le m_k^v(0) - \kappa_{\text{cd}}\kappa_{\text{cn}}\kappa_{\text{J}} \min\left\{\kappa_{\text{J}} v_k, \delta_k^v, 1-\kappa_{\text{fbn}}\right\}. \end{aligned} \tag{4.45}$$

It also follows from Lemma 4.8(ii), the fact that $k \in \mathcal{V}\backslash\mathcal{S}_v$, Assumption 4.3, (3.1), and (4.43) that

$$\delta_k^v > \min\left\{\kappa_{\Delta\text{c1}}, \kappa_{\Delta\text{c2}}\pi_k^v, \kappa_{\Delta\text{c3}}\chi_k^v\right\} \ge \min\left\{\kappa_{\Delta\text{c1}}, \kappa_{\Delta\text{c2}}\kappa_{\text{J}} v_k, \kappa_{\Delta\text{c3}}\kappa_{\text{J}}\right\} = \kappa_{\Delta\text{c2}}\kappa_{\text{J}} v_k.$$

Substituting this into (4.45) ensures with (4.43) the existence of $\kappa_{\text{cld}} \in (0,1)$ independent of k such that

$$\begin{aligned} m_k^v(d_k) &\le m_k^v(0) - \kappa_{\text{cd}}\kappa_{\text{cn}}\kappa_{\text{J}} \min\left\{\kappa_{\text{J}} v_k, \kappa_{\Delta\text{c2}}\kappa_{\text{J}} v_k, 1-\kappa_{\text{fbn}}\right\} \\ &= v_k - \kappa_{\text{cd}}\kappa_{\text{cn}}\kappa_{\text{J}} \min\left\{\kappa_{\text{J}}, \kappa_{\Delta\text{c2}}\kappa_{\text{J}}\right\} v_k \le \kappa_{\text{cld}} v_k. \end{aligned}$$

This is the first desired result. Defining $d_k^P := P_k^{-1} d_k$, we may use the inequality above, the triangle inequality, and $J(x_k, s_k) P_k d_k^P = J(x_k, s_k) P_k \text{Proj}_k(d_k^P)$ to get

$$v_k - \|J(x_k, s_k) P_k \mathrm{Proj}_k(d_k^P)\|_2 \leq \|c(x_k, s_k) + J(x_k, s_k) P_k \mathrm{Proj}_k(d_k^P)\|_2$$
$$= \|c(x_k, s_k) + J(x_k, s_k) P_k d_k^P\|_2 = m_k^v(d_k) \leq \kappa_{\text{cld}} v_k.$$

Combining the above, $k \in \mathcal{N}$, (4.43), (4.40), and norm inequalities shows that

$$\begin{aligned}\|P_k^{-1} n_k\|_2 &\leq \frac{2}{\kappa_{\text{J}}^2} \pi_k^v \leq \frac{2}{\kappa_{\text{J}}^2} \|P_k J(x_k, s_k)^T\|_2\, v_k \\ &\leq \frac{2}{\kappa_{\text{J}}^2} \|P_k J(x_k, s_k)^T\|_2 \frac{\|J(x_k, s_k) P_k \mathrm{Proj}_k(d_k^P)\|_2}{1 - \kappa_{\text{cld}}} \\ &\leq \frac{2}{\kappa_{\text{J}}^2} \|P_k J(x_k, s_k)^T\|_2 \frac{\|J(x_k, s_k) P_k\|_2 \|\mathrm{Proj}_k(d_k^P)\|_2}{1 - \kappa_{\text{cld}}}.\end{aligned}$$

It then follows from the definition of d_k^P, Lemma 4.2, and the fact that $\kappa_{\text{cld}} \in (0, 1)$ that for some $\kappa_{\text{sRn}} \in (0, 1)$ independent of k, we have

$$\|\mathrm{Proj}_k(P_k^{-1} d_k)\|_2 \geq \frac{(1 - \kappa_{\text{cld}})\kappa_{\text{J}}^2}{2\|J(x_k, s_k) P_k\|_2^2} \|P_k^{-1} n_k\|_2 \geq \kappa_{\text{sRn}} \|P_k^{-1} n_k\|_2,$$

which is the second desired result. □

For our next pair of results, we define the constants

$$\varsigma_{\text{tn}} := \kappa_{\text{tn}} \max\left\{1, \frac{2\kappa_{\text{ub}}}{(1 - \kappa_\delta)(\kappa_{\text{tn}} - 1)\kappa_{\text{ct}}(1 - \kappa_{\text{B}})\epsilon_\pi}\right\} > 1 \quad \text{and} \tag{4.46a}$$

$$\varsigma_\delta := \min\left\{1, \frac{\epsilon_\pi}{1 - \kappa_{\text{B}}}, \frac{(1 - \kappa_{\text{fbt}})\kappa_{\text{bfn}}}{1 - \kappa_{\text{B}}}\right\} \in (0, 1]. \tag{4.46b}$$

Lemma 4.25 *If $k \notin \mathcal{Y}$ such that*

$$\pi_k^f \geq \epsilon_\pi > 0, \tag{4.47a}$$

$$\min\{\kappa_{\text{vf}} \delta_k^v, \delta_k^f\} \leq \varsigma_\delta, \quad \textit{and} \tag{4.47b}$$

$$\|P_k^{-1} t_k\|_2 \geq \varsigma_{\text{tn}} \|P_k^{-1} n_k\|_2, \tag{4.47c}$$

then $t_k \neq 0$ and (2.10) holds.

Proof Let $k \notin \mathcal{Y}$ be such that (4.47) holds. If $k \in \mathcal{F}$, the results follow by the definition of the index set $\mathcal{F}$. Thus, for the remainder of the proof, assume $k \in \mathcal{V}$.

If $n_k = 0$, then $t_k \neq 0$ (since otherwise $k \in \mathcal{Y}$ by Lemma 3.3(vi)), so that by (3.19a)/(3.23a) and Lemma 4.3(ii), we have $\Delta m_k^{f,d} = \Delta m_k^{f,t} \geq 0$, meaning that (2.10) holds, as desired. Otherwise, if $n_k \neq 0$, then since $s_k > 0$ and $P_k \succ 0$ for all k

and (4.47c) holds, we have $t_k \neq 0$, which implies $k \in \mathcal{T}\backslash\mathcal{T}_0$ and (3.12) holds. It then follows from the triangle inequality, (4.47c), and (4.46a) that

$$\begin{aligned} \|P_k^{-1}d_k\|_2 &\geq \|P_k^{-1}t_k\|_2 - \|P_k^{-1}n_k\|_2 \\ &= \left(1 - \frac{\|P_k^{-1}n_k\|_2}{\|P_k^{-1}t_k\|_2}\right)\|P_k^{-1}t_k\|_2 \geq \left(\frac{\kappa_{\text{tn}}-1}{\kappa_{\text{tn}}}\right)\|P_k^{-1}t_k\|_2. \end{aligned} \tag{4.48}$$

We also have that

$$\begin{aligned} -\Delta m_k^{f,n} &= \nabla f(x_k, s_k)^T n_k + \tfrac{1}{2}n_k^T G_k n_k \\ &= \left(P_k \nabla f(x_k, s_k)\right)^T P_k^{-1} n_k + \tfrac{1}{2}(P_k^{-1}n_k)^T P_k G_k P_k (P_k^{-1}n_k). \end{aligned} \tag{4.49}$$

Using the triangle and Cauchy-Schwarz inequalities, Lemma 4.2, and the fact that (3.12), (4.47b) and (4.46b) imply $\|P_k^{-1}n_k\|_2 \leq \min\{\kappa_{\text{vf}}\delta_k^v, \delta_k^f\} \leq 1$, we then have

$$|\Delta m_k^{f,n}| \leq \kappa_{\text{ub}}(\|P_k^{-1}n_k\|_2 + \tfrac{1}{2}\|P_k^{-1}n_k\|_2^2) \leq 2\kappa_{\text{ub}}\|P_k^{-1}n_k\|_2. \tag{4.50}$$

Moreover, it follows from the fact that $k \in \mathcal{T}\backslash\mathcal{T}_0$, Lemma 4.3(ii), (4.47a), (3.38), (4.47b), and (4.46b) that

$$\Delta m_k^{f,t} \geq \kappa_{\text{ct}}\epsilon_\pi \min\{\epsilon_\pi, (1-\kappa_{\text{B}})\delta_k^t, (1-\kappa_{\text{fbt}})\kappa_{\text{fbn}}\} = \kappa_{\text{ct}}\epsilon_\pi(1-\kappa_{\text{B}})\delta_k^t. \tag{4.51}$$

Combining (4.51), (4.50), $k \in \mathcal{T}\backslash\mathcal{T}_0$, Lemma 4.4, (4.48), (4.47c), and (4.46a) yields

$$\begin{aligned} \frac{|\Delta m_k^{f,n}|}{\Delta m_k^{f,t}} &\leq \frac{2\kappa_{\text{ub}}\|P_k^{-1}n_k\|_2}{\kappa_{\text{ct}}\epsilon_\pi(1-\kappa_{\text{B}})\delta_k^t} \leq \frac{2\kappa_{\text{ub}}\|P_k^{-1}n_k\|_2}{\kappa_{\text{ct}}\epsilon_\pi(1-\kappa_{\text{B}})\|P_k^{-1}d_k\|_2} \\ &\leq \frac{2\kappa_{\text{ub}}\kappa_{\text{tn}}}{\kappa_{\text{ct}}\epsilon_\pi(1-\kappa_{\text{B}})(\kappa_{\text{tn}}-1)} \frac{\|P_k^{-1}n_k\|_2}{\|P_k^{-1}t_k\|_2} \leq 1 - \kappa_\delta. \end{aligned}$$

Hence, (2.10) holds, which completes the proof. □

We next prove that if the primal iterate is nearly feasible, then certain v-iterations will be successful.

Lemma 4.26 *If $k \in \mathcal{V} \cap \mathcal{D}$,*

$$\|P_k^{-1}t_k\|_2 \leq \varsigma_{\text{tn}}\|P_k^{-1}n_k\|_2, \tag{4.52}$$

and

$$v_k \leq \min\left\{\kappa_{\text{c}}, \frac{\kappa_{\Delta\text{c1}}}{\kappa_{\Delta\text{c2}}\kappa_{\text{J}}}, \frac{\kappa_{\Delta\text{c3}}}{\kappa_{\Delta\text{c2}}}, \frac{1-\kappa_{\text{fbn}}}{\kappa_{\text{J}}}, \frac{1-\kappa_{\text{fbn}}}{\kappa_{\Delta\text{c2}}\kappa_{\text{J}}}, \frac{\kappa_{\text{R1}}\kappa_{\text{J}}^2}{2\kappa_{\text{R2}}\kappa_{\text{sRn}}\kappa_{\text{ub}}}, \frac{\kappa_{\text{J}}^3\kappa_{\text{R2}}\kappa_{\text{sRn}}(1-\eta_1)}{2\kappa_{\text{C}}(1+\varsigma_{\text{tn}})^2\kappa_{\text{ub}}}\right\} \tag{4.53}$$

then $k \in \mathcal{S}_v$ and $\delta_{k+1}^v \geq \delta_k^v$.

Proof Consider $k \in \mathcal{V} \cap \mathcal{D}$ such that (4.52) and (4.53) hold. If $n_k = 0$, then (4.52) implies that $t_k = 0$, which in turn implies by Lemma 3.3(vi) that $k \in \mathcal{Y}$. However, this contradicts the supposition that $k \in \mathcal{V}$, so we must have $n_k \neq 0$. In this case, Lemma 3.3(ii) ensures that $k \in \mathcal{N}$, so that overall we have $k \in \mathcal{N} \cap \mathcal{V} \cap \mathcal{D}$.

To obtain a contradiction, suppose that $k \notin \mathcal{S}_v$, so that overall we have $k \in (\mathcal{N} \cap \mathcal{V} \cap \mathcal{D}) \backslash \mathcal{S}_v$. This and the bound (4.53) imply that the results of Lemmas 4.22 and 4.24 hold, i.e., that (4.41) and (4.44) hold. Moreover, $k \in \mathcal{D}$ and Lemma 3.3(ix) imply that (2.15) holds. Using this and the facts that $n_k \neq 0$ and $k \in \mathcal{V} \backslash \mathcal{S}_v$, it follows from (3.36) that $\rho_k^v < \eta_1$. However, since (4.41) and (4.44) hold,

$$\Delta m_k^{v,d} \geq \kappa_{\text{J}} \min\{\kappa_{\text{R1}}, \kappa_{\text{R2}} \|\text{Proj}_k(P_k^{-1} d_k)\|_2\} \geq \kappa_{\text{J}} \min\{\kappa_{\text{R1}}, \kappa_{\text{R2}} \kappa_{\text{sRn}} \|P_k^{-1} n_k\|_2\}.$$

In fact, it follows from (4.40), Lemma 4.2 and (4.53) that

$$\kappa_{\text{R2}} \kappa_{\text{sRn}} \|P_k^{-1} n_k\|_2 \leq \frac{2\kappa_{\text{R2}} \kappa_{\text{sRn}}}{\kappa_{\text{J}}^2} \pi_k^v \leq \frac{2\kappa_{\text{R2}} \kappa_{\text{sRn}} \kappa_{\text{ub}}}{\kappa_{\text{J}}^2} v_k \leq \kappa_{\text{R1}},$$

and thus

$$\Delta m_k^{v,d} \geq \kappa_{\text{J}} \kappa_{\text{R2}} \kappa_{\text{sRn}} \|P_k^{-1} n_k\|_2. \tag{4.54}$$

Furthermore, by (2.13), (4.2), (4.54), the triangle inequality, (4.52), (4.40), the Cauchy-Schwarz inequality, Lemma 4.2, and (4.53), we have that

$$\begin{aligned} |\rho_k^v - 1| &= \left| \frac{v(x_k + d_k^x, s_k + d_k^s) - m_k^v(d_k)}{\Delta m_k^{v,d}} \right| \leq \frac{\kappa_{\text{C}} \|P_k^{-1} d_k\|_2^2}{\kappa_{\text{J}} \kappa_{\text{R2}} \kappa_{\text{sRn}} \|P_k^{-1} n_k\|_2} \\ &\leq \frac{\kappa_{\text{C}} (1 + \varsigma_{\text{tn}})^2 \|P_k^{-1} n_k\|_2}{\kappa_{\text{J}} \kappa_{\text{R2}} \kappa_{\text{sRn}}} \leq \frac{2\kappa_{\text{C}} (1 + \varsigma_{\text{tn}})^2 \kappa_{\text{ub}}}{\kappa_{\text{J}}^3 \kappa_{\text{R2}} \kappa_{\text{sRn}}} v_k \leq 1 - \eta_1, \end{aligned}$$

and hence $\rho_k^v \geq \eta_1$, which is a contradiction. Thus, we must conclude that $k \in \mathcal{S}_v$. The fact that $\delta_{k+1}^v \geq \delta_k^v$ now follows from the fact that $k \in \mathcal{S}_v$ and (3.34). □

We now prove finite termination when the set of successful v-iterations is finite.

Lemma 4.27 *If $|\mathcal{S}_v| < \infty$, then Algorithm 2 terminates finitely.*

Proof We prove the result by contradiction, and so suppose that $|\mathcal{S}_v| < \infty$, but that Algorithm 2 does not terminate finitely. It then follows from Theorem 4.11 that $|\mathcal{S}| = \infty$, which when combined with the fact that $|\mathcal{S}_v| < \infty$ implies that $|\mathcal{S}_f| = \infty$; i.e., it follows that there are an infinite number of successful iterations, and all belong to $\mathcal{S}_f$ for all sufficiently large k. We may also deduce from these facts—and since the barrier function is decreased for $k \in \mathcal{S}_f$ and the slack reset only possibly decreases the barrier function—that the sequence $\{f(x_k, s_k)\}$ is monotonically decreasing for sufficiently large k. Moreover, since $v_{k+1}^{\max} \leftarrow v_k^{\max}$ for all $k \notin \mathcal{S}_v$ and $|\mathcal{S}_v| < \infty$, we have that there exists a constant $v_\infty^{\max} > 0$ such that

$$v_k^{\max} = v_\infty^{\max} > 0 \ \text{ for all sufficiently large } \ k. \tag{4.55}$$

We now consider two cases depending on whether, for some $\epsilon_f > 0$, (4.11) holds.

Case 1: Suppose that (4.11) holds for some $\epsilon_f > 0$. It then follows from Lemma 4.9 that (4.12) also holds, in which case we have from (3.19a)/(3.23a), the fact that $\mathcal{S}_f \subseteq \mathcal{F} \subseteq \mathcal{T}\backslash\mathcal{T}_0$, Lemma 4.3(ii), (4.11), (4.12), (3.38), and (4.55) that

$$\begin{aligned}\Delta m_k^{f,t} &\geq \kappa_{\rm ct}\pi_k^f \min\{\pi_k^f, (1-\kappa_{\rm B})\delta_k^t, (1-\kappa_{\rm fbt})\kappa_{\rm fbn}\} \\ &\geq \kappa_{\rm ct}\epsilon_f \min\{\epsilon_f, (1-\kappa_{\rm B})\delta_k^t, (1-\kappa_{\rm fbt})\kappa_{\rm fbn}\} \\ &\geq \kappa_{\rm ct}\epsilon_f \min\{\epsilon_f, (1-\kappa_{\rm B})\min\{\kappa_{\rm vf}\delta_k^v, \epsilon_{\mathcal{F}}, \kappa_{\rm v}v_\infty^{\max}\}, (1-\kappa_{\rm fbt})\kappa_{\rm fbn}\} \end{aligned} \tag{4.56}$$

for sufficiently large $k \in \mathcal{S}_f$. We now consider two subcases, deriving contradictions in each, which will prove that the condition of this case (namely, that there exists $\epsilon_f > 0$ such that (4.11) holds) cannot occur.

Subcase 1.1: Suppose there exists an infinite subsequence $\mathcal{K}_f \subseteq \mathcal{S}_f$ such that $\{\delta_k^v\}_{k\in\mathcal{K}_f} \to 0$. Since $\delta_{k+1}^v < \delta_k^v$ only if $k \in \mathcal{V}\backslash\mathcal{S}_v$ and $\delta_{k+1}^v \leftarrow \delta_k^v$ otherwise (and any potential reset of δ_k^v in Step 13 increases its value), it follows that there exists an infinite subsequence $\mathcal{K}_v \subseteq \mathcal{V}\backslash\mathcal{S}_v$ such that $\{\delta_k^v\}_{k\in\mathcal{K}_v} \to 0$. Our goal in the remainder of this subcase is to prove that for all sufficiently large $k \in \mathcal{K}_v \subseteq \mathcal{V}$, we have that all of the conditions of an f-iteration are satisfied, which is a contradiction since $\mathcal{V}\cap\mathcal{F} = \emptyset$. This will prove that such a sequence $\mathcal{K}_f \subseteq \mathcal{S}_f$ cannot exist.

Using the fact that $\{\delta_k^v\}_{k\in\mathcal{K}_v} \to 0$ and Lemma 4.6, we may conclude that, for all sufficiently large $k \in \mathcal{K}_v$, we have $k \in (\mathcal{V}\cap\mathcal{D})\backslash\mathcal{S}_v$. In addition, since $|\mathcal{S}_v| < \infty$ and $\{\delta_k^v\}_{k\in\mathcal{K}_v} \to 0$, we may conclude from Lemma 4.8(ii) and Lemma 4.15 that $\{\pi_k^v\}_{k\in\mathcal{K}_v} \to 0$, which in turn implies with Lemma 4.17 that $\{v_k\}_{k\in\mathcal{K}_v} \to 0$. Now, suppose that there exists an infinite subsequence $\mathcal{K}'_v \subseteq \mathcal{K}_v$ such that $\mathcal{K}'_v \cap \mathcal{N} = \emptyset$. The following then hold for all sufficiently large $k \in \mathcal{K}'_v \subseteq \mathcal{K}_v \subseteq \mathcal{V}\backslash\mathcal{S}_v$:

(a) $n_k = 0$ by Lemma 3.3(ii) (and thus (2.10) holds);
(b) $t_k \neq 0$ by (a), Lemma 3.3(vi), and the fact that $k \in \mathcal{V}$; and
(c) $v_k < \kappa_{\rm vv}v_k^{\max} = \kappa_{\rm vv}v_\infty^{\max}$ by Step 10, (3.2), and (4.55).

It then follows from Assumption 1.1, Lemma 4.4, the fact that $\{\delta_k^v\}_{k\in\mathcal{K}'_v} \to 0$, statement (c) above, and the bound $\kappa_{\rm vv} < 1$ that $v(x_k+d_k^x, s_k+d_k^s) \leq v_k^{\max}$ for all sufficiently large $k \in \mathcal{K}'_v$. Overall, this yields (2.11), and thus we have that all of the conditions of an f-iteration hold, so $k \in \mathcal{F}$. However, this is a contradiction since $k \in \mathcal{K}'_v \subseteq \mathcal{V}$ and $\mathcal{V}\cap\mathcal{F} = \emptyset$. Thus, such an infinite subsequence $\mathcal{K}'_v \subseteq \mathcal{K}_v$ cannot exist, so we may conclude that for all sufficiently large $k \in \mathcal{K}_v$ we have $k \in \mathcal{N}$. To summarize, at this point in this subcase, we may assume without loss of generality that there exists an infinite subsequence $\mathcal{K}_v \subseteq (\mathcal{N}\cap\mathcal{V}\cap\mathcal{D})\backslash\mathcal{S}_v$ over which $\{\delta_k^v\}_{k\in\mathcal{K}_v} \to 0$, $\{\pi_k^v\}_{k\in\mathcal{K}_v} \to 0$, and $\{v_k\}_{k\in\mathcal{K}_v} \to 0$.

It follows from Lemma 4.24, $\mathcal{K}_v \subseteq (\mathcal{N}\cap\mathcal{V}\cap\mathcal{D})\backslash\mathcal{S}_v$, and $\{v_k\}_{k\in\mathcal{K}_v} \to 0$ that $m_k^v(d_k) \leq \kappa_{\rm cld}v_k$ for all sufficiently large $k \in \mathcal{K}_v$. Using this fact, (4.2), the triangle inequality, Lemmas 4.4, 3.7, and (4.55), we have

$$v(x_k^+, s_k^+) \leq \kappa_{\rm cld}v_\infty^{\max} + \kappa_{\rm C}(\delta_k^v)^2 \quad \text{for all sufficiently large} \quad k \in \mathcal{K}_v.$$

This then implies that $v(x_k^+, s_k^+) \leq v_\infty^{\max} = v_k^{\max}$ for all sufficiently large $k \in \mathcal{K}_v$ such that $(\delta_k^v)^2 \leq ((1-\kappa_{\rm cld})/\kappa_{\rm C})v_\infty^{\max}$. Thus, since $\{\delta_k^v\}_{k\in\mathcal{K}_v} \to 0$, we may conclude that (2.11) holds for all sufficiently large $k \in \mathcal{K}_v$.

Next, suppose that for $\varsigma_{\text{tn}} > 0$ defined in (4.46a), we have

$$\|P_k^{-1} t_k\|_2 \le \varsigma_{\text{tn}} \|P_k^{-1} n_k\|_2 \ \text{ for all sufficiently large } \ k \in \mathcal{K}_v. \tag{4.57}$$

We may then use $\mathcal{K}_v \subseteq (\mathcal{N} \cap \mathcal{V} \cap \mathcal{D})$, $\{v_k\}_{k \in \mathcal{K}_v} \to 0$, (4.57), and Lemma 4.26 to conclude that $|\mathcal{S}_v \cap \mathcal{K}_v| = \infty$, which contradicts the fact that $|\mathcal{S}_v| < \infty$. Therefore, there exists an infinite subsequence $\mathcal{K}_v'' \subseteq \mathcal{K}_v$ such that if $k \in \mathcal{K}_v''$ then (4.57) fails.

We now show that with $k \in \mathcal{K}_v'' \subseteq \mathcal{K}_v \subseteq \mathcal{V} \backslash \mathcal{S}_v$, the conditions of Lemma 4.25 hold. Consider $k \in \mathcal{K}_v''$. First, since $k \in \mathcal{K}_v'' \subseteq \mathcal{V}$, we know that $k \notin \mathcal{Y}$. Second, since $k \in \mathcal{K}_v''$, we know from the previous paragraph that (4.57) does not hold, and therefore that $t_k \neq 0$ and r_k was computed to satisfy (3.15a), (3.15b), or (3.15c). Since we have supposed that the algorithm does not terminate finitely, we may use the fact that $\{v_k\}_{k \in \mathcal{K}_v} \to 0$ along with (3.15a) to conclude that (4.47a) holds for all sufficiently large $k \in \mathcal{K}_v''$. Third, since $\{\delta_k^v\}_{k \in \mathcal{K}_v} \to 0$, we have that (4.47b) holds for all sufficiently large $k \in \mathcal{K}_v''$. Fourth, we know from the definition of the set $\mathcal{K}_v''$ that (4.57) fails, which is to say that (4.47c) holds. We may now apply Lemma 4.25 to deduce that $t_k \neq 0$ and (2.10) holds for all sufficiently large $k \in \mathcal{K}_v''$. Thus, along with our previous conclusion that (2.11) holds for all sufficiently large $k \in \mathcal{K}_v$, we conclude that for all sufficiently large $k \in \mathcal{K}_v''$ we have that all of the conditions of an f-iteration are satisfied. However, as previously mentioned, this is impossible since $\mathcal{K}_v'' \subseteq \mathcal{K}_v \subseteq \mathcal{V}$ and $\mathcal{F} \cap \mathcal{V} = \emptyset$. Thus, our supposition for Subcase 1.1 that there is an infinite subsequence $\mathcal{K}_f \subseteq \mathcal{S}_f$ with $\{\delta_k^v\}_{k \in \mathcal{K}_f} \to 0$, is impossible.

Subcase 1.2: Suppose that there exists $\epsilon_* > 0$ such that $\delta_k^v \ge \epsilon_*$ for all $k \in \mathcal{S}_f$, and recall that $|\mathcal{S}_f| = \infty$. We may combine (4.56) and $\delta_k^v \ge \epsilon_*$ for all $k \in \mathcal{S}_f$ to conclude that there exists k' such that, for all $k \ge k'$ with $k \in \mathcal{S}_f$, we have

$$\Delta m_k^{f,t} \ge \kappa_{\text{ct}} \epsilon_f \min\left\{\epsilon_f, (1-\kappa_{\text{B}}) \min\{\kappa_{\text{vf}} \epsilon_*, \epsilon_{\mathcal{F}}, \kappa_{\text{v}} v_\infty^{\max}\}, (1-\kappa_{\text{fbt}})\kappa_{\text{fbn}}\right\} > 0. \tag{4.58}$$

Combining $|\mathcal{S}_v| < \infty$, $|\mathcal{S}_f| = \infty$, (2.12), and (2.10) (which holds for $k \in \mathcal{F}$) yields

$$f(x_{k'}, s_{k'}) - f(x_k, s_k) = \sum_{j=k', j \in \mathcal{S}_f}^{k-1} [f(x_j, s_j) - f(x_{j+1}, s_{j+1})] \ge \eta_1 \kappa_\delta \sum_{j=k', j \in \mathcal{S}_f}^{k-1} \Delta m_j^{f,t}, \tag{4.59}$$

which with (4.58) proves that $\{f(x_k, s_k)\} \to -\infty$. However, this is a contradiction since f is bounded below by Lemma 4.2 and Assumptions 1.1 and 4.1.

Since neither Subcase 1.1 nor 1.2 can occur, it follows that Case 1 cannot occur.

Case 2: Suppose that there exists $\mathcal{K} \subseteq \mathcal{F}$ with

$$\lim_{k \in \mathcal{K}} \pi_k^f = 0. \tag{4.60}$$

For all $k \in \mathcal{K} \subseteq \mathcal{F} \subseteq \mathcal{T} \backslash \mathcal{T}_0$, we have that $t_k \neq 0$ was computed (and not reset to zero), in which case (3.15b) must not hold. Combining this with (4.60) shows that $0 = \lim_{k \in \mathcal{K}} \pi_k^f \ge \lim_{k \in \mathcal{K}} \omega_t(\pi_k^v) \ge 0$, so that $\{\pi_k^v\}_{k \in \mathcal{K}} = 0$. Hence, by Lemma 4.17,

$\{v_k\}_{k\in\mathcal{K}} \to 0$, which when combined with (4.60) shows that (3.15a) will be satisfied for all sufficiently large $k \in \mathcal{K}$. However, this contradicts our supposition that the algorithm does not terminate finitely. □

The previous result proves that if the algorithm does not terminate finitely, then there are an infinite number of successful v-iterations. We now establish an important consequence of this fact.

Lemma 4.28 *If $|\mathcal{S}_v| = \infty$ and (4.52) holds for all sufficiently large $k \in \mathcal{V} \cap \mathcal{D}$, then*

$$\delta_k^v \geq \epsilon_* \ \text{for some} \ \ \epsilon_* > 0 \ \text{for all} \ k. \tag{4.61}$$

Proof First, by Lemma 4.23, the fact that $|\mathcal{S}_v| = \infty$ implies that $\{v_k\} \to 0$. Hence, for sufficiently large $k \in \mathcal{V} \cap \mathcal{D}$, we have that (4.52) and (4.53) hold, which implies by Lemma 4.26 that $\delta_{k+1}^v \geq \delta_k^v$. Second, if $k \in \mathcal{V}\backslash\mathcal{D}$, then it follows from Lemma 4.6 that $\kappa_{\text{vf}}\delta_k^v \geq \min\{\kappa_{\text{vf}}\delta_k^v, \delta_k^f, \kappa_{\text{v}} v_k^{\max}\} > \kappa_{\mathcal{V}}$. Third, if $k \in \mathcal{Y} \cup \mathcal{F}$, then by (3.24), (3.28), and (3.29) we have that $\delta_{k+1}^v \geq \delta_k^v$. The result follows by combining these facts. □

We now prove a result about certain v-iterations that are unsuccessful.

Lemma 4.29 *If $k \in \mathcal{V}\backslash\mathcal{S}_v$, (4.43) holds,*

$$v_k^{\max} \leq \min\left\{\left(\frac{1-\kappa_{\text{cld}}}{\kappa_{\text{C}}}\right)^2, \left(\frac{1-\kappa_{\text{vv}}}{\kappa_{\text{C}}}\right)^2, \left(\frac{\kappa_{\mathcal{V}}}{\kappa_{\text{vf}}}\right)^{\frac{4}{3}}\right\}, \tag{4.62}$$

and

$$\delta_k^v \leq (v_k^{\max})^{\frac{3}{4}} \tag{4.63}$$

then $k \in \mathcal{D}$ and (2.11) holds.

Proof Let $k \in \mathcal{V}\backslash\mathcal{S}_v$ and observe that (4.62) and (4.63) imply that $\kappa_{\text{vf}}\delta_k^v \leq \kappa_{\mathcal{V}}$. Hence, by Lemma 4.6, we have that $k \in \mathcal{D}$. That is, $k \in (\mathcal{V} \cap \mathcal{D})\backslash\mathcal{S}_v$. We now consider two cases depending on whether or not $k \in \mathcal{N}$.

Suppose $k \in \mathcal{N}$ so that $k \in (\mathcal{N}\cap\mathcal{V}\cap\mathcal{D})\backslash\mathcal{S}_v$. It then follows from (4.2), the triangle inequality, the fact that (4.43) holds, and Lemmas 4.4 and 4.24 that

$$v(x_k + d_k^x, s_k + d_k^s) \leq \kappa_{\text{cld}} v_k + \kappa_{\text{C}} (\delta_k^v)^2.$$

Then, from this inequality, Lemma 3.7, (4.63), and (4.62), we have that

$$\begin{aligned} v(x_k + d_k^x, s_k + d_k^s) &\leq \kappa_{\text{cld}} v_k^{\max} + \kappa_{\text{C}}\, (v_k^{\max})^{\frac{3}{2}} \\ &= v_k^{\max}\left(\kappa_{\text{cld}} + \kappa_{\text{C}}\sqrt{v_k^{\max}}\right) \leq v_k^{\max}, \end{aligned}$$

which means that (2.11) holds, as desired.

Now suppose $k \notin \mathcal{N}$ (so that $n_k = 0$). It then follows from (4.2), the triangle inequality, Lemmas 4.4 and 3.7, (3.19d) (which holds since $k \in \mathcal{D}$), and the fact that $v_k < \kappa_{\text{vv}} v_k^{\max}$ (which holds by (3.2) since $k \notin \mathcal{N}$), (4.62), and (4.63) that

$$v(x_k + d_k^x, s_k + d_k^s) \le m_k^v(d_k) + \kappa_{\rm C}(\delta_k^v)^2$$
$$\le \kappa_{\rm vv} v_k^{\rm max} + \kappa_{\rm C}\,(v_k^{\rm max})^{\frac{3}{2}} \le v_k^{\rm max}\left(\kappa_{\rm vv} + \kappa_{\rm C}\sqrt{v_k^{\rm max}}\right) \le v_k^{\rm max},$$

which means that (2.11) holds, as desired. □

We now prove that there are a finite number of successful v-iterations.

Theorem 4.30 *The set $\mathcal{S}_v$ is finite.*

Proof We prove the result by contradiction, and so suppose that $|\mathcal{S}_v| = \infty$. It then follows from Lemma 4.23 that $\{v_k^{\rm max}\} \to 0$, $\{v_k\} \to 0$, $\{\pi_k^v\} \to 0$, and $\{n_k\} \to 0$. Moreover, since $|\mathcal{S}_v| = \infty$, we have that (3.15a) must not hold for all sufficiently large k, or else the algorithm would terminate finitely in Step 21 or 35, which is a contradiction. Thus, since $\{v_k\} \to 0$, we have

$$\pi_k^f \ge \epsilon_\pi > 0 \;\text{ for all sufficiently large }\; k. \tag{4.64}$$

It follows from this fact and Lemma 4.9 that (4.12) holds. Also it follows from the facts that $\{v_k\} \to 0$, $\{v_k^{\rm max}\} \to 0$, and $|\mathcal{S}_v| = \infty$ that there exists k_0 such that (4.43), (4.53), and (4.62) hold for all $k \ge k_0$.

We now prove a lower bound for δ_k^v that holds for all sufficiently large k, written as equation (4.68) below. We prove the bound by considering two cases.

Case 1: Suppose that (4.52) holds for all sufficiently large $k \ge k_0$ such that $k \in \mathcal{V} \cap \mathcal{D}$. Then, since $|\mathcal{S}_v| = \infty$, we may apply Lemma 4.28 to deduce that (4.61) holds for all sufficiently large k.

Case 2: Suppose that there exists an infinite index set

$$\mathcal{K}_1 := \{k \ge k_0 : k \in \mathcal{V} \cap \mathcal{D} \;\text{ and }\; \|P_k^{-1} t_k\|_2 > \varsigma_{\rm tn} \|P_k^{-1} n_k\|_2\}.$$

Since δ_k^v ($v_k^{\rm max}$) is not decreased (increased) for $k \in \mathcal{S}_v \cup \mathcal{Y} \cup \mathcal{F}$, our goal is to provide a lower bound for δ_k^v over $k \in \mathcal{K}_1 \backslash \mathcal{S}_v$. We do this by considering two subcases.
Subcase 1: Consider k such that $k_0 \le k \in \mathcal{K}_1 \backslash (\mathcal{S}_v \cup \mathcal{N})$. Since $k \notin \mathcal{N}$, it follows from Lemma 3.3(ii) that $n_k = 0$. By Lemma 3.3(vi), this means that $t_k \ne 0$ (since otherwise we would have $k \in \mathcal{Y}$), which in turn means by Lemma 3.3(v) that $k \in \mathcal{T} \backslash \mathcal{T}_0$ and that (2.10) holds (since $n_k = 0$). We may then conclude from the fact that $k \in \mathcal{V} \backslash \mathcal{S}_v$, the choice of k_0 being large enough such that (4.43) and (4.62) hold for $k \ge k_0$, and Lemma 4.29 that if (4.63) holds, then (2.11) also holds. However, this would imply that $k \in \mathcal{F}$, which contradicts the definition of $\mathcal{K}_1$ since $\mathcal{V} \cap \mathcal{F} = \emptyset$. Thus, (4.63) must not hold and

$$\delta_k^v > (v_k^{\rm max})^{\frac{3}{4}} \;\text{ for all }\; k \;\text{ such that }\; k_0 \le k \in \mathcal{K}_1 \backslash (\mathcal{S}_v \cup \mathcal{N}). \tag{4.65}$$

Subcase 2: Consider k such that $k_0 \le k \in (\mathcal{K}_1 \cap \mathcal{N}) \backslash \mathcal{S}_v$. By (4.64), we have that (4.47a) holds. Similarly, by the definition of $\mathcal{K}_1$, we have that (4.47c) holds. Now suppose that (4.47b) and (4.63) both hold. Then, since $k \notin \mathcal{Y}$ and (4.47a), (4.47b), and (4.47c) all hold, we may apply Lemma 4.25 to conclude that $t_k \neq 0$ and (2.10) holds. Also, since $k \in \mathcal{V}\backslash\mathcal{S}_v$, we have shown that (4.43) and (4.62) hold, and we have supposed that (4.63) holds, we may apply Lemma 4.29 to conclude that (2.11) holds. Overall, we have shown that all of the conditions of an f-iteration are satisfied so that $k \in \mathcal{F}$. However, this contradicts the fact that $k \in \mathcal{K}_1 \subseteq \mathcal{V}$ and $\mathcal{V} \cap \mathcal{F} = \emptyset$. Therefore, at least one of (4.47b) or (4.63) must not hold, yielding

$$\delta_k^v > \min\left\{\frac{\varsigma_\delta}{\kappa_{\text{vf}}}, (v_k^{\max})^{\frac{3}{4}}\right\} \text{ for all } k \text{ such that } k_0 \leq k \in (\mathcal{K}_1 \cap \mathcal{N})\backslash\mathcal{S}_v. \tag{4.66}$$

Combining (4.65)/(4.66) from Subcases 1/2 shows that, for Case 2, we have

$$\delta_k^v \geq \min\left\{\frac{\varsigma_\delta}{\kappa_{\text{vf}}}, (v_k^{\max})^{\frac{3}{4}}\right\} \text{ for all } k \text{ such that } k_0 \leq k \in \mathcal{K}_1\backslash\mathcal{S}_v. \tag{4.67}$$

Moreover, the fact that $\{v_k\} \to 0$ and Lemma 4.26 implies that for any k with $k_0 \leq k \in (\mathcal{V} \cap \mathcal{D})\backslash\mathcal{K}_1$, we have $k \in \mathcal{S}_v$. Thus, for all $k \geq k_0$ with $k \in (\mathcal{V} \cap \mathcal{D})\backslash\mathcal{S}_v$, we have $k \in \mathcal{K}_1\backslash\mathcal{S}_v$. As a result, the inequality in (4.67) holds for all k with $k_0 \leq k \in (\mathcal{V} \cap \mathcal{D})\backslash\mathcal{S}_v$. This conclusion, along with the deduction that $\kappa_{\text{vf}}\delta_k^v > \kappa_{\mathcal{V}}$ for all $k \in \mathcal{V}\backslash\mathcal{D}$ from Lemma 4.6 yields

$$\delta_k^v \geq \min\left\{\frac{\varsigma_\delta}{\kappa_{\text{vf}}}, (v_k^{\max})^{\frac{3}{4}}, \frac{\kappa_{\mathcal{V}}}{\kappa_{\text{vf}}}\right\} \text{ for all } k \text{ with } k_0 \leq k \in \mathcal{V}\backslash\mathcal{S}_v,$$

which, when combined with the fact that δ_k^v (resp. $v_k^{\max}$) is not decreased (resp. increased) for $k \in \mathcal{S}_v \cup \mathcal{Y} \cup \mathcal{F}$, yields

$$\delta_k^v \geq \min\left\{\frac{\varsigma_\delta}{\kappa_{\text{vf}}}, (v_k^{\max})^{\frac{3}{4}}, \frac{\kappa_{\mathcal{V}}}{\kappa_{\text{vf}}}\right\} \text{ for all } k \geq k_0.$$

Combining the results of Cases 1 and 2, we have that

$$\kappa_{\text{vf}}\delta_k^v \geq \min\left\{\kappa_{\text{vf}}\epsilon_*, \varsigma_\delta, \kappa_{\text{vf}}(v_k^{\max})^{\frac{3}{4}}, \kappa_{\mathcal{V}}\right\} \text{ for all sufficiently large } k. \tag{4.68}$$

Using this fact, (4.12), and $\{v_k^{\max}\} \to 0$ yields

$$\min\{\kappa_{\text{vf}}\delta_k^v, \delta_k^f\} \geq \kappa_{\text{vf}}(v_k^{\max})^{\frac{3}{4}} \text{ for large } k. \tag{4.69}$$

Under our supposition that the set $\mathcal{S}_v$ is infinite, at least one of the following two scenarios must occur. In both, we reach a contradiction to this supposition that $\mathcal{S}_v$ is infinite, which proves the theorem.

Scenario 1: Suppose that $\mathcal{S}_1 := \mathcal{S}_v\backslash\mathcal{T}$ is infinite. For $k \in \mathcal{S}_1$, we have that either (3.12) does not hold or (3.15b) holds. In fact, since (4.64) holds and $\{\pi_k^v\} \to 0$, condition (3.15b) cannot hold infinitely often for $k \in \mathcal{S}_1$, implying that for all sufficiently large $k \in \mathcal{S}_1$ we have that (3.12) does not hold. Then, since $t_k = 0$ for $k \in \mathcal{S}_1 \subseteq \mathcal{V}$, we have by Lemma 3.3(vi) that $n_k \neq 0$ (or else $k \in \mathcal{Y}$). We may now use the facts that

$v_k^{\max} > 0$, $\delta_k^v > 0$, and $\delta_k^f > 0$ for all k, (4.40), (4.69), Lemmas 3.7 and 4.2, and the fact that $\{v_k\} \to 0$ to conclude that, for sufficiently large $k \in \mathcal{S}_1$,

$$\frac{\|P_k^{-1}n_k\|_2}{\min\{\kappa_{\text{vf}}\delta_k^v, \delta_k^f\}} \leq \frac{2\pi_k^v}{\kappa_{\text{J}}^2\kappa_{\text{vf}}(v_k^{\max})^{\frac{3}{4}}} \leq \frac{2\kappa_{\text{ub}}v_k}{\kappa_{\text{J}}^2\kappa_{\text{vf}}(v_k)^{\frac{3}{4}}} = \frac{2\kappa_{\text{ub}}}{\kappa_{\text{J}}^2\kappa_{\text{vf}}} v_k^{\frac{1}{4}} \leq \kappa_{\text{B}}.$$

However, this means that (3.12) holds for all sufficiently large $k \in \mathcal{S}_1$, contradicting our earlier conclusion that it does not. Thus, this scenario cannot occur.

Scenario 2: Suppose that $\mathcal{S}_2 = \mathcal{S}_v \cap \mathcal{T}$ is infinite. Our goal is to show that for all sufficiently large $k \in \mathcal{S}_2$, we have that all of the conditions of an f-iteration are satisfied, which is impossible since $\mathcal{S}_2 \subseteq \mathcal{V}$ and $\mathcal{V} \cap \mathcal{F} = \emptyset$. We begin by showing that (2.10) holds for all sufficiently large $k \in \mathcal{S}_2$. To do this, first note that since $\mathcal{S}_2 \subseteq \mathcal{S}_v \subseteq \mathcal{N}$ and $\{v_k\} \to 0$, we may apply the result of Lemma 4.22 for sufficiently large $k \in \mathcal{S}_2$. Then, using (4.49), the triangle and Cauchy-Schwarz inequalities, Lemma 4.2, (3.1b), and that $\{\pi_k^v\} \to 0$ (implying in turn that $2\pi_k^v \leq \kappa_{\text{J}}^2$ and thus, in view of (4.40), that $\|P_k^{-1}n_k\|_2 \leq 1$ for all sufficiently large k), it follows as in the proof of Lemma 4.25 (see (4.50)) that

$$|\Delta m_k^{f,n}| \leq \kappa_{\text{ub}}(\|P_k^{-1}n_k\|_2 + \tfrac{1}{2}\|P_k^{-1}n_k\|_2^2) \leq \frac{4\kappa_{\text{ub}}}{\kappa_{\text{J}}^2}\pi_k^v \leq \frac{4\kappa_{\text{ub}}^2}{\kappa_{\text{J}}^2} v_k \tag{4.70}$$

for all sufficiently large $k \in \mathcal{S}_2$. It also follows from $\{v_k^{\max}\} \to 0$, $\mathcal{S}_2 \subseteq \mathcal{V}$, and Lemma 4.6 that $k \in \mathcal{D}$ for all sufficiently large $k \in \mathcal{S}_2$. Moreover, since $\mathcal{S}_2 \subseteq \mathcal{T}$, it follows that for all $k \in \mathcal{S}_2$ a tangential step $t_k \neq 0$ was computed to satisfy either (3.19) or (3.23). However, for all $k \in \mathcal{S}_2$, it follows from (2.15) that $n_k \neq 0$, and then from Lemma 3.3(xi) that $k \in \mathcal{T}_{\mathcal{D}}$, i.e., that (3.19) holds. This implies by (3.38) that $\delta_k^t = \min\{\kappa_{\text{vf}}\delta_k^v, \delta_k^f\}$ for all sufficiently large $k \in \mathcal{S}_2$. Combining this with $k \in \mathcal{T}_{\mathcal{D}}$, (3.19a), Lemma 4.3(ii), (4.64), (4.69), $\{v_k^{\max}\} \to 0$, and Lemma 3.7 gives, for all sufficiently large $k \in \mathcal{S}_2$,

$$\begin{aligned}\Delta m_k^{f,t} &\geq \kappa_{\text{ct}}\epsilon_\pi \min\left\{\epsilon_\pi, (1-\kappa_{\text{B}})\delta_k^t, (1-\kappa_{\text{fbt}})\kappa_{\text{fbn}}\right\}\\ &= \kappa_{\text{ct}}\epsilon_\pi \min\left\{\epsilon_\pi, (1-\kappa_{\text{B}})\min\{\kappa_{\text{vf}}\delta_k^v, \delta_k^f\}, (1-\kappa_{\text{fbt}})\kappa_{\text{fbn}}\right\}\\ &\geq \kappa_{\text{ct}}\epsilon_\pi(1-\kappa_{\text{B}})\kappa_{\text{vf}}(v_k^{\max})^{\frac{3}{4}} \geq \kappa_{\text{ct}}\epsilon_\pi(1-\kappa_{\text{B}})\kappa_{\text{vf}}v_k^{\frac{3}{4}}.\end{aligned}$$

Combining this with (4.70) and $\{v_k\} \to 0$ shows that

$$\frac{|\Delta m_k^{f,n}|}{\Delta m_k^{f,t}} \leq \frac{4\kappa_{\text{ub}}^2 v_k^{\frac{1}{4}}}{\kappa_{\text{ct}}\epsilon_\pi(1-\kappa_{\text{B}})\kappa_{\text{vf}}\kappa_{\text{J}}^2} \leq 1-\kappa_\delta \quad \text{for all sufficiently large} \quad k \in \mathcal{S}_2.$$

Hence, (2.10) holds for sufficiently large $k \in \mathcal{S}_2$, as desired. From here, it follows from Step 30 that the computed tangential step is not reset to zero, i.e., $k \in \mathcal{T}_{\mathcal{D}} \backslash \mathcal{T}_0$ for all sufficiently large $k \in \mathcal{S}_2$, from which it follows that $t_k \neq 0$ for all sufficiently large $k \in \mathcal{S}_2$. Moreover, since $k \in \mathcal{S}_v$ implies by Lemma 3.7 that (2.11) holds, we have from the fact that $\mathcal{S}_2 \subseteq \mathcal{S}_v$ that (2.11) holds for all $k \in \mathcal{S}_2$. To summarize, we have shown that for all sufficiently large $k \in \mathcal{S}_2$, all conditions of an f-iteration are satisfied, which is a contradiction. Thus, this scenario cannot occur.

Overall, we have shown that under our supposition that $|\mathcal{S}_v| = \infty$, neither Scenario 1 nor 2 may occur. However, since one of them must occur when $|\mathcal{S}_v| = \infty$, we have reached a contradiction to our supposition, and the result is proved. □

We conclude by summarizing our convergence results.

Theorem 4.31 *The following hold for Algorithm 2:*

(i) *If Assumptions 1.1, 3.1, and 4.1 hold, then either Algorithm 2 terminates finitely or there exists an infinite index set $\mathcal{K}$ such that $\lim_{k \in \mathcal{K}} \min\{v_k, \chi_k^v\} = \lim_{k \in \mathcal{K}} \pi_k^v = 0$. In the latter case, any limit point (x_*, s_*) of $\{(x_k, s_k)\}_{k \in \mathcal{K}}$ satisfies $\pi^v(x_*, s_*) = 0$ and is therefore a critical point of minimizing $\frac{1}{2}v(x,s)^2$ subject to $s \geq 0$.*

(ii) *If Assumptions 1.1, 3.1, 4.1, and 4.2 hold, then either Algorithm 2 terminates finitely or there exists an infinite index set $\mathcal{K}$ such that $\lim_{k \in \mathcal{K}} \min\{v_k, \chi_k^v\} = \lim_{k \in \mathcal{K}} \pi_k^v = 0$. In the latter case, any limit point (x_*, s_*) of $\{(x_k, s_k)\}_{k \in \mathcal{K}}$ satisfies $v(x_*, s_*) = 0$ so that (x_*, s_*) is feasible for (NPs).*

(iii) *If Assumptions 1.1, 3.1, 4.1, 4.2, and 4.3 hold, then either Algorithm 2 terminates finitely in Step 9 with an infeasible stationary point (x_k, s_k) with $v_k > \kappa_c$ or it terminates finitely in Step 21 or 35 with an approximate first-order KKT point (x_k, s_k, y_k) for the barrier problem (BSP).*

Proof Part (i) follows from Lemmas 4.13, 4.15, and 4.16. Part (ii) follows from part (i) and Lemma 4.17. Also, it follows from Theorem 4.30 and Lemma 4.27 that Algorithm 2 terminates finitely. Thus, part (iii) follows since, under Assumption 4.3, a subsequence cannot converge to an infeasible stationary point with $v_k \leq \kappa_c$. (For this last conclusion, recall Remark 4.20.) □

5 A trust-funnel algorithm for the nonlinear optimization problem

The previous section considers the global convergence properties of our trust-funnel algorithm when applied to solve the barrier subproblem (BSP). This section describes how a sequence of barrier subproblems with decreasing values for the barrier parameter may be solved to find a first-order KKT point for (NPs).

To achieve our stated goal, we require the constants ϵ_π and ϵ_v in Algorithm 2 to depend on μ. Moreover, for practical reasons, it is advisable to make other constants in Algorithm 2 depend on μ as well. In the previous section, for ease of exposition, we did not explicitly state these dependencies since μ was fixed. This does not pose a problem in this section since we use Algorithm 2 to solve a sequence of barrier problems where for each particular instance the barrier parameter is fixed and therefore our previous analysis still holds. A summary of the constants that depend on μ and precisely where they are used is given in Table 1. In addition to requiring them to be positive, it is appropriate to have them satisfy

$$\lim_{\mu \to 0} \epsilon_\pi(\mu) = \lim_{\mu \to 0} \epsilon_v(\mu) = \lim_{\mu \to 0} \kappa_{\text{fbn}}(\mu) = \lim_{\mu \to 0} \kappa_{\text{fbt}}(\mu) = 0 \quad \text{and} \tag{5.1}$$

Table 1 Parameters for Algorithm 2 that depend on μ

Parameter	Used	Parameter	Used	Parameter	Used
$\kappa_y = \kappa_y(\mu)$	(3.10)	$\kappa_D = \kappa_D(\mu)$	(3.11)	$\epsilon_\pi = \epsilon_\pi(\mu)$	(3.15a)
$\kappa_{fbt} = \kappa_{fbt}(\mu)$	(3.19b)/(3.23b)	$\kappa_{fbn} = \kappa_{fbn}(\mu)$	(2.2)/(3.5)	$\epsilon_v = \epsilon_v(\mu)$	(3.15a)

$$\lim_{\mu\to 0} \kappa_y(\mu) = \lim_{\mu\to 0} \kappa_D(\mu) = \infty. \tag{5.2}$$

Moreover, the convergence result that we present additionally assumes that

$$\epsilon_\pi(\mu_j) \le \zeta_1 \mu_j^\alpha \quad \text{and} \quad \epsilon_v(\mu_j) \le \zeta_2 \mu_j^\beta \tag{5.3}$$

for some $\zeta_1 \in (0, 1)$, $\{\zeta_2, \beta\} \subset (0, \infty)$, $\alpha \ge 1$, and that a particular choice for the positive-definite matrix D_k in (3.11) is used; specifically, for each $1 \le i \le m$, let

$$[d_k]_i := [D_k]_{ii} := \begin{cases} \kappa_D(\mu_j) & \text{if} \quad \mu_j [s_k]_i^{-2} > \kappa_D(\mu_j), \\ \mu_j [s_k]_i^{-2} & \text{otherwise.} \end{cases} \tag{5.4}$$

Other choices are possible, e.g., based on the primal-dual update $D_k = Y_k S_k^{-1}$, and only require a small modification in the proof.
With these requirements, we now state our method for solving problem (NPs).

Algorithm 3 Trust-funnel algorithm for solving (NPs).

1: **Input**: (x_0, s_0, y_0, μ_0) satisfying $(s_0, y_0, \mu_0) > 0$.
2: Choose a parameter $\gamma_\mu \in (0, 1)$ and forcing functions $\epsilon_\pi(\cdot)$ and $\epsilon_v(\cdot)$.
3: Set $(x_0^{\text{start}}, s_0^{\text{start}}, y_0^{\text{start}}) \leftarrow (x_0, s_0, y_0)$ and $j \leftarrow 0$.
4: **for** $j = 0, 1, \ldots$ **do**
5: Obtain $(x_{j+1}, s_{j+1}, y_{j+1})$=BSP$(x_j^{\text{start}}, s_j^{\text{start}}, y_j^{\text{start}}, \mu_j, \epsilon_\pi(\mu_j), \epsilon_v(\mu_j))$ from Algorithm 2.
6: **if** Algorithm 2 terminated in Step 9 **then**
7: Return the infeasible stationary point (x_{j+1}, s_{j+1}).
8: Set $\mu_{j+1} \in (0, \gamma_\mu \mu_j]$.
9: Use μ_j, μ_{j+1}, and $(x_{j+1}, s_{j+1}, y_{j+1})$ to compute the starting point $(x_{j+1}^{\text{start}}, s_{j+1}^{\text{start}}, y_{j+1}^{\text{start}})$.

Theorem 5.1 *If Assumptions 1.1, 3.1, 4.1, 4.2, and 4.3 hold with (5.3)–(5.4), then*

(i) *Algorithm 3 returns an infeasible stationary point in Step 7, or*
(ii) *there exists a limit point (x_*, s_*, y_*) of the iterates $\{(x_{j+1}, s_{j+1}, y_{j+1})\}$ computed by Algorithm 3 such that (x_*, s_*, y_*) is a first-order KKT point for problem* (NPs).

Proof If statement (i) occurs, then there is nothing left to prove. Therefore, suppose that statement (i) does not occur, in which case we have that Algorithm 2 never terminates in Step 9, which by (3.15a) and (5.3) means that for all $j \ge 0$ we have

$$\pi_{j+1}^f(y_{j+1}) \le \epsilon_\pi(\mu_j) \le \zeta_1 \mu_j^\alpha \quad \text{and} \quad v_{j+1} \le \epsilon_v(\mu_j) \le \zeta_2 \mu_j^\beta. \tag{5.5}$$

In particular, we have that the sequence $\{(x_{j+1}, s_{j+1}, y_{j+1})\}$ is infinite, and from the second part of (5.5), the triangle inequality, and Assumption 4.1, that $\{s_{j+1}\}$ is bounded. Combining this fact with Assumption 4.1 implies the existence of an infinite index set $\mathcal{J}$ and a point (x_*, s_*) with $s_* \geq 0$ such that

$$\lim_{j\in\mathcal{J}} (x_{j+1}, s_{j+1}) = (x_*, s_*). \tag{5.6}$$

It follows from this fact, (5.5), $\mu_j \to 0$, and Assumption 1.1 that

$$\lim_{j\in\mathcal{J}} v_{j+1} = v(x_*, s_*) = 0. \tag{5.7}$$

We comment that for the remainder of the proof, the quantities P_{j+1}, n_{j+1}, etc. are used to represent the final values of the relevant quantities computed in Algorithm 2 when it is called in line 5 during iteration j of Algorithm 3; they are the complementary quantities to $(x_{j+1}, s_{j+1}, y_{j+1})$.

It follows from norm inequalities, the definition of P_{j+1}, (4.40), the fact that $n_j = 0$ if $j \notin \mathcal{N}$ (see Lemma 3.3(ii)), (3.1), (5.6), (5.7), Assumption 1.1, and (5.5) that, for all $i \in \{1, 2, \ldots, m\}$, we have

$$\begin{aligned}\left|\frac{[n^s_{j+1}]_i}{[s_{j+1}]_i}\right| &\leq \|S^{-1}_{j+1} n^s_{j+1}\|_2 \leq \|P^{-1}_{j+1} n_{j+1}\|_2 \leq \frac{2}{\kappa_{\text{J}}^2}\pi^v_{j+1} \\ &= \mathcal{O}(v_{j+1}) = \mathcal{O}(\mu_j^\beta) \quad \text{for} \quad j \in \mathcal{J}.\end{aligned}$$

Since we maintain positive slacks throughout Algorithm 2, we may conclude that

$$|[n^s_{j+1}]_i| = \mathcal{O}(\mu_j^\beta [s_{j+1}]_i) \;\; \text{for all} \quad 1 \leq i \leq m \text{ and } j \in \mathcal{J}. \tag{5.8}$$

We now develop a crucial bound by considering two cases motivated by (5.4). First, suppose that for a given i we have $\mu_j [s_{j+1}]_i^{-2} \leq \kappa_{\text{D}}(\mu_j)$, so that from (5.4) we have $[d_{j+1}]_i = \mu_j [s_{j+1}]_i^{-2}$. It then follows from this fact and (5.8) that

$$|[s_{j+1}]_i [d_{j+1}]_i [n^s_{j+1}]_i| = \mathcal{O}(\mu_j^{1+\beta}) \quad \text{for} \quad j \in \mathcal{J}.$$

Second, suppose that for a given i we have $\mu_j [s_{j+1}]_i^{-2} > \kappa_{\text{D}}(\mu_j)$, so that from (5.4) we have $[d_{j+1}]_i = \kappa_{\text{D}}(\mu_j) < \mu_j [s_{j+1}]_i^{-2}$, and thus $[s_{j+1}]_i^2 [d_{j+1}]_i < \mu_j$. Combining this fact with (5.8) shows that

$$|[s_{j+1}]_i [d_{j+1}]_i [n^s_{j+1}]_i| = \mathcal{O}(\mu_j^\beta [s_{j+1}]_i^2 [d_{j+1}]_i) = \mathcal{O}(\mu_j^{1+\beta}) \quad \text{for} \quad j \in \mathcal{J}. \tag{5.9}$$

Therefore, (5.9) holds in both cases, i.e., (5.9) holds for all $1 \leq i \leq m$ and $j \in \mathcal{J}$. We may now use the same proof as for Lemma 4.19, combined with (5.7), (5.9), and the first part of (5.5) to deduce that $\lim_{j\in\mathcal{J}} y_{j+1} = y_*$ for some y_* satisfying

$g(x_*) + J(x_*)^T y_* = 0$ and $S_* y_* = 0$. To prove that (x_*, s_*, y_*) is a first-order KKT point for problem (NPs), it only remains to prove that $y^* \geq 0$, as we do next.

From the first part of (5.5), we know that

$$\begin{aligned}\zeta_1 \mu_j^\alpha &\geq \left\| \begin{pmatrix} g(x_{j+1}) + \nabla_{xx}\mathcal{L}(x_{j+1}, y_{j+1}^{\mathrm{B}}) n_{j+1}^x + J(x_{j+1})^T y_{j+1} \\ -\mu_j e + S_{j+1} D_{j+1} n_{j+1}^s + S_{j+1} y_{j+1} \end{pmatrix} \right\|_2 \\ &\geq \left\| -\mu_j e + S_{j+1} D_{j+1} n_{j+1}^s + S_{j+1} y_{j+1} \right\|_2 \\ &\geq | -\mu_j + [s_{j+1}]_i [d_{j+1}]_i [n_{j+1}^s]_i + [s_{j+1}]_i [y_{j+1}]_i | \quad \text{for all} \quad 1 \leq i \leq m. \end{aligned} \tag{5.10}$$

We now consider two cases. First, suppose that i is such that $[s_*]_i > 0$. In this case it follows from (5.10), (5.9), the fact that $\mu_j \to 0$, and (5.6) that $\lim_{j \in \mathcal{J}} [y_{j+1}]_i = [y^*]_i = 0$, as desired. Second, suppose that i is such that $[s_*]_i = 0$. It may be observed from (5.10) that $-\zeta_1 \mu_j^\alpha \leq -\mu_j + [s_{j+1}]_i [d_{j+1}]_i [n_{j+1}^s]_i + [s_{j+1}]_i [y_{j+1}]_i$, so

$$[y_{j+1}]_i \geq \frac{-\zeta_1 \mu_j^\alpha + \mu_j - [s_{j+1}]_i [d_{j+1}]_i [n_{j+1}^s]_i}{[s_{j+1}]_i}. \tag{5.11}$$

It follows from (5.11), $\zeta_1 \in (0, 1)$, $\alpha \geq 1$, $\beta > 0$, $\mu_j \to 0$, (5.9), and the positivity of the slack variables as imposed in Algorithm 2, that $[y_{j+1}]_i > 0$ for all sufficiently large $j \in \mathcal{J}$. Combining this with $\lim_{j \in \mathcal{J}} y_{j+1} = y_*$ shows that $[y_*]_i \geq 0$. □

6 Conclusion and discussion

In this paper, we have presented a new algorithm for solving constrained nonlinear optimization problems. The algorithm is of the inexact barrier-SQP variety, i.e., it approximately solves a sequence of barrier subproblems using an inexact SQP method. In Sects. 3 and 4, we proved that each barrier subproblem could be solved approximately using a new inexact-SQP method based on a trust-funnel mechanism (not requiring a filter or penalty function). The algorithm is extremely flexible in that, during each iteration, it automatically determines the types of steps and updates that are expected to be most productive, where potential productivity is determined by available criticality measures. In each iteration, each subproblem may be solved approximately using matrix-free iterative methods, which means that the algorithm is viable for solving large-scale barrier subproblems. We then proved in Sect. 5 that an approximate solution of the original nonlinear optimization problem may be obtained by approximately solving a sequence of barrier subproblems for a decreasing sequence of barrier parameters.

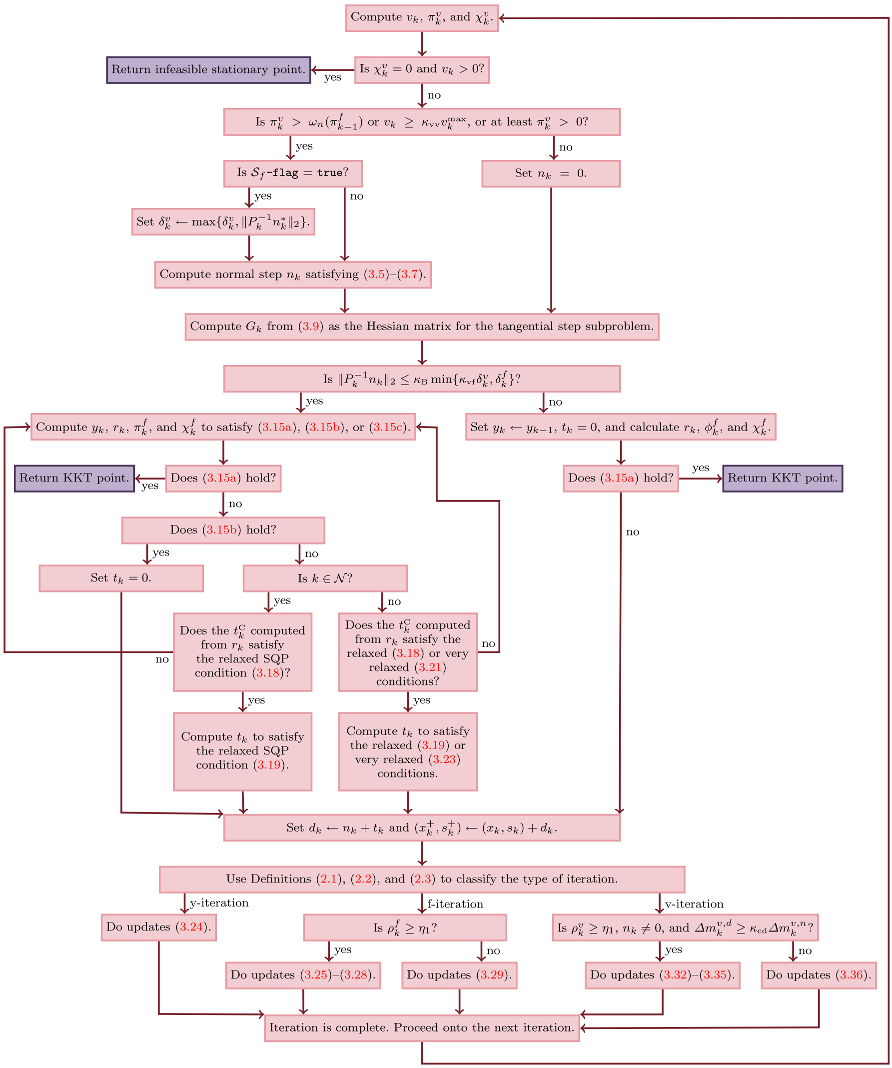
Although we have not considered them explicitly in this paper, we remark that equality constraints, call them $c_{\mathrm{E}}(x) = 0$, may easily be included in our algorithm. To do this, one may simply redefine

$$c(x, s) := \begin{pmatrix} c(x) + s \\ c_{\mathrm{E}}(x) \end{pmatrix}$$

and adjust the barrier problem (BSP), violation measure (1.3) and v-criticality measure (3.1) in obvious ways. Clearly, two-sided bounds on inequality constraints may also be incorporated in a similar fashion.

7 Appendix

The following is a flow diagram of our trust-funnel method stated as Algorithm 2.



References

1. Argáez, M., Tapia, R.: On the global convergence of a modified augmented Lagrangian linesearch interior-point Newton method for nonlinear programming. J Optim Theory Appl **114**, 1–25 (2002)
2. Byrd, R.H., Curtis, F.E., Nocedal, J.: An inexact SQP method for equality constrained optimization. SIAM J. Optim. **19**, 351–369 (2008)
3. Byrd, R.H., Gilbert, J.C., Nocedal, J.: A trust region method based on interior point techniques for nonlinear programming. Math. Program. **89**, 149–185 (2000)
4. Byrd, R.H., Hribar, M.E., Nocedal, J.: An interior point algorithm for large-scale nonlinear programming. SIAM J. Optim. **9**, 877–900 (1999)
5. Conn, A.R., Gould, N.I.M., Toint, Ph.L.: Trust-Region Methods. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA (2000)
6. Curtis, F.E., Schenk, O., Wächter, A.: An interior-point algorithm for large-scale nonlinear optimization with inexact step computations. SIAM J. Sci. Comput. **32**, 3447–3475 (2010)
7. Czyzyk, J., Fourer, R., Mehrotra, S.: Using a massively parallel processor to solve large sparse linear programs by an interior-point method. SIAM J. Sci. Comput. **19**, 553–565 (1998)
8. Fletcher, R.: Practical Methods of Optimization. Wiley-Interscience (Wiley), New York (2001)
9. Fletcher, R., Gould, N.I.M., Leyffer, S., Toint, Ph.L., Wächter, A.: Global convergence of a trust-region SQP-filter algorithm for general nonlinear programming. SIAM J. Optim. **13**, 635–659 (2002). **[(electronic) (2003)]**
10. Fletcher, R., Leyffer, S.: Nonlinear programming without a penalty function. Math. Program. **91**, 239–269 (2002)
11. Fletcher, R., Leyffer, S., Toint, Ph.L.: On the global convergence of a filter-SQP algorithm. SIAM J. Optim. **13**, 44–59 (2002)
12. Fourer, R., Mehrotra, S.: Performance of an augmented system approach for solving least-squares problems in an interior-point method for linear programming. Math. Program. **19**, 26–31 (1991)
13. Fourer, R., Mehrotra, S.: Solving symmetric indefinite systems in an interior-point method for linear programming. Math. Program. **62**, 15–39 (1993)
14. Gertz, E.M., Gill, P.E.: A primal-dual trust region algorithm for nonlinear optimization. Math. Program Ser. B **100**, 49–94 (2004)
15. Gill, P.E., Murray, W., Saunders, M.A.: SNOPT: an SQP algorithm for large-scale constrained optimization. SIAM Rev. **47**, 99–131 (2005)
16. Gill, P.E., Murray, W., Wright, M.H.: Practical Optimization. Academic Press Inc. (Harcourt Brace Jovanovich Publishers), London (1981)
17. Gondzio, J.: Interior point methods 25 years later. Eur. J. Oper. Res. **218**, 587–601 (2012)
18. Gould, N.I.M., Orban, D., Toint, Ph.L.: GALAHAD, a library of thread-safe Fortran 90 packages for large-scale nonlinear optimization. ACM Trans. Math. Softw. **29**, 353–372 (2003)
19. Gould, N.I.M., Robinson, D.P.: A second derivative SQP method: global convergence. SIAM J. Optim. **20**, 2023–2048 (2010)
20. Gould, N.I.M., Robinson, D.P.: A second derivative SQP method: local convergence and practical issues. SIAM J. Optim. **20**, 2049–2079 (2010)
21. Gould, N.I.M., Robinson, D.P.: A second derivative SQP method with a "trust-region-free" predictor step. IMA J. Numer. Anal. **32**, 580–601 (2012)
22. Gould, N.I.M., Robinson, D.P., Thorne, H.S.: On solving trust-region and other regularised subproblems in optimization. Math. Program. Comput. **2**, 21–57 (2010)
23. Gould, N.I.M., Toint, Ph.L.: Nonlinear programming without a penalty function or a filter. Math. Program. **122**, 155–196 (2010)
24. Karmarkar, N.: A new polynomial-time algorithm for linear programming. Combinatorica **4**, 373–395 (1984)
25. Karush, W.: Minima of Functions of Several Variables with Inequalities as Side Conditions. Master's thesis, Department of Mathematics, University of Chicago, Illinois, USA (1939)
26. Kuhn, H.W., Tucker, A.W.: Nonlinear programming. In: Neyman, J. (ed.) Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. University of Berkeley Press, California (1951)
27. Lalee, M., Nocedal, J., Plantenga, T.: On the implementation of an algorithm for large-scale equality constrained optimization. SIAM J. Optim. **8**, 682–706 (1998)

28. Mehrotra, S.: On the implementation of a primal-dual interior point method. SIAM J. Optim. **2**, 575–601 (1992)
29. Morales, J.L., Nocedal, J., Wu, Y.: A sequential quadratic programming algorithm with an additional equality constrained phase. IMA J. Numer. Anal. **32**, 553–579 (2012)
30. Orban, D., Gould, N.I.M., Robinson, D.P.: Trajectory-following methods for large-scale degenerate convex quadratic programming. Math. Program. Comput. **5**, 113–142 (2013)
31. Vanderbei, R.J.: LOQO: an interior point code for quadratic programming. Optim. Methods Softw. **11**, 451–484 (1999)
32. Wächter, A., Biegler, L.T.: On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. Math. Program. Ser. A **106**, 25–57 (2006)
33. Yabe, H., Yamashita, H.: Q-superlinear convergence of primal-dual interior point quasi-Newton methods for constrained optimization. J. Oper. Res. Soc. Jpn. **40**, 415–436 (1997)
34. Yamashita, H., Yabe, H.: Superlinear and quadratic convergence of some primal-dual interior point methods for constrained optimization. Math. Program. **75**, 377–397 (1996)
35. Yamashita, H., Yabe, H.: An interior point method with a primal-dual quadratic barrier penalty function for nonlinear optimization. SIAM J. Optim. **14**, 479–499 (2003)
36. Yamashita, H., Yabe, H., Tanabe, T.: A globally and superlinearly convergent primal-dual interior point trust region method for large scale constrained optimization. Math. Program. Ser. A **102**, 111–151 (2005)