

AN INEXACT PERTURBED PATH-FOLLOWING METHOD FOR LAGRANGIAN DECOMPOSITION IN LARGE-SCALE SEPARABLE CONVEX OPTIMIZATION

QUOC TRAN DINH^{*†}, ION NECOARA[‡], CARLO SAVORGNAN^{*} AND MORITZ DIEHL^{*}

Abstract. This paper studies an inexact perturbed path-following algorithm in the framework of Lagrangian dual decomposition for solving large-scale separable convex programming problems. Unlike the exact versions considered in the literature, we propose to solve the primal subproblems inexactly up to a given accuracy. This leads to an inexactness of the gradient vector and the Hessian matrix of the smoothed dual function. An inexact perturbed algorithm is then applied to minimize the smoothed dual function. The algorithm consists of two phases and both make use of the inexact derivative information of the smoothed dual problem. The convergence of the algorithm is analyzed and the worst-case complexity is estimated. As a special case, an exact path-following decomposition algorithm is obtained and its worst-case complexity is estimated. Implementation details are discussed and preliminary numerical results are reported.

Key words. Smoothing technique, self-concordant barrier, Lagrangian decomposition, inexact perturbed Newton-type method, separable convex optimization, parallel algorithm.

1. Introduction. Many optimization problems arising in networked systems, image processing, data mining, economics, distributed control and multi-stage stochastic optimization can be formulated as separable convex optimization problems, see, e.g. [5, 10, 12, 14, 20, 24, 25, 28]. For a centralized setup and problems of moderate size there exist many standard iterative algorithms to solve them such as Newton, quasi-Newton or projected gradient methods. But in many applications, we encounter separable convex programming problems which may not be easy to solve by standard optimization algorithms due to the high dimensionality, the hierarchical, multistage or dynamical structure, the existence of multiple decision-makers or the distributed locations of the data and devices. Decomposition methods can be an appropriate choice for solving these problems. Moreover, decomposition approaches also benefit if the subproblems generated from the components of the problem can be solved in a closed form or lower computational cost than the full problem.

In this paper, we are interested in the following separable convex programming problem (SCPP):

$$(1.1) \quad \begin{aligned} \max_{x \in \mathbb{R}^n} \quad & \{\phi(x) := \sum_{i=1}^M \phi_i(x_i)\} \\ \text{s.t.} \quad & \sum_{i=1}^M (A_i x_i - b_i) = 0, \\ & x_i \in X_i, \quad i = 1, \dots, M, \end{aligned}$$

where $x = (x_1^T, \dots, x_M^T)^T$ with $x_i \in \mathbb{R}^{n_i}$ is a vector of decision variables, each $\phi_i : \mathbb{R}^{n_i} \rightarrow \mathbb{R}$ is concave, X_i is a nonempty, closed convex subset in \mathbb{R}^{n_i} , $A_i \in \mathbb{R}^{m \times n_i}$, $b_i \in \mathbb{R}^m$ for all $i = 1, \dots, M$, and $n_1 + n_2 + \dots + n_M = n$. The first constraint is usually referred to as a *linear coupling constraint*.

Several methods have been proposed to solve problem (1.1) by decomposing it into smaller subproblems that can be solved separately by standard optimization

^{*}Department of Electrical Engineering (ESAT-SCD) and Optimization in Engineering Center (OPTEC), K.U. Leuven, Kasteelpark Arenberg 10, B-3001 Leuven, Belgium (`{quoc.trandinh, carlo.savorgnan, moritz.diehl}@esat.kuleuven.be`).

[†] Department of Mathematics-Mechanics-Informatics, VNU University of Science, Hanoi, Vietnam.

[‡] Automation and Systems Engineering Department, University Politehnica Bucharest, 060042 Bucharest, Romania; (`ion.necoara@acse.pub.ro`)

techniques, see e.g. [2, 4, 13, 19, 22]. One standard technique to decompose a separable programming problem is Lagrangian dual decomposition [2]. However, using such a technique generally leads to a nonsmooth optimization problem. There are several attempts to overcome this difficulty by smoothing the dual function. One can add an augmented Lagrangian term [19] or a proximal term [4] to the objective function of the primal problem. Unfortunately, the first approach breaks the separability of the original problem due to the cross terms between the components. Therefore, the second approach is more attractive to solve this type of problems.

Recently, smoothing techniques in convex optimization have attracted increasing interest and have found many applications [17]. In the framework of the Lagrangian dual decomposition, there are two relevant approaches. The first one is regularization. By adding a regularization term such as a proximal term to the objective function, the primal subproblems become strongly convex. Consequently, the dual master problem is smooth which allows one to apply smoothing optimization techniques [4, 13, 22]. The second approach is using barrier functions, this technique is suitable for problems with conic constraints [7, 9, 11, 14, 21, 27, 28]. Several methods in this direction have used a fundamental property that by using self-concordant log-barrier functions, the family of the dual functions depending on a barrier parameter is strongly self-concordant in the sense of Nesterov and Nemirovski [15]. Consequently, path-following methods can be applied to solve the dual master problem. Until now, those methods required the crucial assumption that the primal subproblems are solved exactly. In practice, solving the primal subproblems exactly to construct the dual function is only conceptual. Any numerical optimization method provides an approximate solution and, consequently, the dual function is also approximated. In this paper, we study an inexact perturbed path-following decomposition method for solving (1.1) which employs approximate gradient vectors and approximate Hessian matrices of the smoothed dual function.

Contribution. The contribution of this paper is as follows:

1. By applying a smoothing technique via self-concordant barrier functions, we prove a local and a global smooth approximation to the dual function and estimate the approximation error.
2. A new two-phase inexact perturbed path-following decomposition algorithm is proposed for solving (1.1). Both phases allow one to solve the primal subproblems approximately. The whole algorithm is highly parallelizable.
3. The convergence and the worst-case complexity of the algorithm are investigated under standard assumptions used in any interior point method.
4. As a special case, an exact path-following decomposition algorithm studied in [11, 14, 21, 28] is obtained. However, for this variant we obtain better values for the radius of the central path neighborhood compared to those from existing methods.

Let us emphasize some differences between the proposed method and existing similar methods. First, although smoothing techniques via self-concordant barriers are not new [11, 14, 21, 28], in this paper we prove a new local and global estimate for the dual function. These estimates are only based on the convexity of the objective function (not necessarily its smoothness). Since the smoothed dual function is continuously differentiable, smooth optimization techniques can be used to minimize such a function. Second, the new algorithm allows us to solve the primal subproblems inexactly, where the inexactness in the early iterations of the path-following algorithm can be high, resulting in significant time savings when the solution of the primal subprob-

lems requires a high computational cost. Note that the proposed algorithm is different from the one considered in [26] for linear programming, where the inexactness of the primal subproblems was defined in a different way. Third, by analyzing directly the convergence of the algorithm based on a recent monograph [16], the theory in this paper is self-contained. Moreover, it also allows us to optimally choose the parameters and to trade-off between the convergence rate of the master problem and the accuracy of the primal subproblems. Finally, in the exact case, the radius of the central path neighborhood is $(3 - \sqrt{5})/2 \approx 0.38197$ which is larger than $2 - \sqrt{3} \approx 0.26795$ of previous methods [11, 14, 21, 28]. Moreover, since the performance of an interior point algorithm crucially depends on the parameters of the algorithm, we analyze directly the path-following iteration to select these parameters in an appropriate way.

The rest of this paper is organized as follows. In the next section, we briefly recall the Lagrangian dual decomposition method in separable convex optimization. Section 3 is devoted to constructing smooth approximations for the dual function via self-concordant barriers and investigates the main properties of these approximations. Section 4 presents an inexact perturbed path-following decomposition algorithm and investigates the convergence and the worst-case complexity of the algorithm. Section 5 deals with an exact variant of the algorithm presented in Section 4. Section 6 discusses implementation details and Section 7 presents preliminary numerical tests. The proofs of the technical statements are given in the appendix.

Notation and Terminology. Throughout the paper, we shall consider the Euclidean space \mathbb{R}^n endowed with an inner product $x^T y$ for $x, y \in \mathbb{R}^n$ and the Euclidean norm $\|x\| = \sqrt{x^T x}$. The notation $x = (x_1, \dots, x_M)$ defines a vector in \mathbb{R}^n formed from M sub-vectors $x_i \in \mathbb{R}^{n_i}$, $i = 1, \dots, M$, where $n_1 + \dots + n_M = n$.

For a given symmetric real matrix P , the expression $P \succeq 0$ (resp. $P \succ 0$) means that P is positive semidefinite (resp. positive definite); $P \preceq Q$ means that $Q - P \succeq 0$.

For a proper, lower semi-continuous convex function f , $\text{dom}(f)$ denotes the domain of f , $\overline{\text{dom}(f)}$ is the closure of $\text{dom}(f)$ and $\partial f(x)$ denotes the subdifferential of f at x . For a concave function f we also denote by $\partial f(x)$ the ‘‘super-differential’’ of f at x , i.e. $\partial f(x) := -\partial\{-f(x)\}$. Let f be twice continuously differentiable and convex on \mathbb{R}^n . For a given vector u , the local norm of u w.r.t. f at x , where $\nabla^2 f(x) \succ 0$, is defined as $\|u\|_x := [u^T \nabla^2 f(x) u]^{1/2}$ and its dual norm is $\|u\|_x^* := \max\{u^T v \mid \|v\|_x \leq 1\} = [u^T \nabla^2 f(x)^{-1} u]^{1/2}$. Clearly, $|u^T v| \leq \|u\|_x \|v\|_x^*$. The set $\mathcal{N}_X(x) := \{w \in \mathbb{R}^n \mid w^T(x - u) \geq 0, u \in X\}$ if $x \in X$ and $\mathcal{N}_X(x) := \emptyset$, otherwise, is called the normal cone of a closed convex set X at x .

The notation \mathbb{R}_+ (resp. \mathbb{R}_{++}) defines the set of nonnegative (resp. positive) numbers. The function $\omega : \mathbb{R}_+ \rightarrow \mathbb{R}$ is defined by $\omega(t) := t - \ln(1 + t)$ and its dual $\omega_* : [0, 1] \rightarrow \mathbb{R}$ is defined by $\omega_*(t) := -t - \ln(1 - t)$. Note that both functions are convex, nonnegative and increasing. For a real number x , $\lfloor x \rfloor$ denotes the largest integer number which is less than or equal to x , and ‘‘:=’’ means ‘‘equal by definition’’.

2. Lagrangian dual decomposition in convex optimization. A classical technique to address coupling constraints in SCPP is Lagrangian dual decomposition [2]. We briefly recall such a technique in this section.

Let $A := [A_1, \dots, A_M]$ and $b := \sum_{i=1}^M b_i$. The linear coupling constraint $\sum_{i=1}^M (A_i x_i - b_i) = 0$ can be written as $Ax = b$. The Lagrange function associated with the constraint $Ax = b$ for the problem (1.1) is defined as $\mathcal{L}(x, y) := \phi(x) + y^T (Ax - b) = \sum_{i=1}^M [\phi_i(x_i) + y^T (A_i x_i - b_i)]$, where $y \in \mathbb{R}^m$ is the corresponding Lagrange multi-

plier. The dual problem of (1.1) is formulated as:

$$(2.1) \quad d_0^* := \min_{y \in \mathbb{R}^m} d_0(y),$$

where d_0 is the dual function defined by:

$$(2.2) \quad d_0(y) := \max_{x \in X} \mathcal{L}(x, y) = \max_{x \in X} \left\{ \sum_{i=1}^M [\phi_i(x_i) + y^T (A_i x_i - b_i)] \right\}.$$

We say that the problem (1.1) satisfies the Slater condition if $\text{ri}(X) \cap \{x \in \mathbb{R}^n \mid Ax = b\} \neq \emptyset$, where $\text{ri}(X)$ is the relative interior of the convex set X [3]. Let us denote by X^* and Y^* the solution sets of (1.1) and (2.1), respectively. Throughout the paper, we assume that the following fundamental assumptions hold.

ASSUMPTION A.1.

- (a) *The solution set X^* of (1.1) is nonempty and either the Slater condition for (1.1) is satisfied or X is polyhedral.*
- (b) *For $i = 1, \dots, M$, the function ϕ_i is proper, upper semicontinuous and concave on X_i .*
- (c) *The matrix A is full-row rank.*

Note that Assumptions A.1.a) and A.1.b) are standard in convex optimization, which guarantee the solvability of the primal-dual problems and strong duality. Assumption A.1.c) is not restrictive since it can be guaranteed by applying standard linear algebra techniques to eliminate redundant constraints.

Under Assumption A.1, the solution set Y^* of the dual problem (2.1) is nonempty, convex and bounded. Moreover, *strong duality* holds, i.e.:

$$d_0^* = d_0(y_0^*) = \min_{y \in \mathbb{R}^m} d_0(y) = \max_{x \in X} \{\phi(x) \mid Ax = b\} = \phi(x_0^*) = \phi^*, \quad \forall (x_0^*, y_0^*) \in X^* \times Y^*.$$

Finally, the dual function $d_0(\cdot)$ can be computed separately by:

$$(2.3) \quad d_0(y) = \sum_{i=1}^M d_{0,i}(y), \quad \text{where } d_{0,i}(y) := \max_{x_i \in X_i} \{\phi_i(x_i) + y^T (A_i x_i - b_i)\}, \quad i = 1, \dots, M.$$

We denote by $x_{0,i}^*(y)$ a solution of the maximization problem in (2.3) for $i = 1, \dots, M$ and $x_0^*(y) := (x_{0,1}^*(y), \dots, x_{0,M}^*(y))$.

3. Smoothing via self-concordant barriers. Let us assume that the feasible set X_i possesses a ν_i -self-concordant barrier F_i for $i = 1, \dots, M$, see [15, 16]. In other words, we assume the following.

ASSUMPTION A.2. *For each $i \in \{1, \dots, M\}$, the feasible set X_i is bounded in \mathbb{R}^{n_i} with $\text{int}(X_i) \neq \emptyset$ and possesses a self-concordant barrier F_i with a parameter $\nu_i > 0$.*

The assumption on the boundedness of X_i is not restrictive. In principle, we can bound the set of desired solutions by a sufficiently large compact set such that all the sample points generated by an optimization algorithm belong to this set. A more general case, $\text{ri}(X_i) \neq \emptyset$, will be discussed in Subsection 6.1.

Let us denote by x_i^c the *analytic center* of X_i , which is defined as:

$$x_i^c := \operatorname{argmin} \{F_i(x_i) \mid x_i \in \text{int}(X_i)\}, \quad i = 1, \dots, M.$$

Under Assumption A.2, $x^c := (x_1^c, \dots, x_M^c)$ is well-defined due to [18, Corollary 2.3.6]. To compute x^c , one can apply the algorithms proposed in [16, pp. 204–205]. Moreover,

the following estimates hold:

$$(3.1) \quad F_i(x_i) - F_i(x_i^c) \geq \omega(\|x_i - x_i^c\|_{x_i^c}) \text{ and } \|x_i - x_i^c\|_{x_i^c} \leq \nu_i + 2\sqrt{\nu_i},$$

for all $x_i \in \overline{\text{dom}}(F_i)$ and $i = 1, \dots, M$ [16, Theorems 4.1.13 and 4.2.6].

3.1. A smooth approximation of the dual function. Let us define the following function:

$$(3.2) \quad d(y, t) := \sum_{i=1}^M d_i(y, t), \quad d_i(y, t) := \max_{x_i \in \text{int}(X_i)} \{ \phi_i(x_i) + y^T(A_i x_i - b_i) - t[F_i(x_i) - F_i(x_i^c)] \},$$

where $t > 0$ is referred to as a smoothness or barrier parameter. Similar to [9, 14, 21, 28], we can show that $d(\cdot, t)$ is well-defined and smooth due to strict convexity of F_i . We denote by $x_i(y, t)$ the unique solution of the maximization problem in (3.2) for $i = 1, \dots, M$ and $x^*(y, t) = (x_1^*(y, t), \dots, x_M^*(y, t))$. We refer to d as a *smoothed dual function* of d_0 and to the maximization problem in (3.2) as *primal subproblem*. The optimality condition for the primal subproblem (3.2) is:

$$(3.3) \quad 0 \in \partial \phi_i(x_i^*(y, t)) + A_i^T y - t \nabla F_i(x_i^*(y, t)), \quad i = 1, \dots, M$$

where $\partial \phi_i(x_i^*(y, t))$ is the super-differential of ϕ_i at $x_i^*(y, t)$. Since the problem (3.2) is unconstrained and convex, this condition is necessary and sufficient.

Associated with $d(\cdot, t)$, we consider the following smoothed dual *master problem*:

$$(3.4) \quad d^*(t) := \min_{y \in Y} d(y, t).$$

We denote by $y^*(t)$ a solution of (3.4) if it exists and by $x^*(t) := x^*(y^*(t), t)$.

Let $F(x) := \sum_{i=1}^M F_i(x_i)$. Then the function F is also a self-concordant barrier of X with a parameter $\nu := \sum_{i=1}^M \nu_i$. For a given $\beta \in (0, 1)$, we define a neighbourhood in \mathbb{R}^m w.r.t. F and $t > 0$ as:

$$\mathcal{N}_t^F(\beta) := \left\{ y \in \mathbb{R}^m \mid \lambda_{F_i}(x_i^*(y, t)) := \|\nabla F_i(x_i^*(y, t))\|_{x_i^*(y, t)}^* \leq \beta, \quad i = 1, \dots, M \right\}.$$

Since $x^c \in \mathcal{N}_t^F(\beta)$, this set is nonempty. Let $\underline{\omega}(x^*(y, t)) := \sum_{i=1}^M \omega(\|x_i^*(y, t) - x_i^c\|_{x_i^c})$ and $\bar{\omega}(x^*(y, t)) := \sum_{i=1}^M \nu_i \omega^{-1}(\nu_i^{-1} \omega_*(\lambda_{F_i}(x_i^*(y, t))))$. The following lemma provides a local estimate for $d_0(\cdot)$, whose proof can be found in the appendix.

LEMMA 3.1. *Suppose that Assumptions A.1 and A.2 are satisfied and $\beta \in (0, 1)$. Then the function $d(\cdot, t)$ defined by (3.2) satisfies:*

$$(3.5) \quad 0 \leq t \underline{\omega}(x^*(y, t)) \leq d_0(y) - d(y, t) \leq t [\bar{\omega}(x^*(y, t)) + \nu],$$

for all $y \in \mathcal{N}_t^F(\beta)$. Consequently, one has:

$$0 \leq d_0(y) - d(y, t) \leq t [\bar{\omega}_\beta + \nu], \quad \forall y \in \mathcal{N}_t^F(\beta),$$

where $\bar{\omega}_\beta := \sum_{i=1}^M \nu_i \omega^{-1}(\nu_i^{-1} \omega_*(\beta))$ and ω^{-1} is the inverse function of ω .

Lemma 3.1 implies that, for a given $\varepsilon_d > 0$, if we choose $t_f := (\bar{\omega}_\beta + \nu)^{-1} \varepsilon_d$, then $d(y, t_f) \leq d_0(y) \leq d(y, t_f) + \varepsilon_d$ for all $y \in \mathcal{N}_{t_f}^F(\beta)$.

Under Assumption A.1, the solution set Y^* of the dual problem (2.1) is bounded. Let Y be a compact set in \mathbb{R}^m such that $Y^* \subseteq Y$. We define:

$$(3.6) \quad K_i := \max_{y \in Y} \max_{\xi_i \in \partial \phi_i(x_i^c)} \{ \|\xi_i + A_i^T y\|_{x_i^c}^* \} \in [0, +\infty), \quad i = 1, \dots, M.$$

The following lemma provides a global estimate of the dual function $d_0(\cdot)$. The proof of this lemma can be found in the appendix.

LEMMA 3.2. *Suppose that Assumptions A.1 and A.2 are satisfied and the constants K_i , $i = 1, \dots, M$, are defined by (3.6). Then, for any $t > 0$, we have:*

$$(3.7) \quad t\omega(x^*(y, t)) \leq d_0(y) - d(y, t) \leq tD_X(t), \quad \forall y \in Y,$$

where $D_X(t) := \sum_{i=1}^M \bar{\zeta}(K_i; \nu_i, t)$ and $\bar{\zeta}(\tau; a, b) := a(1 + \max\{0, \ln(\frac{\tau}{ab})\})$.

Consequently, for a given tolerance $\varepsilon_d > 0$ and a constant $\kappa \in (0, 1)$, if:

$$(3.8) \quad 0 < t \leq \bar{t} := \min_{1 \leq i \leq M} \left\{ \frac{K_i}{\nu_i} \kappa^{1/\kappa}, \left(\frac{\varepsilon}{\sum_{i=1}^M [\nu_i + \nu_i^{1-\kappa} K_i^\kappa]} \right)^{1/(1-\kappa)} \right\},$$

then $d(y, t) \leq d_0(y) \leq d(y, t) + \varepsilon_d$ for all $y \in Y$.

Lemma 3.2 shows that if we fix $t_f \in (0, \bar{t}]$ and minimize $d(\cdot, t_f)$ over Y , then the obtained solution $y^*(t_f)$ is an ε_d -solution of (2.1). Since $d(\cdot, t_f)$ is continuously differentiable, smooth optimization techniques such as gradient-based methods can be applied to minimize $d(\cdot, t_f)$ over Y .

3.2. The self-concordance of the smoothed dual function. If the function $-\phi_i$ is self-concordant on $\text{dom}(-\phi_i)$ with a parameter M_{ϕ_i} , then the family of the functions $\phi_i(\cdot, t) := tF(\cdot) - \phi_i(\cdot)$ is also self-concordant on $\text{dom}(-\phi_i) \cap \text{dom}(F_i)$. Consequently, the smooth dual function $d(\cdot, t)$ is self-concordant due to Legendre transformation as stated in the following lemma, see, e.g. [11, 14, 21, 28].

LEMMA 3.3. *Suppose that Assumptions A.1 and A.2 are satisfied. Suppose further that $-\phi_i$ is M_{ϕ_i} -self-concordant. Then, for $t > 0$, the function $d_i(\cdot, t)$ defined by (3.2) is self-concordant with the parameter $M_{d_i} := \max\{M_{\phi_i}, 2/\sqrt{t}\}$, $i = 1, \dots, M$. Consequently, $d(\cdot, t)$ is self-concordant with the parameter $M_d := \max_{1 \leq i \leq M} M_{d_i}$.*

Similar to standard path-following methods [15, 16], in the following discussion, we assume that ϕ_i is linear as stated in Assumption A.3 below.

ASSUMPTION A.3. *The function ϕ_i is linear, i.e. $\phi_i(x_i) := c_i^T x_i$ for $i = 1, \dots, M$. Let $c := (c_1, \dots, c_M)$ be a column vector formed from c_i ($i = 1, \dots, M$). Assumption A.3 and Lemma 3.3 imply that $d(\cdot, t)$ is $\frac{2}{\sqrt{t}}$ -self-concordant. Since ϕ_i is linear, the optimality condition (3.3) is rewritten as:*

$$(3.9) \quad c + A^T y - t\nabla F(x^*(y, t)) = 0.$$

The following lemma provides an explicit formula to compute the derivatives of $d(\cdot, t)$. The proof can be found in [14, 28].

LEMMA 3.4. *Suppose that Assumptions A.1-A.3 are satisfied. Then the gradient vector and the Hessian matrix of $d(\cdot, t)$ on Y are respectively given as:*

$$(3.10) \quad \nabla d(y, t) = Ax^*(y, t) - b \quad \text{and} \quad \nabla^2 d(y, t) = t^{-1} A \nabla^2 F(x^*(y, t))^{-1} A^T,$$

where $x^*(y, t)$ is the solution of the primal subproblem (3.2).

Note that since A is full-row rank and $\nabla^2 F(x^*(y, t)) \succ 0$, we can see that $\nabla^2 d(y, t) \succ 0$ for any $y \in Y$. Now, since $d(\cdot, t)$ is $\frac{2}{\sqrt{t}}$ self-concordant, if we define:

$$(3.11) \quad \tilde{d}(y, t) := t^{-1} d(y, t),$$

then $\tilde{d}(\cdot, t)$ is standard self-concordant, i.e. $M_{\tilde{d}} = 2$, due to [16, Corollary 4.1.2]. For a given vector $v \in \mathbb{R}^m$, we define the local norm $\|v\|_y$ of v w.r.t. $\tilde{d}(\cdot, t)$ as $\|v\|_y := [v^T \nabla^2 \tilde{d}(y, t) v]^{1/2}$.

3.3. Optimality and feasibility recovery. It remains to show the relations between the master problem (3.4), the dual problem (2.1) and the original primal problem (1.1). We first prove the following lemma.

LEMMA 3.5. *Let Assumptions A.1-A.3 be satisfied. Then:*

- a) *For a given $y \in Y$, $d(y, \cdot)$ is nonincreasing in \mathbb{R}_{++} .*
- b) *The function $d^*(\cdot)$ defined by (3.4) is nonincreasing and differentiable in \mathbb{R}_{++} . Moreover, $d^*(t) \leq d_0^* = \phi^*$ and $\lim_{t \downarrow 0^+} d^*(t) = \phi^*$.*
- c) *The point $x^*(t) := x^*(y^*(t), t)$ is feasible to (1.1) and $\lim_{t \downarrow 0^+} x^*(t) = x_0^*$.*

Proof. Since the function $\xi(x, y, t) := \phi(x) + y^T(Ax - b) - t[F(x) - F(x^c)]$ is strictly concave in x and linear in t , it is well-known that $d(y, t) = \max\{\xi(x, y, t) \mid x \in \text{int}(X)\}$ is differentiable w.r.t. t and its derivative is given by $\frac{\partial d(y, t)}{\partial t} = -[F(x^*(y, t)) - F(x^c)] \leq -\omega(\|x^*(y, t) - x^c\|_{x^c}) \leq 0$ due to (3.1). Thus $d(y, \cdot)$ is nonincreasing in t as stated in a). From the definitions of $d^*(\cdot)$, $d(y, \cdot)$ and $y^*(\cdot)$ in (3.4), and strong duality, we have:

$$\begin{aligned}
 d^*(t) &= \min_{y \in Y} d(y, t) \stackrel{\text{strong duality}}{=} \max_{x \in \text{int}(X)} \min_{y \in Y} \{\phi(x) + y^T(Ax - b) - t[F(x) - F(x^c)]\} \\
 (3.12) \quad &= \max_{x \in \text{int}(X)} \{\phi(x) - t[F(x) - F(x^c)] \mid Ax = b\} \\
 &= \phi(x^*(t)) - t[F(x^*(t)) - F(x^c)].
 \end{aligned}$$

It follows from the second line of (3.12) that $d^*(\cdot)$ is differentiable and nonincreasing in \mathbb{R}_{++} . From the second line of (3.12), we also deduce that $x^*(t)$ is feasible to (1.1). The limit in c) was proved in [28, Proposition 2]. Since $x^*(t)$ is feasible to (1.1) and $F(x^*(t)) - F(x^c) \geq 0$, the last line of (3.12) implies that $d^* \leq d_0^*$. We also obtain the limit $\lim_{t \downarrow 0^+} d^*(t) = d_0^* = \phi^*$. \square

Let us define the Newton decrement of $\tilde{d}(\cdot, t)$ as follows:

$$(3.13) \quad \lambda = \lambda_{\tilde{d}(\cdot, t)}(y) := \|\nabla \tilde{d}(y, t)\|_y^* = \left[\nabla \tilde{d}(y, t) \nabla^2 \tilde{d}(y, t)^{-1} \nabla \tilde{d}(y, t) \right]^{1/2}.$$

The following lemma shows the gap between $d(y, t)$ and $d^*(t)$.

LEMMA 3.6. *Suppose that Assumptions A.1-A.3 are satisfied. Then, for any $y \in Y$ and $t > 0$ such that $\lambda_{\tilde{d}(\cdot, t)}(y) \leq \beta < 1$, we have:*

$$(3.14) \quad 0 \leq t\omega(\lambda_{\tilde{d}(\cdot, t)}(y)) \leq d(y, t) - d^*(t) \leq t\omega_*(\lambda_{\tilde{d}(\cdot, t)}(y)).$$

Moreover, it holds that:

$$(3.15) \quad (c + A^T y)^T(u - x^*(y, t)) \leq t\nu \text{ and } \|Ax^*(y, t) - b\|_y^* \leq t\beta,$$

for all $u \in X$.

Proof. Since $\tilde{d}(\cdot, t)$ is standard self-concordant and $y^*(t) = \text{argmin}\{\tilde{d}(y, t) \mid y \in Y\}$, for any $y \in Y$ such that $\lambda \leq \beta < 1$, by applying [16, Theorem 4.1.13, inequality 4.1.17], we have $0 \leq \omega(\lambda) \leq \tilde{d}(y, t) - \tilde{d}(y^*(t), t) \leq \omega_*(\lambda)$. By (3.11), these inequalities are equivalent to (3.14). It follows from the optimality condition (3.9) that $c + A^T y = t\nabla F(x^*(y, t))$. Hence, by [16, Theorem 4.2.4], we have $(c + A^T y)^T(u - x^*(y, t)) = t\nabla F(x^*(y, t))^T(u - x^*(y, t)) \leq t\nu$ for any $u \in \text{dom}F$. Since $X \subseteq \text{dom}F$, the last inequality implies the first condition in (3.15). Furthermore, from (3.10) we have $\nabla d(y, t) = Ax^*(y, t) - b$. Therefore, $\|Ax^*(y, t) - b\|_y^* = t\|\nabla \tilde{d}(y^*(t), t)\|_{y^*}^* = t\lambda_{\tilde{d}(\cdot, t)}(y) \leq t\beta$. \square

Let us recall the optimality condition for the primal-dual problems (1.1)-(2.1) as:

$$(3.16) \quad 0 \in c + A^T y_0^* - \mathcal{N}_X(x_0^*) \text{ and } Ax_0^* - b = 0, \quad \forall (x_0^*, y_0^*) \in \mathbb{R}^n \times \mathbb{R}^m,$$

where $\mathcal{N}_X(x)$ is the normal cone of X at x . Here, since X^* is nonempty, the first inclusion also indicates implicitly that $x_0^* \in X$. Moreover, if $x_0^* \in X$ then (3.16) can be expressed equivalently as $(c + A^T y_0^*)^T (u - x_0^*) \leq 0$ for all $u \in X$. Now, we define an approximate solution of (1.1)-(2.1) as follows.

DEFINITION 3.7. *For a given tolerance $\varepsilon_p \geq 0$, a point $(\hat{x}, \hat{y}) \in X \times \mathbb{R}^m$ is said to be an ε_p -solution of (1.1)-(2.1) if $(c + A^T \hat{y})^T (u - \hat{x}) \leq \varepsilon_p$ for all $u \in X$ and $\|A\hat{x} - b\|_{\hat{y}}^* \leq \varepsilon_p$.*

It is clear that for any point $x \in \text{int}(X)$, $\mathcal{N}_X(x) = \{0\}$. Furthermore, according to (3.16), the conditions in Definition (3.7) are well-defined.

Finally, we note that $\nu \geq 1$, $\beta < 1$ and $x^*(y, t) \in \text{int}(X)$. By (3.15), if we choose the tolerance $\varepsilon_p := \nu t$ then $(x^*(y, t), y)$ is an ε_p -solution of (1.1)-(2.1) in the sense of Definition 3.7. We denote the feasibility gap by $\mathcal{F}(y, t) := \|Ax^*(y, t) - b\|_y^*$ for further references.

4. Inexact perturbed path-following method. This section presents an inexact perturbed path-following decomposition algorithm for solving (2.1).

4.1. Inexact solution of the primal subproblems. Firstly, we define an inexact solution of (3.2) by using local norms. For a given $y \in Y$ and $t > 0$, suppose that we solve approximately (3.2) up to a given accuracy $\bar{\delta} \geq 0$. More precisely, we define this approximation as follows.

DEFINITION 4.1. *For given $\bar{\delta} \geq 0$, a vector $\bar{x}_{\bar{\delta}}(y, t)$ is said to be a $\bar{\delta}$ -approximate solution of $x^*(y, t)$ if:*

$$(4.1) \quad \|\bar{x}_{\bar{\delta}}(y, t) - x^*(y, t)\|_{x^*(y, t)} \leq \bar{\delta}.$$

Associated with $\bar{x}_{\bar{\delta}}(\cdot)$, we define the following function:

$$(4.2) \quad d_{\bar{\delta}}(y, t) := c^T \bar{x}_{\bar{\delta}}(y, t) + y^T (A\bar{x}_{\bar{\delta}}(y, t) - b) - t[F(\bar{x}_{\bar{\delta}}(y, t)) - F(x^c)].$$

This function can be considered as an *inexact version* of d . Next, we introduce two quantities:

$$(4.3) \quad \nabla d_{\bar{\delta}}(y, t) := A\bar{x}_{\bar{\delta}}(y, t) - b, \quad \text{and} \quad \nabla^2 d_{\bar{\delta}}(y, t) := t^{-1} A \nabla^2 F(\bar{x}_{\bar{\delta}}(y, t))^{-1} A^T.$$

Since $x^*(y, t) \in \text{dom}(F)$, we can choose an appropriate $\bar{\delta} \geq 0$ such that $\bar{x}_{\bar{\delta}}(y, t) \in \text{dom}(F)$. Hence, $\nabla^2 F(\bar{x}_{\bar{\delta}}(y, t))$ is positive definite which means that $\nabla^2 d_{\bar{\delta}}$ is well-defined. Note that $\nabla d_{\bar{\delta}}$ and $\nabla^2 d_{\bar{\delta}}$ are not the gradient vector and Hessian matrix of $d_{\bar{\delta}}(\cdot, t)$. However, due to Lemma 3.4 and (4.1), we can consider these quantities as an approximate gradient vector and Hessian matrix of $d(\cdot, t)$, respectively.

Let:

$$(4.4) \quad \tilde{d}_{\bar{\delta}}(y, t) := t^{-1} d_{\bar{\delta}}(y, t),$$

and $\bar{\lambda}$ be the inexact Newton decrement of $\tilde{d}_{\bar{\delta}}$ which is defined by:

$$(4.5) \quad \bar{\lambda} = \bar{\lambda}_{\tilde{d}_{\bar{\delta}}(\cdot, t)}(y) := \|\nabla \tilde{d}_{\bar{\delta}}(y, t)\|_y^* = \left[\nabla \tilde{d}_{\bar{\delta}}(y, t) \nabla^2 \tilde{d}_{\bar{\delta}}(y, t)^{-1} \nabla \tilde{d}_{\bar{\delta}}(y, t) \right]^{1/2}.$$

Here, we use the norm $\|\cdot\|_y$ to distinguish it from $\|\cdot\|_y$.

4.2. The algorithmic framework. From Lemma 3.6 we see that if we can generate a sequence $\{(y^k, t_k)\}_{k \geq 0}$ such that $\lambda_k := \lambda_{\tilde{d}(\cdot, t_k)}(y^k) \leq \beta < 1$, then:

$$d(y^k, t_k) \uparrow d_0^* = \phi^* \text{ and } \mathcal{F}(y^k, t_k) \rightarrow 0, \text{ as } t_k \downarrow 0^+.$$

Therefore, the aim of the algorithm is to generate $\{(y^k, t_k)\}_{k \geq 0}$ such that $\lambda_k \leq \beta < 1$ and $t_k \downarrow 0^+$. First, we fix $t = t_0 > 0$ and find a point $y^0 \in Y$ such that $\lambda_{\tilde{d}(\cdot, t_0)}(y^0) \leq \beta$. Then we simultaneously update y^k and t_k such that $t_k \downarrow 0^+$. The algorithmic framework is presented as follows.

INEXACT-PERTURBED PATH-FOLLOWING ALGORITHMIC FRAMEWORK.

Initialization. Choose an appropriate $\beta \in (0, 1)$ and a tolerance $\varepsilon_d > 0$. Fix $t = t_0 > 0$ a priori.

Phase 1. (Determine a starting point $y^0 \in Y$ such that $\lambda_{\tilde{d}(\cdot, t_0)}(y^0) \leq \beta$).

Choose an initial vector $y^{0,0} \in Y$.

For $j = 0, 1, \dots, j_{\max}$ **perform**

1. If $\lambda_j := \lambda_{\tilde{d}(\cdot, t_0)}(y^{0,j}) \leq \beta$ then set $y^0 := y^{0,j}$ and terminate.
2. Solve (3.2) *in parallel* to obtain an approximation of $x^*(y^{0,j}, t_0)$.
3. Evaluate $\nabla d_{\bar{\delta}}(y^{0,j}, t_0)$ and $\nabla^2 d_{\bar{\delta}}(y^{0,j}, t_0)$ by (4.3).
4. Perform the inexact-perturbed damped Newton step: $y^{0,j+1} := y^{0,j} - \alpha_j \nabla^2 d_{\bar{\delta}}(y^{0,j}, t_0)^{-1} \nabla d_{\bar{\delta}}(y^{0,j}, t_0)$, where $\alpha_j \in (0, 1]$ is a given step size.

End For

Phase 2. (Path-following iterations).

Compute an appropriate value $\sigma \in (0, 1)$.

For $k = 0, 1, \dots, k_{\max}$ **perform**:

1. If $t_k \leq \varepsilon_d / \omega_*(\beta)$ then terminate.
2. Update $t_{k+1} := (1 - \sigma)t_k$.
3. Solve (3.2) *in parallel* to obtain an approximation of $x^*(y^k, t_{k+1})$.
4. Evaluate the quantities $\nabla d_{\bar{\delta}}(y^k, t_{k+1})$ and $\nabla^2 d_{\bar{\delta}}(y^k, t_{k+1})$ as in (3.10).
5. Perform the inexact-perturbed full-step Newton step as $y^{k+1} := y^k - \nabla^2 d_{\bar{\delta}}(y^k, t_{k+1})^{-1} \nabla d_{\bar{\delta}}(y^k, t_{k+1})$.

End For

Output. An ε_d -approximate solution y^k of (3.4), i.e. $0 \leq d(y^k, t_k) - d^*(t_k) \leq \varepsilon_d$.

END.

This algorithm is still conceptual. In the following subsections, we shall discuss each step of this algorithmic framework in detail. We notice that the proposed algorithm provides an ε_d -approximate solution y^k such that $t_k \leq \varepsilon_t := \omega_*(\beta)^{-1} \varepsilon_d$. Now, by solving the primal subproblem (3.2), we obtain $x^*(y^k, t_k)$ as an ε_p -solution of (1.1) in the sense of Definition 3.7, where $\varepsilon_p := \nu \varepsilon_t$.

4.3. Computing inexact solutions. The condition (4.1) can not be used in practice to compute $\bar{x}_{\bar{\delta}}$ since $x^*(y, t)$ is unknown. We need to show how to compute $\bar{x}_{\bar{\delta}}$ practically such that (4.1) holds.

For the sake of notational simplicity, we abbreviate by $\bar{x}_{\bar{\delta}} := \bar{x}_{\bar{\delta}}(y, t)$ and $x^* := x^*(y, t)$. The error of the approximate solution $\bar{x}_{\bar{\delta}}$ to x^* is defined as:

$$(4.6) \quad \delta(\bar{x}_{\bar{\delta}}, x^*) := \|\bar{x}_{\bar{\delta}}(y, t) - x^*(y, t)\|_{x^*(y, t)}.$$

The following lemma gives a criterion to ensure that the condition (4.1) holds.

LEMMA 4.2. *Let $\delta(\bar{x}_{\bar{\delta}}, x^*)$ be defined by (4.6) such that $\delta(\bar{x}_{\bar{\delta}}, x^*) < 1$. Then:*

$$(4.7) \quad 0 \leq t\omega(\delta(\bar{x}_{\bar{\delta}}, x^*)) \leq d(y, t) - d_{\bar{\delta}}(y, t) \leq t\omega_*(\delta(\bar{x}_{\bar{\delta}}, x^*)).$$

Moreover, if:

$$(4.8) \quad E_{\bar{\delta}}^c := \|c + A^T y - t \nabla F(\bar{x}_{\bar{\delta}})\|_{x^c}^* \leq \varepsilon_d := [(\nu + 2\sqrt{\nu})(1 + \bar{\delta})]^{-1} \bar{\delta} t$$

then $\bar{x}_{\bar{\delta}}(y, t)$ satisfies (4.1). Consequently, if $t \leq \omega_*(\beta)^{-1} \varepsilon_d$ and $\bar{\delta} < 1$ then:

$$(4.9) \quad |d_{\bar{\delta}}(y, t) - d^*(t)| \leq [1 + \omega_*(\beta)^{-1} \omega_*(\bar{\delta})] \varepsilon_d.$$

Proof. It follows from the definitions of $d(\cdot, t)$ and $d_{\bar{\delta}}(\cdot, t)$, and (3.9) that:

$$\begin{aligned} d(y, t) - d_{\bar{\delta}}(y, t) &= [c + A^T y](x^* - \bar{x}_{\bar{\delta}}) - t[F(x^*) - F(\bar{x}_{\bar{\delta}})] \\ &= -t[F(x^*) + \nabla F(x^*)^T(\bar{x}_{\bar{\delta}} - x^*) - F(\bar{x}_{\bar{\delta}})]. \end{aligned}$$

Since F is self-concordant, by applying [16, Theorems 4.1.7 and 4.1.8], and the definition of $\delta(\bar{x}_{\bar{\delta}}, x^*)$, the above equality implies that:

$$0 \leq t\omega(\delta(\bar{x}_{\bar{\delta}}, x^*)) \leq d(y, t) - d_{\bar{\delta}}(y, t) \leq t\omega_*(\delta(\bar{x}_{\bar{\delta}}, x^*)),$$

which is indeed (4.7).

Next, by using again (3.9) and the definition of $E_{\bar{\delta}}^c$ we have:

$$E_{\bar{\delta}}^c \stackrel{(3.9)}{=} t \|\nabla F(\bar{x}_{\bar{\delta}}) - \nabla F(x^*)\|_{x^c}^* \geq (\nu + 2\sqrt{\nu})^{-1} t \|\nabla F(\bar{x}_{\bar{\delta}}) - \nabla F(x^*)\|_{x^*}^*,$$

where the last inequality follows from [16, Corollary 4.2.1]. Combining this inequality and [16, Theorem 4.1.7], we obtain:

$$\begin{aligned} \frac{\delta(\bar{x}_{\bar{\delta}}, x^*)^2}{1 + \delta(\bar{x}_{\bar{\delta}}, x^*)} &\leq [\nabla F(\bar{x}_{\bar{\delta}}) - \nabla F(x^*)]^T (\bar{x}_{\bar{\delta}} - x^*) \leq \|\nabla F(\bar{x}_{\bar{\delta}}) - \nabla F(x^*)\|_{x^*}^* \|\bar{x}_{\bar{\delta}} - x^*\|_{x^*} \\ &\leq t^{-1} (\nu + 2\sqrt{\nu}) E_{\bar{\delta}}^c \delta(\bar{x}_{\bar{\delta}}, x^*). \end{aligned}$$

Hence, we get:

$$(4.10) \quad \delta(\bar{x}_{\bar{\delta}}, x^*) \leq [t - (\nu + 2\sqrt{\nu}) E_{\bar{\delta}}^c]^{-1} (\nu + 2\sqrt{\nu}) E_{\bar{\delta}}^c,$$

provided that $t > (\nu + 2\sqrt{\nu}) E_{\bar{\delta}}^c$. Let us define an accuracy ε_p for the primal subproblem (3.2) as $\varepsilon_p := [(\nu + 2\sqrt{\nu})(1 + \bar{\delta})]^{-1} \bar{\delta} t \geq 0$. Then it follows from (4.10) that if $E_{\bar{\delta}}^c \leq [(\nu + 2\sqrt{\nu})(1 + \bar{\delta})]^{-1} \bar{\delta} t$ then $\bar{x}_{\bar{\delta}}(y, t)$ satisfies (4.1). It remains to consider the distance from $d_{\bar{\delta}}$ to $d^*(t)$ when t is sufficiently small. Suppose that $t \leq \omega_*(\beta)^{-1} \varepsilon_d$. Then, by combining (3.14) and (4.7) we obtain (4.9). \square

REMARK 1. Since $E_{\bar{\delta}} := \|c + A^T y - t \nabla F(\bar{x}_{\bar{\delta}})\|_{\bar{x}_{\bar{\delta}}}^* \geq (1 - \bar{\delta}) \|c + A^T y - t \nabla F(\bar{x}_{\bar{\delta}})\|_{x^*}^*$, by the same argument as in the proof of Lemma 4.2, we can show that if $E_{\bar{\delta}} \leq \hat{\varepsilon}_p$, where $\hat{\varepsilon}_p := \frac{\bar{\delta}(1 - \bar{\delta})t}{1 + \bar{\delta}}$ then (4.1) holds. This condition can be used to terminate the algorithm presented in the next section.

4.4. Phase 2 - The path-following scheme with inexact-perturbed full-step Newton iterations. Now, we analyze Steps 2-5 in Phase 2 of the algorithmic framework. In the path-following fashion, we only perform one inexact-perturbed full-step Newton (IPFNT) iteration for each value of the parameter t . Thus one iteration of this scheme is specified as follows:

$$(4.11) \quad \begin{cases} t_+ := t - \Delta t, \\ y_+ := y - \nabla^2 d_{\bar{\delta}}(y, t_+)^{-1} \nabla d_{\bar{\delta}}(y, t_+). \end{cases}$$

Since the Newton method is invariant under linear transformations, by (4.2), the second line of (4.11) is equivalent to:

$$(4.12) \quad y_+ := y - \nabla^2 \tilde{d}_{\bar{\delta}}(y, t_+)^{-1} \nabla \tilde{d}_{\bar{\delta}}(y, t_+).$$

For the sake of notational simplicity, we denote all the functions at (y_+, t_+) and (y, t_+) by the sub-index “+” and “1”, respectively, and at (y, t) without index in the following analysis. More precisely, we denote by:

$$(4.13) \quad \begin{aligned} \bar{\lambda}_+ &:= \bar{\lambda}_{\bar{d}_{\bar{\delta}}(\cdot, t_+)}(y_+), & \delta_+ &:= \|\bar{x}_{\bar{\delta}}(y_+, t_+) - x^*(y_+, t_+)\|_{x^*(y_+, t_+)}, \\ \bar{\lambda}_1 &:= \bar{\lambda}_{\bar{d}_{\bar{\delta}}(\cdot, t_+)}(y), & \delta_1 &:= \|\bar{x}_{\bar{\delta}}(y, t_+) - x^*(y, t_+)\|_{x^*(y, t_+)}, \\ \bar{\lambda} &:= \bar{\lambda}_{\bar{d}_{\bar{\delta}}(\cdot, t)}(y), & \delta &:= \|\bar{x}_{\bar{\delta}}(y, t) - x^*(y, t)\|_{x^*(y, t)}, \end{aligned}$$

and by

$$(4.14) \quad \Delta := \|\bar{x}_{\bar{\delta}}(y, t_+) - \bar{x}_{\bar{\delta}}(y, t)\|_{\bar{x}_{\bar{\delta}}(y, t)} \text{ and } \Delta^* := \|x^*(y, t_+) - x^*(y, t)\|_{x^*(y, t)}.$$

Note that the above notation does not cause any confusion since it can be recognized from the context.

4.4.1. The main estimate. Now, by using the notation in (4.13) and (4.14), we provide a main estimate which will be used to analyze the convergence of the algorithm presented in Subsection 4.4.4 whose proof can be found in the appendix.

LEMMA 4.3. *Let $y \in Y$ be given and $t > 0$. Let (y_+, t_+) be a pair generated by (4.11). Suppose that $\delta_1 + 2\Delta + \bar{\lambda} < 1$, $\delta_+ < 1$ and $\xi := \frac{\Delta + \bar{\lambda}}{1 - \delta_1 - 2\Delta - \bar{\lambda}}$. Then:*

$$(4.15) \quad \bar{\lambda}_+ \leq (1 - \delta_+)^{-1} \left\{ \delta_+ + \delta_1 + \xi^2 + \delta_1 [(1 - \delta_1)^{-2} + 2(1 - \delta_1)^{-1}] \xi \right\}.$$

Moreover, the right-hand side of (4.15) is nondecreasing w.r.t. all variables δ_+ , δ_1 , Δ and $\bar{\lambda}$.

In particular, if we set $\delta_+ = 0$ and $\delta_1 = 0$, i.e. the primal subproblem (3.2) is assumed to be solved exactly, then $\bar{\lambda}_+ = \lambda_+$, $\bar{\lambda} = \lambda$ and (4.15) reduces to:

$$(4.16) \quad \lambda_+ \leq (1 - 2\Delta^* - \lambda)^{-2} (\lambda + \Delta^*)^2,$$

provided that $\lambda + 2\Delta^* < 1$.

4.4.2. Maximum central path neighborhood. The key point of the path-following algorithm is to determine the maximum central path neighborhood $(\beta_*, \beta^*) \subseteq (0, 1)$ such that for any $\beta \in (\beta_*, \beta^*)$, if $\bar{\lambda} \leq \beta$ then $\bar{\lambda}_+ \leq \beta$. Now, we analyze the estimate (4.15) to find the parameters $\bar{\delta}$ and Δ such that the last condition holds.

Suppose that $\bar{\delta} \geq 0$ as in Definition 4.1. First, we construct the following parametric cubic polynomial:

$$(4.17) \quad \mathcal{P}_{\bar{\delta}}(\beta) := c_0(\bar{\delta}) + c_1(\bar{\delta})\beta + c_2(\bar{\delta})\beta^2 + c_3(\bar{\delta})\beta^3,$$

where the coefficients are given by $c_0(\bar{\delta}) := -2\bar{\delta}(1 - \bar{\delta})^2 \leq 0$, $c_1(\bar{\delta}) := (1 - \bar{\delta})^{-1}[1 - 3\bar{\delta} + \bar{\delta}^4]$, $c_2(\bar{\delta}) := \bar{\delta}[(1 - \bar{\delta})^{-2} + 2(1 - \bar{\delta})^{-1}] - 3 + 2\bar{\delta}(1 - \bar{\delta})$ and $c_3(\bar{\delta}) := 1 - \bar{\delta} > 0$. Then we define:

$$(4.18) \quad p := \bar{\delta}[(1 - \bar{\delta})^{-2} + 2(1 - \bar{\delta})^{-1}], \quad q := (1 - \bar{\delta})\beta - 2\bar{\delta} \text{ and } \theta := 0.5(\sqrt{p^2 + 4q} - p).$$

The following theorem provides the conditions such that if $\bar{\lambda} \leq \beta$ then $\bar{\lambda}_+ \leq \beta$.

THEOREM 4.4. *Suppose that $\bar{\delta} \in [0, \bar{\delta}_{\max}] := [0, 0.043286]$ is fixed and θ is defined by (4.18). Then the polynomial $\mathcal{P}_{\bar{\delta}}$ defined by (4.17) has three nonnegative real roots*

$0 \leq \beta_* < \beta^* < \beta_3$. Moreover, if we choose $\beta \in (\beta_*, \beta^*)$ and compute $\bar{\Delta} := \frac{\theta(1-\bar{\delta}-\beta)-\beta}{1+2\theta}$ then $\bar{\Delta} > 0$ and, for $0 \leq \delta_+ \leq \bar{\delta}$, $0 \leq \delta_1 \leq \bar{\delta}$ and $0 \leq \Delta \leq \bar{\Delta}$, the condition $\lambda \leq \beta$ implies $\bar{\lambda}_+ \leq \beta$.

The proof of this theorem is postponed to the appendix. Now, we illustrate the variation of the values of β_* , β^* and $\bar{\Delta}$ w.r.t. $\bar{\delta}$ in Figure 4.1. The left figure shows the values of β_* (solid) and β^* (dash) and the right one plots the value of $\bar{\Delta}$ when β is chosen by $\beta := \frac{\beta_* + \beta^*}{2}$ (dash) and $\beta := \frac{\beta^*}{4}$ (solid), respectively.

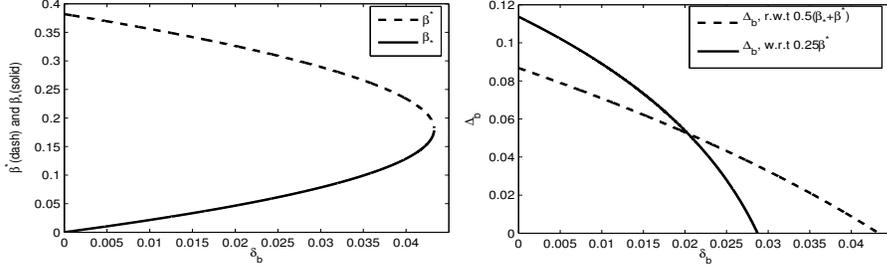


FIG. 4.1. The values of β_* , β^* and $\bar{\Delta}$ varying w.r.t. $\bar{\delta}$.

4.4.3. The update rule of the barrier parameter. It remains to quantify the decrement Δt of the barrier parameter t in (4.11). The following lemma shows how to update t .

LEMMA 4.5. Let $\bar{\delta}$ and $\bar{\Delta}$ be defined as in Theorem 4.4 and let:

$$(4.19) \quad \bar{\Delta}^* := \frac{1}{2} \left[(1 - \bar{\delta})\bar{\Delta} - \bar{\delta} + 1 - \sqrt{((1 - \bar{\delta})\bar{\Delta} - \bar{\delta} - 1)^2 + 4\bar{\delta}} \right].$$

Then the barrier parameter t can be decreased linearly, i.e. $t_+ := (1 - \sigma)t$, where $\sigma := [\sqrt{\nu} + \bar{\Delta}^*(\sqrt{\nu} + 1)]^{-1} \bar{\Delta}^* \in (0, 1)$.

Proof. It follows from (3.9) that $c + A^T y - t \nabla F(x^*) = 0$ and $c + A^T y - t_+ \nabla F(x_1^*) = 0$, where $x^* := x^*(y, t)$ and $x_1^* := x^*(y, t_+)$. Subtracting these equalities and then using $t_+ = t - \Delta t$, we have $t_+ [\nabla F(x_1^*) - \nabla F(x^*)] = \Delta t \nabla F(x^*)$. Using this relation together with [16, Theorem 4.1.7] and $\|\nabla F(x^*)\|_{x^*}^* \leq \sqrt{\nu}$ (see [16, inequality 4.2.4]), we have:

$$\begin{aligned} \frac{t_+ \|x_1^* - x^*\|_{x^*}^2}{1 + \|x_1^* - x^*\|_{x^*}^*} &\leq t_+ [\nabla F(x_1^*) - \nabla F(x^*)]^T (x_1^* - x^*) = \Delta t \nabla F(x^*)^T (x_1^* - x^*) \\ &\leq \Delta t \|\nabla F(x^*)\|_{x^*}^* \|x_1^* - x^*\|_{x^*} \leq \Delta t \sqrt{\nu} \|x_1^* - x^*\|_{x^*}. \end{aligned}$$

By the definition of Δ^* in (4.14), if $t > (\sqrt{\nu} + 1)\Delta t$, then the above inequality leads to:

$$(4.20) \quad \Delta^* \leq \bar{\Delta}^* := t [t - (\sqrt{\nu} + 1)\Delta t]^{-1} \sqrt{\nu} \Delta.$$

Therefore,

$$(4.21) \quad \Delta t = t [\sqrt{\nu} + (\sqrt{\nu} + 1)\bar{\Delta}^*]^{-1} \bar{\Delta}^*.$$

On the other hand, using the definitions of Δ and δ , we have:

$$\begin{aligned} \Delta &:= \|\bar{x}_{\delta_1} - \bar{x}_{\bar{\delta}}\|_{\bar{x}_{\bar{\delta}}} \stackrel{(A.10)}{\leq} (1 - \delta)^{-1} \left[\|\bar{x}_{\delta_1} - x_1^*\|_{x^*} + \|x_1^* - x^*\|_{x^*} + \|x^* - \bar{x}_{\bar{\delta}}\|_{x^*} \right] \\ (4.22) \quad &\leq (1 - \delta)^{-1} [(1 - \Delta^*)^{-1} \delta_1 + \Delta^* + \delta] \\ &\stackrel{(4.20), \delta, \delta_1 \leq \bar{\delta}}{\leq} (1 - \bar{\delta})^{-1} [(1 - \bar{\Delta}^*)^{-1} \bar{\delta} + \bar{\Delta}^* + \bar{\delta}]. \end{aligned}$$

Now, we need to find a condition such that $\Delta \leq \bar{\Delta}$, where $\bar{\Delta}$ is given in Theorem 4.4. It follows from (4.22) that $\Delta \leq \bar{\Delta}$ if $\frac{\bar{\delta}}{1-\Delta^*} + \Delta^* \leq (1-\bar{\delta})\bar{\Delta} - \bar{\delta}$. The last condition holds if:

$$(4.23) \quad 0 \leq \bar{\Delta}^* \leq \frac{1}{2} \left[(1-\bar{\delta})\bar{\Delta} - \bar{\delta} + 1 - \sqrt{((1-\bar{\delta})\bar{\Delta} - \bar{\delta} - 1)^2 + 4\bar{\delta}} \right],$$

provided that $\bar{\delta} \leq \frac{\bar{\Delta}}{1+\bar{\Delta}}$ due to (4.20). Thus, we can fix $\bar{\Delta}^*$ at the upper bound as defined in (4.19). By (4.21), the update rule for the barrier parameter t becomes $t_+ := t - \sigma t = (1-\sigma)t$ where $\sigma := \frac{\bar{\Delta}^*}{\sqrt{\nu+\bar{\Delta}^*}(\sqrt{\nu+1})} \in (0, 1)$. \square

Finally, we show that the conditions given in Theorem 4.4 and Lemma 4.5 are well-defined. Indeed, let us fix $\bar{\delta} := 0.01$. Then we can compute the values of β_* and β^* as $\beta_* \approx 0.021371 < \beta^* \approx 0.356037$. Therefore, if we choose $\beta := \frac{\beta^*}{4} \approx 0.089009 > \beta_*$ then $\bar{\Delta} \approx 0.089012$ and $\bar{\Delta}^* \approx 0.067399$.

4.4.4. The algorithm and its convergence. Before presenting the algorithm, we need to find a stopping criterion. By using Lemma A.1c) with $\Delta \leftarrow \delta$, we have:

$$(4.24) \quad \lambda \leq (1-\delta)^{-1}(\bar{\lambda} + \delta),$$

provided that $\delta < 1$ and $\bar{\lambda} \leq \beta < 1$. Consequently, if $\bar{\lambda} \leq (1-\bar{\delta})\beta - \bar{\delta}$ then $\lambda \leq \beta$. Let us define $\vartheta := (1-\bar{\delta})\beta - \bar{\delta}$, where $0 < \bar{\delta} < \beta/(\beta+1)$. It follows from Lemma 3.6 that if $t\omega_*(\vartheta) \leq \varepsilon_d$ for a given tolerance $\varepsilon_d > 0$, then y is an ε_d -solution of (3.4).

The second phase of the algorithmic framework presented in Subsection 4.2 is now described in detail as follows.

ALGORITHM 1. (*Path-following algorithm with IPFNT iterations*).

Initialization: Choose $\bar{\delta} \in [0, \bar{\delta}_{\max}]$ and compute β_* and β^* as in Theorem 4.4.

Phase 1. Apply Algorithm 2 presented in Subsection 4.5 below to find $y^0 \in Y$ such that $\lambda_{\bar{\delta}(\cdot, t_0)}(y^0) \leq \beta$.

Phase 2.

Initialization of Phase 2: Perform the following steps:

1. Given a tolerance $\varepsilon_d > 0$.
2. Compute $\bar{\Delta}$ as in Theorem 4.4. Then, compute $\bar{\Delta}^*$ by (4.19).
3. Compute $\sigma := \frac{\bar{\Delta}^*}{\sqrt{\nu+(\sqrt{\nu+1})\bar{\Delta}^*}}$ and the accuracy factor $\gamma := \frac{\bar{\delta}}{(\nu+2\sqrt{\nu})(1+\bar{\delta})}$.

Iteration: For $k = 0, 1, \dots, k_{\max}$ perform the following steps:

1. If $t_k \leq \frac{\varepsilon_d}{\omega_*(\vartheta)}$, where $\vartheta := (1-\bar{\delta})\beta - \bar{\delta}$, then terminate.
2. Compute an accuracy $\varepsilon_k := \gamma t_k$ for the primal subproblems.
3. Update $t_{k+1} := (1-\sigma)t_k$.
4. Solve approximately (2.2) *in parallel* up to the accuracy ε_k to obtain $\bar{x}_{\bar{\delta}}(y^k, t_{k+1})$.
5. Compute $\nabla d_{\bar{\delta}}(y^k, t_{k+1})$ and $\nabla^2 d_{\bar{\delta}}(y^k, t_{k+1})$ as in (4.3).
6. Update y^{k+1} as $y^{k+1} := y^k - \nabla^2 d_{\bar{\delta}}(y^k, t_{k+1})^{-1} \nabla d_{\bar{\delta}}(y^k, t_{k+1})$.

End of For.

END.

The core steps of Phase 2 in Algorithm 1 are Steps 4 and 6, where we need to solve M convex primal subproblems *in parallel* and computing the IPFNT direction, respectively. Note that Step 6 requires one to solve a linear equation system. In addition, the quantity $\nabla^2 F(\bar{x}_{\bar{\delta}}(y^k, t_{k+1}))$ can also be computed *in parallel*.

The parameter t at Step 3 can be updated adaptively as $t_{k+1} := (1-\sigma_k)t_k$, where $\sigma_k := \frac{\bar{\Delta}^*}{R_{\bar{\delta}}+(\sqrt{\nu+1})\bar{\Delta}^*}$ and $R_{\bar{\delta}} := (1-\bar{\delta})^{-1}[\bar{\delta}(1-\bar{\delta})^{-1} + \|\nabla F(\bar{x}_{\bar{\delta}})\|_{\bar{x}_{\bar{\delta}}}^*]$. The

stopping criterion at Step 1 can be replaced by $\omega_*(\vartheta_k)t_k \leq \varepsilon_d$, where $\vartheta_k := (1 - \bar{\delta})^{-1}[\lambda_{\bar{d}_{\bar{\delta}}(\cdot, t_k)}(y^k) + \bar{\delta}]$ due to Lemma 3.6 and (4.24).

Let us define $\lambda_{k+1} := \lambda_{\bar{d}_{\bar{\delta}}(\cdot, t_{k+1})}(y^{k+1})$ and $\lambda_k := \lambda_{\bar{d}_{\bar{\delta}}(\cdot, t_k)}(y^k)$. Then the local convergence of Algorithm 1 is stated in the following theorem.

THEOREM 4.6. *Let $\{(y^k, t_k)\}$ be a sequence generated by Algorithm 1. Then the number of iterations to obtain an ε_d -solution of (3.4) does not exceed:*

$$(4.25) \quad k_{\max} := \left\lceil [\ln(1 - \sigma)]^{-1} \ln \left(\frac{\varepsilon_d}{t_0 \omega_*(\vartheta)} \right) \right\rceil + 1,$$

where $\sigma := \frac{\bar{\Delta}^*}{\sqrt{\nu} + (\sqrt{\nu} + 1)\bar{\Delta}^*} \in (0, 1)$ and $\vartheta := (1 - \bar{\delta})\beta - \bar{\delta} \in (0, 1)$.

Proof. Note that y^k is an ε_d -solution of (3.4) if $t_k \leq \frac{\varepsilon_d}{\omega_*(\vartheta)}$ due to Lemma 3.6, where $\vartheta = (1 - \bar{\delta})\beta - \bar{\delta}$. Since $t_k = (1 - \sigma)^k t_0$ due to Step 3, we require $(1 - \sigma)^k \leq \frac{\varepsilon_d}{t_0 \omega_*(\vartheta)}$. Consequently, we obtain (4.25). \square

REMARK 2 (The worst-case complexity). *Since $(1 - \sigma)^{-1} = 1 + \frac{\bar{\Delta}^*}{\sqrt{\nu}(\bar{\Delta}^* + 1)}$, we have $-\ln(1 - \sigma) \approx \sigma = \frac{\bar{\Delta}^*}{\sqrt{\nu}(\bar{\Delta}^* + 1)}$. It follows from Theorem 4.6 that the complexity of Algorithm 1 is $O(\sqrt{\nu} \ln \frac{t_0}{\varepsilon_d})$.*

REMARK 3 (Linear convergence). *The sequence $\{t_k\}$ linearly converges to zero with a contraction factor not greater than $1 - \sigma$. When $\lambda_{\bar{d}_{\bar{\delta}}(\cdot, t)}(y) \leq \beta$, it follows from (3.11) that $\lambda_{d_{\bar{\delta}}(\cdot, t)}(y) \leq \beta\sqrt{t}$. Thus the sequence of the Newton decrements $\{\lambda_{d(\cdot, t_k)}(y^k)\}_k$ of d also converges linearly to zero with a contraction factor not greater than $\sqrt{1 - \sigma}$.*

REMARK 4 (The inexactness of the IPFNT direction). *Note that we can also apply an inexact method to solve the linear system for computing an IPFNT direction in (4.11). For more details of this approach, one can refer to [23].*

Finally, as a consequence of Theorem 4.6, the following corollary shows how to recover the optimality and feasibility of the original primal-dual problems (1.1)-(2.1).

COROLLARY 4.7. *Suppose that (y^k, t_k) is the output of Algorithm 1 and $x^*(y^k, t_k)$ is the solution of the primal subproblem (3.2). Then $(x^*(y^k, t_k), y^k)$ is an ε_p -solution of (1.1)-(2.1), where $\varepsilon_p := \nu \omega_*(\beta)^{-1} \varepsilon_d$.*

4.5. Phase 1 - Finding a starting point. Phase 1 of the algorithmic framework aims to find $y^0 \in Y$ such that $\lambda_{\bar{d}_{\bar{\delta}}(\cdot, t)}(y^0) \leq \beta$. In this subsection, we consider an inexact perturbed damped Newton (IPDNT) method for finding such a point y^0 .

4.5.1. Inexact perturbed damped Newton iteration. For a given $t = t_0 > 0$ and an accuracy $\bar{\delta} \geq 0$, let us assume that the current point $y \in Y$ is given, we compute the new point y_+ by applying the IPDNT iteration as follows:

$$(4.26) \quad y_+ := y - \alpha(y) \nabla^2 d_{\bar{\delta}}(y, t_0)^{-1} \nabla d_{\bar{\delta}}(y, t_0),$$

where $\alpha := \alpha(y) > 0$ is the step size which will be chosen appropriately. Note that since (4.26) is invariant under linear transformations, it is equivalent to:

$$(4.27) \quad y_+ := y - \alpha(y) \nabla^2 \tilde{d}_{\bar{\delta}}(y, t_0)^{-1} \nabla \tilde{d}_{\bar{\delta}}(y, t_0),$$

It follows from (3.11) that $\tilde{d}(\cdot, t_0)$ is standard self-concordant, and by [16, Theorem 4.1.8], we have:

$$(4.28) \quad \tilde{d}(y_+, t_0) \leq \tilde{d}(y, t_0) + \nabla \tilde{d}(y, t_0)^T (y_+ - y) + \omega_*(\|y_+ - y\|_y),$$

provided that $\|y_+ - y\|_y < 1$. On the other hand, (4.7) implies that:

$$(4.29) \quad 0 \leq \omega(\delta(\bar{x}_{\bar{\delta}}, x^*)) \leq \tilde{d}(y, t_0) - \tilde{d}_{\bar{\delta}}(y, t_0) \leq \omega_*(\delta(\bar{x}_{\bar{\delta}}, x^*)),$$

which bounds the error between $\tilde{d}(\cdot, t_0)$ and $\tilde{d}_{\bar{\delta}}(\cdot, t_0)$. In order to analyze the convergence of the IPDNT iteration (4.26) we denote by:

$$(4.30) \quad \begin{aligned} \hat{\delta}_+ &:= \|\bar{x}_{\bar{\delta}}(y_+, t_0) - x^*(y_+, t_0)\|_{x^*(y_+, t_0)}, \\ \hat{\delta} &:= \|\bar{x}_{\bar{\delta}}(y, t_0) - x^*(y, t_0)\|_{x^*(y, t_0)}, \\ \bar{\lambda}_0 &:= \lambda_{\tilde{d}_{\bar{\delta}}(\cdot, t_0)}(y) = \alpha(y)\|y_+ - y\|_y, \end{aligned}$$

the solution differences of $d(\cdot, t_0)$ and $d_{\bar{\delta}}(\cdot, t_0)$ and the Newton decrement of $\tilde{d}_{\bar{\delta}}(\cdot, t_0)$, respectively.

4.5.2. Finding the step size. The following lemma provides a formula to update the step size $\alpha(y)$ in (4.26).

LEMMA 4.8. *Let $0 < \bar{\delta} < \hat{\delta}^* := \beta(2 + \beta + 2\sqrt{\beta + 1})^{-1}$ and $\underline{\eta}$ be defined as:*

$$(4.31) \quad \underline{\eta} := \beta \left[(1 + \bar{\delta})\beta + \sqrt{(1 - \bar{\delta})^2\beta^2 - 4\bar{\delta}\beta} \right]^{-1} \left[(1 - \bar{\delta})\beta - 2\bar{\delta} + \sqrt{(1 - \bar{\delta})^2\beta^2 - 4\bar{\delta}\beta} \right].$$

Then $\underline{\eta} \in (0, 1)$. Furthermore, if we choose the step size $\alpha(y)$ as:

$$(4.32) \quad \alpha(y) := [2\bar{\lambda}_0(1 + \bar{\lambda}_0)]^{-1} \left[(1 - \bar{\delta})\bar{\lambda}_0 - 2\bar{\delta} + \sqrt{(1 - \bar{\delta})^2\bar{\lambda}_0^2 - 4\bar{\delta}\bar{\lambda}_0} \right].$$

then $\alpha(y) \in (0, 1)$ and:

$$(4.33) \quad \tilde{d}_{\bar{\delta}}(y_+, t_0) \leq \tilde{d}_{\bar{\delta}}(y, t_0) - \omega(\underline{\eta}).$$

As a consequence, if $\bar{\delta} = 0$ then $\underline{\eta} = \beta$ and $\alpha(y) := (1 + \bar{\lambda}_0)^{-1}$.

Proof. Let $p := y_+ - y$. From (4.28) and (4.29), we have:

$$(4.34) \quad \begin{aligned} \tilde{d}_{\bar{\delta}}(y_+, t_0) &\stackrel{(4.29)}{\leq} \tilde{d}(y_+, t_0) \stackrel{(4.28)}{\leq} \tilde{d}(y, t_0) + \nabla \tilde{d}(y, t_0)^T (y_+ - y) + \omega_*(\|y_+ - y\|_y) \\ &\stackrel{(4.29)}{\leq} \tilde{d}_{\bar{\delta}}(y, t_0) + \nabla \tilde{d}(y, t_0)^T (y_+ - y) + \omega_*(\|y_+ - y\|_y) + \omega_*(\hat{\delta}) \\ &= \tilde{d}_{\bar{\delta}}(y, t_0) + \nabla \tilde{d}_{\bar{\delta}}(y, t_0)^T p + [\nabla \tilde{d}(y, t_0) - \nabla \tilde{d}_{\bar{\delta}}(y, t_0)]^T p + \omega_*(\|p\|_y) + \omega_*(\hat{\delta}) \\ &\stackrel{(4.26)}{\leq} \tilde{d}_{\bar{\delta}}(y, t_0) - \alpha\bar{\lambda}_0^2 + \|\nabla \tilde{d}(y, t_0) - \nabla \tilde{d}_{\bar{\delta}}(y, t_0)\|_y^* \|p\|_y + \omega_*(\|p\|_y) + \omega_*(\hat{\delta}) \\ &\stackrel{(A.9)}{\leq} \tilde{d}_{\bar{\delta}}(y, t_0) - \alpha\bar{\lambda}_0^2 + \hat{\delta}\|p\|_y + \omega_*(\|p\|_y) + \omega_*(\hat{\delta}). \end{aligned}$$

Furthermore, from (A.11) and the definition of $\nabla^2 \tilde{d}$ and $\nabla^2 \tilde{d}_{\bar{\delta}}$, we have:

$$(1 - \hat{\delta})\nabla^2 \tilde{d}_{\bar{\delta}}(y, t_0) \preceq \nabla^2 \tilde{d}(y, t_0) \preceq (1 - \hat{\delta})^{-2} \nabla^2 \tilde{d}_{\bar{\delta}}(y, t_0).$$

These inequalities imply $(1 - \hat{\delta})\|p\|_y \leq \|p\|_y \leq (1 - \hat{\delta})^{-1}\|p\|_y$. Combining the previous inequalities, (4.27) and the definition of $\bar{\lambda}_0$ in (4.30) we get:

$$\alpha(1 - \hat{\delta})\bar{\lambda}_0 \leq \|p\|_y \leq \alpha(1 - \hat{\delta})^{-1}\bar{\lambda}_0.$$

Let us assume that $\alpha\bar{\lambda}_0 + \hat{\delta} < 1$. By substituting the second inequality into (4.34) and observing that the right hand side of (4.34) is nondecreasing w.r.t. $\|p\|_y$, we get:

$$(4.35) \quad \tilde{d}_{\bar{\delta}}(y_+, t_0) \leq \tilde{d}_{\bar{\delta}}(y, t_0) - \alpha\bar{\lambda}_0^2 + (1 - \hat{\delta})^{-1}\alpha\bar{\lambda}\hat{\delta} + \omega_* \left((1 - \hat{\delta})^{-1}\alpha\bar{\lambda}_0 \right) + \omega_*(\hat{\delta}).$$

Now, let us simplify the last four terms of (4.35) which we denote by $[\cdot]_{[1]}$ as follows:

$$(4.36) \quad \begin{aligned} [\cdot]_{[1]} &:= -\alpha\bar{\lambda}_0^2 + (1 - \hat{\delta})^{-1}\alpha\bar{\lambda}_0\hat{\delta} + \omega_* \left((1 - \hat{\delta})^{-1}\alpha\bar{\lambda}_0 \right) + \omega_*(\hat{\delta}) \\ &= -\alpha\bar{\lambda}_0^2 - (\alpha\bar{\lambda}_0 + \hat{\delta}) - \ln \left[1 - (\alpha\bar{\lambda}_0 + \hat{\delta}) \right] \\ &= -\alpha\bar{\lambda}_0^2 + \omega_*(\alpha\bar{\lambda}_0 + \hat{\delta}). \end{aligned}$$

Suppose that we can choose $\eta > 0$ such that $\alpha\bar{\lambda}_0^2 - \omega_*(\alpha\bar{\lambda}_0 + \hat{\delta}) = \omega(\eta)$. This condition leads to $\alpha\bar{\lambda}_0^2 = (\alpha\bar{\lambda}_0 + \hat{\delta}) \left[\alpha(\bar{\lambda}_0 + \bar{\lambda}_0) + \hat{\delta} \right]$ which is equivalent to:

$$(4.37) \quad \alpha = [2\bar{\lambda}_0(1 + \bar{\lambda}_0)]^{-1} \left[(1 - \hat{\delta})\bar{\lambda}_0 - 2\hat{\delta} + \sqrt{(1 - \hat{\delta})^2\bar{\lambda}_0^2 - 4\hat{\delta}\bar{\lambda}_0} \right],$$

provided that $0 \leq \hat{\delta} < \bar{\delta} := \frac{2 + \bar{\lambda}_0 - 2\sqrt{1 + \bar{\lambda}_0}}{\bar{\lambda}_0}$. Consequently, we deduce:

$$\eta = \bar{\lambda}_0 \left[(1 + \hat{\delta})\bar{\lambda}_0 + \sqrt{(1 - \hat{\delta})^2\bar{\lambda}_0^2 - 4\hat{\delta}\bar{\lambda}_0} \right]^{-1} \left[(1 - \hat{\delta})\bar{\lambda}_0 - 2\hat{\delta} + \sqrt{(1 - \hat{\delta})^2\bar{\lambda}_0^2 - 4\hat{\delta}\bar{\lambda}_0} \right].$$

We assume that $\bar{\lambda}_0 \geq \beta$ for a given $\beta \in (0, 1)$. Let us fix $\bar{\delta}$ such that:

$$0 < \bar{\delta} < \hat{\delta}^* := \beta^{-1} [2 + \beta - 2\sqrt{1 + \beta}] = [2 + \beta + 2\sqrt{1 + \beta}]^{-1}\beta.$$

If we choose the step size $\alpha(y)$ as in (4.32) for the IPDNT iteration (4.26) then we obtain (4.33) with η defined by (4.31). \square

Finally, we estimate the constant $\underline{\eta}$ for the case $\beta \approx 0.089009$. We first obtain $\hat{\delta}^* \approx 0.021314$. Let $\bar{\delta} = \frac{1}{2}\hat{\delta}^* \approx 0.010657$. Then we get $\underline{\eta} \approx 0.075496$ and $\omega(\underline{\eta}) \approx 0.003002$.

4.5.3. The algorithm and its worst-case complexity. In summary, the algorithm for finding $y^0 \in Y$ is presented in detail as follows.

ALGORITHM 2. (*Finding a starting point $y^0 \in Y$*).

Initialization: Perform the following steps:

1. Select $\beta \in (\beta_*, \beta^*)$ and $t_0 > 0$ as desired (e.g. $\beta = \frac{1}{4}\beta^* \approx 0.089009$).
2. Take an arbitrary point $y^{0,0} \in Y$.
3. Compute $\hat{\delta}^* := \beta[2 + \beta + 2\sqrt{1 + \beta}]^{-1}$ and fix $\bar{\delta} \in (0, \hat{\delta}^*)$ (e.g. $\bar{\delta} = 0.5\hat{\delta}^*$).
4. Compute an accuracy $\varepsilon_0 := \frac{t_0\bar{\delta}}{2(\nu + 2\sqrt{\nu})(1 + \bar{\delta})}$.

Iteration: For $j = 0, 1, \dots, j_{\max}$, perform the following steps:

1. Solve approximately (3.2) *in parallel* up to the accuracy ε_0 to obtain $\bar{x}_{\bar{\delta}}(y^{0,j}, t_0)$.
2. Compute $\bar{\lambda}_j := \bar{\lambda}_{\bar{\delta}(\cdot, t_0)}(y^{0,j})$.
3. If $\bar{\lambda}_j \leq \beta$ then set $y^0 := y^{0,j}$ and terminate.
4. Update $y^{0,j+1}$ as $y^{0,j+1} := y^{0,j} - \alpha_j \nabla^2 d_{\bar{\delta}}(y^{0,j}, t_0)^{-1} \nabla d_{\bar{\delta}}(y^{0,j}, t_0)$, where $\alpha_j := [2\bar{\lambda}_j(1 + \bar{\lambda}_j)]^{-1} \left[(1 - \bar{\delta})\bar{\lambda}_j - 2\bar{\delta} + \sqrt{(1 - \bar{\delta})^2\bar{\lambda}_j^2 - 4\bar{\delta}\bar{\lambda}_j} \right] \in (0, 1)$.

End of For.

END.

The convergence of this algorithm is stated in the following theorem.

THEOREM 4.9. *The number of iterations required in Algorithm 2 does not exceed:*

$$(4.38) \quad j_{\max} := \left\lceil [t_0 \omega(\underline{\eta})]^{-1} [d_{\bar{\delta}}(y^{0,0}, t_0) - d^*(t_0) + \omega_*(\bar{\delta})] \right\rceil + 1,$$

where $d^*(t_0) = \min_{y \in Y} d(y, t_0)$ and $\underline{\eta}$ is given by (4.31).

Proof. Summing up (4.33) from $j = 0$ to $j = k$ and then using (4.29) we have $0 \leq \tilde{d}(y^{0,k}, t_0) - \tilde{d}^*(t_0) \leq \tilde{d}_{\bar{\delta}}(y^{0,k}, t_0) + \omega_*(\bar{\delta}) - \tilde{d}^*(t_0) \leq \tilde{d}_{\bar{\delta}}(y^{0,0}, t_0) + \omega_*(\bar{\delta}) - \tilde{d}^*(t_0) - k\omega(\underline{\eta})$. This inequality together with (3.11) and (4.4) imply:

$$j \leq [t_0 \omega(\underline{\eta})]^{-1} [d_{\bar{\delta}}(y^{0,0}, t_0) - d^*(t_0) + \omega_*(\bar{\delta})].$$

Hence, the maximum iteration number in Algorithm 2 does not exceed j_{\max} defined by (4.38). \square

Since $d^*(t_0)$ is unknown, the constant j_{\max} in (4.38) only gives an upper bound for Algorithm 2. However, in this algorithm, we do not use j_{\max} as a stopping criterion.

5. Path-following decomposition algorithm with exact Newton iterations. In Algorithm 1, if we set $\bar{\delta} = 0$, then this algorithm reduces to the ones considered in [9, 14, 21, 27, 28] as a special case. Note that, in [9, 14, 21, 27, 28], the primal subproblem (3.2) is assumed to be solved exactly so that the family $\{d(\cdot, t)\}_{t>0}$ of the smoothed dual functions is strongly self-concordant due to the Legendre transformation. Consequently, the standard theory of interior point methods in [15] can be applied to minimize this function. In contrast to those methods, in this section we analyze directly the path-following iterations to select appropriate parameters for implementation. Moreover, the radius of the central path neighbourhood in Algorithm 3 below is $(3 - \sqrt{5})/2 \approx 0.381966$ compared to the one in the literature, $2 - \sqrt{3} \approx 0.267949$.

5.1. The exact path-following iteration. Let us assume that the primal subproblem (3.2) is solved exactly, i.e. $\bar{\delta} = 0$ in Definition 4.1. Then, we have $\bar{x}_{\bar{\delta}} \equiv x^*$ and $\delta(\bar{x}_{\bar{\delta}}, x^*) = 0$ for all $y \in Y$ and $t > 0$. Moreover, it follows from (4.20) that $\Delta = \Delta^* = \|x^*(y, t_+) - x^*(y, t)\|_{x^*(y, t)}$. We consider one step of the path-following scheme with exact full-step Newton iterations:

$$(5.1) \quad \begin{cases} t_+ := t - \Delta t, \quad \Delta t > 0, \\ y_+ := y - \nabla^2 d(y, t_+)^{-1} \nabla d(y, t_+) \equiv y - \nabla^2 \tilde{d}(y, t_+)^{-1} \nabla \tilde{d}(y, t_+). \end{cases}$$

For the sake of notational simplicity, we denote by $\tilde{\lambda} := \lambda_{\tilde{d}(\cdot, t)}(y)$, $\tilde{\lambda}_1 := \lambda_{\tilde{d}(\cdot, t_+)}(y)$ and $\tilde{\lambda}_+ := \lambda_{\tilde{d}(\cdot, t_+)}(y_+)$. It follows from (4.16) of Lemma 4.3 that:

$$(5.2) \quad \tilde{\lambda}_+ \leq (1 - 2\Delta^* - \tilde{\lambda})^{-2} (\tilde{\lambda} + \Delta^*)^2.$$

Now, we fix $\beta \in (0, 1)$ and assume that $\tilde{\lambda} \leq \beta$. We need to find a condition on Δ such that $\tilde{\lambda}_+ \leq \beta$. Indeed, since the right-hand side of (5.2) is nondecreasing w.r.t. $\tilde{\lambda}$, it implies that $\tilde{\lambda}_+ \leq (1 - 2\Delta^* - \beta)^{-2} (\Delta^* + \beta)^2$. Thus if $\frac{\Delta^* + \beta}{1 - 2\Delta^* - \beta} \leq \sqrt{\beta}$ then $\tilde{\lambda}_+ \leq \beta$. The last condition leads to:

$$(5.3) \quad 0 \leq \Delta^* \leq \bar{\Delta}^* := (1 + 2\sqrt{\beta})^{-1} \sqrt{\beta} (1 - \sqrt{\beta} - \beta),$$

provided that:

$$(5.4) \quad 0 < \beta < \beta^* := (3 - \sqrt{5})/2 \approx 0.381966.$$

In particular, if we choose $\beta = \frac{\beta^*}{4} \approx 0.095492$ then $\bar{\Delta}^* \approx 0.113729$. Since $\Delta \equiv \Delta^*$, according to (4.21) and (5.1), we can update t as:

$$(5.5) \quad t_+ := (1 - \sigma)t, \quad \text{where } \sigma := [\sqrt{\nu} + (\sqrt{\nu} + 1)\bar{\Delta}]^{-1}\bar{\Delta} \in (0, 1).$$

5.2. The algorithm and its convergence. The exact variant of Algorithms 1 and 2 is presented in detail as follows.

ALGORITHM 3. (*Path-following algorithm with exact Newton iterations*).

Initialization: Given a tolerance $\varepsilon_d > 0$ and choose an initial value $t_0 > 0$. Fix a constant $\beta \in (0, \beta^*)$, where $\beta^* = \frac{3-\sqrt{5}}{2} \approx 0.381966$. Then, compute $\bar{\Delta} := \frac{\sqrt{\beta}(1-\sqrt{\beta-\beta})}{1+2\sqrt{\beta}}$ and $\sigma := \frac{\bar{\Delta}}{\sqrt{\nu}+(\sqrt{\nu}+1)\bar{\Delta}}$.

Phase 1. (*Finding a starting point*).

Choose an arbitrary starting point $y^{0,0} \in Y$.

For $j = 0, 1, \dots, \tilde{j}_{\max}$, perform the following steps:

1. Solve *exactly* the primal subproblem (3.2) *in parallel* to obtain $x^*(y^{0,j}, t_0)$.
2. Evaluate $\nabla d(y^{0,j}, t_0)$ and $\nabla^2 d(y^{0,j}, t_0)$ as in (3.10). Then compute the Newton decrement $\tilde{\lambda}_j = \lambda_{\tilde{d}(\cdot, t_0)}(y^{0,j})$.
3. If $\tilde{\lambda}_j \leq \beta$ then set $y^0 := y^{0,j}$ and terminate.
4. Update $y^{0,j+1}$ as $y^{0,j+1} := y^{0,j} - (1 + \tilde{\lambda}_j)^{-1} \nabla^2 d(y^{0,j}, t_0)^{-1} \nabla d(y^{0,j}, t_0)$.

End of For.

Phase 2. (*Path-following iterations*).

For $k = 0, 1, \dots, \tilde{k}_{\max}$ perform the following steps:

1. If $t_k \leq \frac{\varepsilon_d}{\omega_*(\beta)}$ then terminate.
2. Update t_k as $t_{k+1} := (1 - \sigma)t_k$.
3. Solve *exactly* the primal subproblem (3.2) *in parallel* to obtain $x^*(y^k, t_{k+1})$.
4. Evaluate $\nabla d(y^k, t_{k+1})$ and $\nabla^2 d(y^k, t_{k+1})$ as in (3.10).
5. Update y^{k+1} as $y^{k+1} := y^k + \Delta y^k = y^k - \nabla^2 d(y^k, t_{k+1})^{-1} \nabla d(y^k, t_{k+1})$.

End of For.

END.

Since $\tilde{d}(\cdot, t_0)$ is standard self-concordant due to Lemma 3.3. By [16, Theorem 4.1.12], the number of iterations required in Phase 1 does not exceed:

$$(5.6) \quad \tilde{j}_{\max} := \left\lceil \left[\tilde{d}(y^{0,0}, t_0) - \tilde{d}^*(t_0) \right] \omega(\beta)^{-1} \right\rceil + 1 = \left\lceil \left[d(y^{0,0}, t_0) - d^*(t_0) \right] [t_0 \omega(\beta)]^{-1} \right\rceil + 1.$$

The convergence of Phase 2 in Algorithm 3 is stated in the following theorem.

THEOREM 5.1. *The maximum number of iterations needed in Phase 2 of Algorithm 3 to obtain an ε_d -solution $y^{\tilde{k}}$ of (3.4) does not exceed:*

$$(5.7) \quad \tilde{k}_{\max} := \left\lceil \left[\ln \left(\frac{t_0 \omega_*(\beta)}{\varepsilon_d} \right) \right] \left[\ln \left(1 + \frac{\bar{\Delta}^*}{\sqrt{\nu}(\bar{\Delta}^* + 1)} \right) \right]^{-1} \right\rceil + 1,$$

where $\bar{\Delta}^*$ is defined by (5.3).

Proof. From Step 2 of Algorithm 3, we have $t_k = (1 - \sigma)^k t_0 = \left(1 + \frac{\bar{\Delta}^*}{\sqrt{\nu}(\bar{\Delta}^* + 1)} \right)^k t_0$. Algorithm 3 is terminated if $t_k \leq \frac{\varepsilon_d}{\omega_*(\beta)}$. Thus $\left(1 + \frac{\bar{\Delta}^*}{\sqrt{\nu}(\bar{\Delta}^* + 1)} \right)^k \leq \frac{\varepsilon_d}{t_0 \omega_*(\beta)}$, which leads

to (5.7). \square

REMARK 5 (The worst-case complexity). Since $\ln\left(1 + \frac{\bar{\Delta}^*}{\sqrt{\nu}(\bar{\Delta}^*+1)}\right) \approx \frac{\bar{\Delta}^*}{\sqrt{\nu}(\bar{\Delta}^*+1)}$, the worst-case complexity of Algorithm 3 is $O(\sqrt{\nu} \ln(t_0/\varepsilon_d))$.

REMARK 6 (Damped Newton iteration). Note that, at Step 5 of Algorithm 3, we can use a damped Newton iteration $y^{k+1} := y^k - \alpha_k \nabla^2 d(y^k, t_{k+1})^{-1} \nabla d(y^k, t_{k+1})$ instead of the full-step Newton iteration, where $\alpha_k = (1 + \lambda_{\bar{d}(\cdot, t_{k+1})}(y^k))^{-1}$. In this case, with the same argument as before, we can compute $\beta^* = 0.5$ and $\Delta^* = \frac{\sqrt{0.5\beta} - \beta}{1 + \sqrt{0.5\beta}}$.

6. Discussion on the implementation. In this section, we further discuss the implementation issues of the proposed algorithms.

6.1. Handling nonlinear objective function and local equality constraints.

If the objective function ϕ_i in (1.1) is nonlinear, concave and its epigraph is endowed with a self-concordant log-barrier for some $i \in \{1, \dots, M\}$ then we propose to use a slack variable to move the objective function into the constraints and reformulate it as an optimization problem with linear objective function. By elimination of variables, it is not difficult to show that the optimality condition of the resulting problem collapses to the optimality condition of the original problem, i.e. $\nabla \phi_i(x_i) + A_i^T y - t \nabla F_i(x_i) = 0$. The algorithms developed in the previous sections can be applied to solve such a problem without moving the nonlinear objective function into the constraints.

We also note that, in Algorithms 1 and 2, we need to solve approximately the primal subproblems in (3.2) up to a desired accuracy. Instead of solving directly these primal subproblems, we can treat them from the optimality condition (3.3). Since the objective function associated with this optimality condition is self-concordant, Newton-type methods can be applied to solve such a problem, see, e.g. [3, 16].

If local equality constraints $E_i x_i = f_i$ are considered in (1.1) for some $i \in \{1, \dots, M\}$, then the KKT conditions of the primal subproblem i become:

$$(6.1) \quad \begin{cases} c_i + A_i^T y + E_i^T z_i - t \nabla F_i(x_i) = 0, \\ E_i x_i - f_i = 0. \end{cases}$$

Instead of the full KKT system (6.1), we consider a reduced KKT condition as follows:

$$Z_i^T (c_i + A_i^T y) - t Z_i^T \nabla F_i(Z_i x_i^z + R_i^{-T} f_i) = 0.$$

Here, (Q_i, R_i) is a QR-factorization of E_i^T and $Q_i = [Y_i, Z_i]$ is a basis of the range space and the null space of E_i^T , respectively. Due to the invariance of the norm $\|\cdot\|_{x^*}$, we can show that $\|\bar{x}_{\bar{\delta}} - x^*\|_{x^*} = \|\bar{x}_{\bar{\delta}}^z - x^{z*}\|_{x^{z*}}$. Therefore, the condition (4.1) coincides with $\|\bar{x}_{\bar{\delta}}^z - x^{z*}\|_{x^{z*}} \leq \bar{\delta}$. However, the last condition is satisfied if:

$$\|Z_i^T (c_i + A_i^T y) - t Z_i^T \nabla F_i(Z_i x_i^z + R_i^{-T} f_i)\|_{x_i^{z^*}} \leq \varepsilon_i(t), \quad i = 1, \dots, M.$$

Note that the QR-factorization of E_i^T is only computed one time, a priori.

6.2. Computing the inexact perturbed Newton direction. Regarding to the Newton direction in Algorithms 1 and 2, one has to solve the linear system:

$$(6.2) \quad \nabla^2 d_{\bar{\delta}}(y^k, t) \Delta y^k = -\nabla d_{\bar{\delta}}(y^k, t).$$

Here, the gradient vector $\nabla d_{\bar{\delta}}(y^k, t) = A \bar{x}_{\bar{\delta}}(y^k, t) - b = \sum_{i=1}^M (A_i \bar{x}_i(y^k, t) - b_i) := g_k$ and the Hessian matrix $\nabla^2 d_{\bar{\delta}}(y^k, t)$ is obtained from:

$$\nabla^2 d_{\bar{\delta}}(y^k, t) = t^{-1} \sum_{i=1}^M A_i \nabla^2 F_i(\bar{x}_i(y^k, t))^{-1} A_i^T := \sum_{i=1}^M A_i G_i^k A_i^T.$$

Note that each block $G_i^k := t^{-1}A_i\nabla^2 F_i(\bar{x}_i(y^k, t))^{-1}A_i^T$ can be computed *in parallel*. Then, the linear system (6.2) can be written as:

$$(6.3) \quad \left(\sum_{i=1}^M A_i G_i^k A_i^T \right) \Delta y^k = -g_k.$$

Since matrix $G^k := \sum_{i=1}^M A_i G_i^k A_i^T \succ 0$, one can apply either Cholesky-type factorizations or conjugate gradient (CG) methods to solve (6.3). Note that the CG method only requires matrix-vector operations. More details on parallel solution of (6.3) can be found, e.g., in [14, 28].

7. Numerical Tests. In this paper, we test the algorithms developed in the previous sections by solving a routing problem with congestion cost. This problem appears in many areas including telecommunications, network and transportation [8].

Let $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ be a network of $n_{\mathcal{N}}$ nodes and $n_{\mathcal{A}}$ links, and \mathcal{C} be a set of $n_{\mathcal{C}}$ commodities to be sent through the network \mathcal{G} , where each commodity $k \in \mathcal{C}$ has a source $s_k \in \mathcal{N}$, a destination $d_k \in \mathcal{N}$ and a certain amount of demand r_k . The optimization model of the routing problem with congestion (RPC) can be formulated as follows (see, e.g. [8] for more details):

$$(7.1) \quad \begin{aligned} \min_{u_{ijk}, v_{ij}} \quad & \sum_{k \in \mathcal{C}} \sum_{(i,j) \in \mathcal{A}} c_{ij} u_{ijk} + \sum_{(i,j) \in \mathcal{A}} w_{ij} g_{ij}(v_{ij}) \\ \text{s.t.} \quad & \sum_{j: (i,j) \in \mathcal{A}} u_{ijk} - \sum_{j: (j,i) \in \mathcal{A}} u_{jik} = \begin{cases} r_k & \text{if } i = s_k, \\ -r_k & \text{if } i = d_k, \\ 0 & \text{otherwise,} \end{cases} \\ & \sum_{k \in \mathcal{C}} u_{ijk} - v_{ij} = b_{ij}, \quad (i, j) \in \mathcal{A}, \\ & u_{ijk} \geq 0, \quad v_{ij} \geq 0, \quad (i, j) \in \mathcal{A}, \end{aligned}$$

where $w_{ij} \geq 0$ is the weighting of the additional cost function g_{ij} for $(i, j) \in \mathcal{A}$.

In this example we assume that the additional cost function g_{ij} is given by either a) $g_{ij}(v_{ij}) = -\ln(v_{ij})$, the logarithmic function or b) $g_{ij}(v_{ij}) = v_{ij} \ln(v_{ij})$, the entropy function. It was shown in [16] that the epigraph of g_{ij} possesses a standard self-concordant barrier a) $F_{ij}(v_{ij}, s_{ij}) = -\ln v_{ij} - \ln(\ln v_{ij} + s_{ij})$ or b) $F_{ij}(v_{ij}, s_{ij}) = -\ln v_{ij} - \ln(s_{ij} - v_{ij} \ln v_{ij})$, respectively.

By using slack variables s_{ij} , we can move the nonlinear terms of the objective function to the constraints. The objective function of the resulting problem becomes:

$$(7.2) \quad f(u, v, s) := \sum_{k \in \mathcal{C}} \sum_{(i,j) \in \mathcal{A}} c_{ij} u_{ijk} + \sum_{(i,j) \in \mathcal{A}} w_{ij} s_{ij},$$

with additional constraints $g_{ij}(v_{ij}) \leq s_{ij}$, $(i, j) \in \mathcal{A}$. It is clear that problem (7.1) is separably convex. Let:

$$(7.3) \quad X_{ij} := \left\{ v_{ij} \geq 0, \sum_{k \in \mathcal{C}} u_{ijk} - v_{ij} = b_{ij}, g_{ij}(v_{ij}) \leq s_{ij}, (i, j) \in \mathcal{A}, k \in \mathcal{C} \right\}, \quad (i, j) \in \mathcal{A}.$$

Then problem (7.1) can be reformulated in the form of (1.1) with linear objective function (7.2) and the local constraint set (7.3). Moreover, the resulting problem has $M := n_{\mathcal{A}}$ components, $n := n_{\mathcal{C}} n_{\mathcal{A}} + 2n_{\mathcal{A}}$ variables including u_{ijk} , v_{ij} and s_{ij} ; and $m := n_{\mathcal{C}} n_{\mathcal{N}}$ coupling constraints. Each primal subproblem (3.2) has $n_i := n_{\mathcal{C}} + 2$ variables and one local linear equality constraint.

The aim is to compare the effect of the inexactness on the performance of the algorithms. We consider two variants of Algorithm 1, where we set $\bar{\delta} = 0.5\bar{\delta}^*$ and $\bar{\delta} = 0.25\bar{\delta}^*$ in Phase 1 and $\bar{\delta} = 0.01$ and $\bar{\delta} = 0.005$ in Phase 2, respectively. We denote these variants by **A1-v1** and **A1-v2**, respectively. For Algorithm 3, we also consider two cases. In the first case we set the tolerance of the primal subproblems to $\varepsilon_p = 10^{-6}$, and the second one is 10^{-10} , where we call them by **A3-v1** and **A3-v2**, respectively. All variants are terminated with the same tolerance $\varepsilon_d = 10^{-4}$. The initial barrier parameter value is set to $t_0 := 0.25$.

We benchmarked four variants with performance profiles [6]. Recall that a performance profile is built based on a set \mathcal{S} of n_s algorithms (solvers) and a collection \mathcal{P} of n_p problems. Suppose that we build a profile based on computational time. We denote by $T_{p,s} :=$ computational time required to solve problem p by solver s . We compare the performance of algorithm s on problem p with the best performance of any algorithm on this problem; that is we compute the performance ratio $r_{p,s} := \frac{T_{p,s}}{\min\{T_{p,\bar{s}} \mid \bar{s} \in \mathcal{S}\}}$. Now, let $\rho_s(\tau) := \frac{1}{n_p} \text{size}\{p \in \mathcal{P} \mid r_{p,s} \leq \tau\}$ for $\tau \in \mathbb{R}_+$. The function $\rho_s : \mathbb{R} \rightarrow [0, 1]$ is the probability for solver s that a performance ratio is within a factor τ of the best possible ratio. We use the term ‘‘performance profile’’ for the distribution function ρ_s of a performance metric. We can also plot the performance profiles in log-scale, i.e. $\rho_s(\tau) := \frac{1}{n_p} \text{size}\{p \in \mathcal{P} \mid \log_2(r_{p,s}) \leq \log_2 \tau\}$.

All the algorithms were implemented in C++ running on a PC Desktop Intel® Core TM2, Quad-Core Processor Q6600 (2.4GHz) and 3Gb RAM and were parallellized by using **OpenMP**. The input data is generated randomly, where the nodes of the network are generated in a rectangle $[0, 100] \times [0, 300]$, the demand r_k is in $[50, 500]$, the weighting vector w is set to 10, the congestion b_{ij} is in $[10, 100]$ and the linear cost c_{ij} is the Euclidean length of the link $(i, j) \in \mathcal{A}$. The nonlinear cost function g_{ij} is chosen randomly between two functions in a) and b) defined above with the same probability.

We tested the algorithms on a collection of 108 random problems. The size of these problems varies from $M = 6$ to 14, 280 components, $n = 84$ to 77, 142 variables and $m = 15$ to 500 coupling constraints. The performance profiles of the four algorithms in terms of computational time (in seconds) are shown in Figure 7.1, where the x -axis is the factor $\log_2 \tau$ (not more than 2^τ -times worse than the best one) and the y -axis is the probability function values $\rho_s(\tau)$ (problems ratio).

As we can see from Figure 7.1 that Algorithm 1 performs better than Algorithm 3 both in the total computational time and the time for solving the primal subproblems. This provides an evidence on the effect of the inexactness to the performance of the algorithm. We also observed that the numbers of iterations for solving the master problem in Phase 1 of all variants are almost similar, while they are different in Phase 2. However, since Phase 2 is performed when the approximate point is in the quadratic convergence region so that it requires few iterations toward the desired approximate solution. Therefore, the computational time of Phase 1 dominates Phase 2. We notice that, in this particular example, the structure of the master problem is almost dense and we did not use any sparse linear algebra solver.

We also compared the total number of iterations for solving the primal subproblems in Figure 7.2. It shows that Algorithm 1 is superior to Algorithm 3 in terms of iteration number, although the accuracy of solving the primal subproblem in Algorithm 3 is only set to 10^{-6} which is not exact as theoretically required. This performance profile also reveals the effect of the inexactness on the number of iterations. In our numerical results, the inexact version **A1-v1** saves 22% (resp. 23%) of

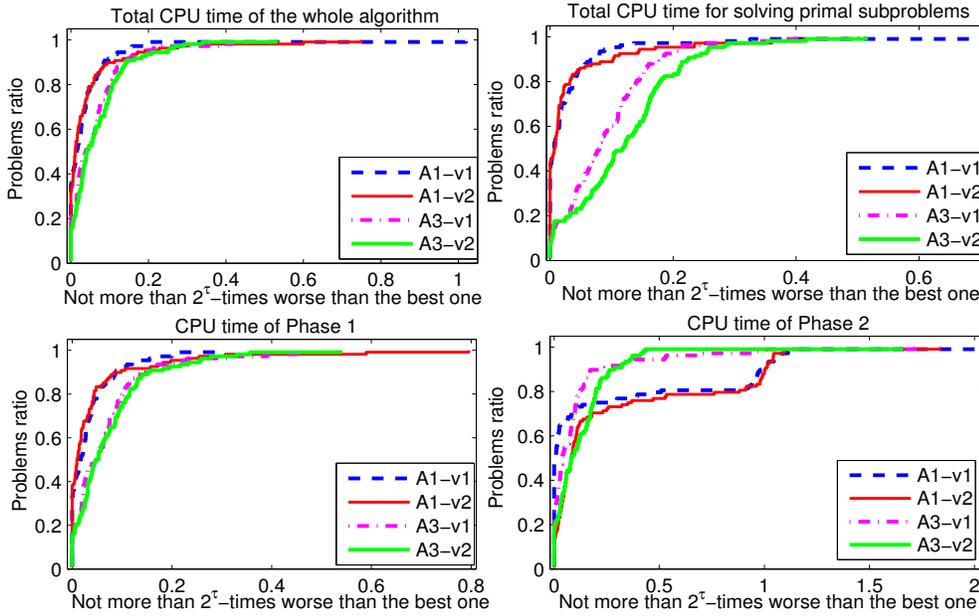


FIG. 7.1. The performance profiles of the four variants in terms of computational time.

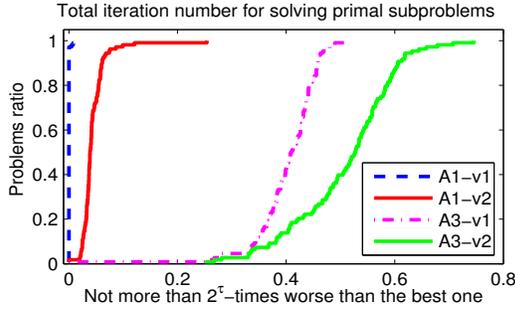


FIG. 7.2. The performance profile of the four variants in terms of iteration number.

the total number of iterations to solve the primal subproblems compared to A3-v1 (resp. A3-v2); while the variant A1-v2 saves 20% (reps. 21%) compared to A3-v1 (resp. A3-v2).

8. Concluding remarks. We have proposed a smoothing technique for Lagrangian decomposition using self-concordant barriers in large-scale separable convex optimization. We have proved a global and a local approximation for the dual function. Then, we proposed a path-following algorithm with inexact perturbed Newton iterations. The convergence of the algorithm has been analyzed and its complexity has been estimated. The theory presented in this paper is significant in practice, since it allows us to solve the primal subproblems inexactly. Moreover, we allow one to balance between the accuracy of solving the primal subproblem and the convergence rate of the path-following algorithm. As a special case, we have obtained again the path-following methods studied by Mehrotra [11] *et al* and Shida [21] with some additional advantages. Preliminary numerical tests confirm the advantages of the inexactness. Extensions to distributed implementation of linear algebra in the master problem are an interesting and significant future research direction.

Acknowledgments. The authors would like to thank the associate editor and two anonymous referees for their insightful comments and suggestions that helped us to improve this paper.

Research supported by Research Council KUL: CoE EF/05/006 Optimization in Engineering(OPTEC), IOF-SCORES4CHEM, GOA/10/009 (MaNet), GOA /10/11, several PhD/postdoc and fellow grants; Flemish Government: FWO: PhD / postdoc grants, projects G.0452.04, G.0499.04, G.0211.05, G.0226.06, G.0321.06, G.0302.07, G.0320.08, G.0558.08, G.0557.08, G.0588.09, G.0377.09, G.0712.11, research communities (ICCoS, ANMMM, MLDM); IWT: PhD Grants, Belgian Federal Science Policy Office: IUAP P6/04; EU: ERNSI; FP7-HDMPC, FP7-EMBOCON no. 248940, ERC-HIGHWIND, Contract Research: AMINAL. Other: Helmholtz-viCERP, COMET-ACCM, CNCS-UEFISCDI (TE, no. 19/11.08.2010); CNCS (PN II, no. 80EU/2010); POSDRU (no. 89/1.5/S/62557).

Appendix A. The proof of the technical statements. In this appendix, we provide a complete proof of Lemmas 3.1, 3.2 and 4.3, and Theorem 4.4.

A.1. The proof of Lemma 3.1. *Proof.* For notational simplicity, we denote by $x_i^* := x_i^*(y, t)$. The left-hand side of (3.5) follows from $F_i(x_i) - F_i(x_i^c) \geq \omega(\|x_i - x_i^c\|_{x_i^c}) \geq 0$ due to (3.1). We prove the right-hand side of (3.5). Since F_i is standard self-concordant and $x_i^c = \operatorname{argmin}_{x_i \in \operatorname{int}(X_i)} F_i(x_i)$, according to [16, Theorem 4.1.13] we have:

$$(A.1) \quad F_i(x_i^*) - F_i(x_i^c) \leq \omega_*(\lambda_{F_i}(x_i^*)).$$

provided that $\lambda_{F_i}(x_i^*) < 1$. Now, we prove (3.5). Let $x_i(\alpha) := x_i^* + \alpha(x_{0i}^*(y) - x_i^*)$ for $\alpha \in [0, 1]$. Since $x_i^* \in \operatorname{int}(X_i)$ and $\alpha < 1$, $x_i(\alpha) \in \operatorname{int}(X_i)$. By applying [15, inequality 2.3.3], we have $F_i(x_i(\alpha)) \leq F_i(x_i^*) - \nu_i \ln(1 - \alpha)$ which is equivalent to:

$$(A.2) \quad F_i(x_i(\alpha)) - F_i(x_i^c) \leq F_i(x_i^*) - F_i(x_i^c) - \nu_i \ln(1 - \alpha).$$

From the definition of $d_i(\cdot, t)$ and $d_{0i}(\cdot)$, the concavity of ϕ_i and (A.1) we have:

$$(A.3) \quad \begin{aligned} d_i(y, t) &\geq \max_{\alpha \in [0, 1]} \{ \phi_i(x_i(\alpha)) + y^T(A_i x_i(\alpha) - b_i) - t[F_i(x_i(\alpha)) - F_i(x_i^c)] \} \\ &\geq \max \left\{ \alpha[\phi_i(x_{0i}^*(y)) + y^T(A_i x_{0i}^*(y) - b_i)] + (1 - \alpha)[\phi_i(x_i^*) + y^T(A_i x_i^* - b_i)] \right. \\ &\quad \left. - t[F_i(x_i^*) - F_i(x_i^c)] + \nu_i t \ln(1 - \alpha) \mid \alpha \in [0, 1] \right\} \\ &\stackrel{(A.1)}{\geq} \max \left\{ \alpha d_{0i}(y) + (1 - \alpha)d_i(y, t) + t\nu_i \ln(1 - \alpha) - t\omega_*(\lambda_{F_i}(x_i^*)) \mid \alpha \in [0, 1] \right\}. \end{aligned}$$

By solving the last maximization problem in (A.3) we obtain the solution $\alpha^* = 0$ if $d_{0i}(y) - d_i(y, t) \leq t\nu_i$ and $\alpha^* = 1 - [d_{0i}(y) - d_i(y, t)]^{-1}\nu_i t$, otherwise. Substituting this solution into (A.3) we get:

$$(A.4) \quad d_{0i}(y) - d_i(y, t) \leq t\nu_i \left\{ 1 + \ln \left[(d_{0i}(y) - d_i(y, t)) / (t\nu_i) \right] + \omega_*(\lambda_{F_i}(x_i^*)) / \nu_i \right\},$$

provided that $d_{0i}(y) - d_i(y, t) > t\nu_i$. By rearranging (A.4) we obtain $d_{0i}(y) - d_i(y, t) \leq t\nu_i(1 + \omega^{-1}(\omega_*(\lambda_{F_i}(x_i^*)) / \nu_i))$. Summing up the last inequalities from $i = 1$ to M we obtain the right-hand side of (3.5). \square

A.2. The proof of Lemma 3.2. *Proof.* The first inequality in (3.7) was proved in Lemma 3.1. We now prove the second one. Let us denote by $x_i^\tau(y) := x_i^c + \tau(x_{0i}^*(y) - x_i^c)$, where $\tau \in [0, 1]$ and $d_i^c(y) := \phi_i(x_i^c) + y^T(A_i x_i^c - b_i)$. Since F_i is ν_i -self-concordant, it follows from [15, inequality (2.3.3)] that:

$$F_i(x_i^\tau(y)) \leq F_i(x_i^c) - \nu_i \ln(1 - \tau), \quad \tau \in [0, 1].$$

Combining this inequality and the concavity of ϕ_i and then using the definitions of d_i^c and $d_i(\cdot)$ we have:

$$\begin{aligned}
d_i(y, t) &\geq \max_{\tau \in [0, 1]} \{ \phi_i(x_i^\tau(y)) + y^T A_i(x_i^\tau(y) - b_i) - t[F_i(x_i^\tau(y)) - F_i(x_i^c)] \} \\
&\geq \max \left\{ (1 - \tau)[\phi_i(x_i^c) + y^T(A_i x_i^c - b_i)] + \tau[\phi_i(x_{0i}^*(y)) + y^T(A_i x_{0i}^*(y) - b_i)] \right. \\
\text{(A.5)} \quad &\quad \left. + t\nu_i \ln(1 - \tau) \mid \tau \in [0, 1] \right\} \\
&= \max_{\tau \in [0, 1]} \{ (1 - \tau)d_i^c(y) + \tau d_{0i}(y) + t\nu_i \ln(1 - \tau) \}.
\end{aligned}$$

Now, we maximize the function $\xi(\tau) := (1 - \tau)d_i^c(y) + \tau d_{0i}(y) + t\nu_i \ln(1 - \tau)$ in last line of (A.5) w.r.t. $\tau \in [0, 1]$ to obtain $\tau^* = \left[1 - \frac{t\nu_i}{d_{0i}(y) - d_i^c(y)} \right]_+$, where $[a]_+ := \max\{0, a\}$. Therefore, if $d_{0i}(y) - d_i^c(y) \leq t\nu_i$, i.e. $\tau^* = 0$, then $d_{0i}(y) - d_i^c(y) \leq t\nu_i$. Otherwise, by substituting τ^* into the last line of (A.5), we obtain:

$$\text{(A.6)} \quad d_{0i}(y) \leq d_i(y, t) + t\nu_i \left(1 + \left[\ln((t\nu_i)^{-1}(d_{0i}(y) - d_i^c(y))) \right]_+ \right).$$

Furthermore, we note that $d_{0i}(y) - d_i^c(y) = \max_{x_i \in X_i} \{ \phi_i(x_i) + y^T(A_i x_i - b_i) \} - [\phi_i(x_i^c) + y^T(A_i x_i^c - b_i)] \geq 0$ for all $y \in Y$ and:

$$\begin{aligned}
d_{0i}(y) - d_i^c(y) &\stackrel{\phi \text{ is concave}}{\leq} \max_{x_i \in X_i} \left\{ \max_{\xi_i \in \partial \phi_i(x_i^c)} \left\{ [\xi_i + A_i^T y]^T (x_i - x_i^c) \right\} \right\} \\
\text{(A.7)} \quad &\leq \max_{x_i \in X_i} \left\{ \max_{\xi_i \in \partial \phi_i(x_i^c)} \left\{ \|\xi_i + A_i^T y\|_{x_i^c}^* \|x_i - x_i^c\|_{x_i^c} \right\} \right\} \\
&\stackrel{(3.1)}{\leq} (\nu_i + 2\sqrt{\nu_i}) \max_{\xi_i \in \partial \phi_i(x_i^c)} \left\{ \|\xi_i + A_i^T y\|_{x_i^c}^* \right\} \\
&\leq K_i < +\infty, \quad \forall y \in Y.
\end{aligned}$$

Summing up the inequalities (A.6) for $i = 1, \dots, M$ and then using (A.7) we get (3.7).

Finally, for fixed $\kappa \in (0, 1)$, since $\ln(x^{-1}) \leq x^{-\kappa}$ for $0 < x \leq \kappa^{1/\kappa}$, we have:

$$\nu_i t \left(1 + \left[\ln \frac{K_i}{\nu_i t} \right]_+ \right) \leq \nu_i t \left[1 + \left(\frac{K_i}{\nu_i t} \right)^\kappa \right] \leq [\nu_i + K_i^\kappa \nu_i^{1-\kappa}] t^{1-\kappa}, \quad \forall t \leq \frac{K_i}{\nu_i} \kappa^{1/\kappa}.$$

Consequently, if $t \leq \min \left\{ \frac{K_i}{\nu_i} \kappa^{1/\kappa}, \left(\frac{\varepsilon}{\sum_{i=1}^M [\nu_i + \nu_i^{1-\kappa} K_i^\kappa]} \right)^{1/(1-\kappa)}, i = 1, \dots, M \right\}$ then $D_X(t) \leq \varepsilon$, where $D_X(t)$ is defined in Lemma 3.2. Combining this condition and (3.7) we get the last conclusion of Lemma 3.2. \square

A.3. The proof of Lemma 4.3. First, we prove the following lemma which will be used to prove the main inequality in Lemma 4.3.

LEMMA A.1. *Suppose that Assumptions A.1-A.3 are satisfied. Then:*

a) $\nabla^2 \tilde{d}$ and $\nabla^2 \tilde{d}_{\bar{\delta}}$ defined by (3.10) and (4.3), respectively, guarantee:

$$\text{(A.8)} \quad (1 - \delta_+)^2 \nabla^2 \tilde{d}(y_+, t_+) \preceq \nabla^2 \tilde{d}_{\bar{\delta}}(y_+, t_+) \preceq (1 - \delta_+)^{-2} \nabla^2 \tilde{d}(y_+, t_+),$$

where $\delta_+ < 1$ defined by (4.6).

b) Moreover, one has:

$$\text{(A.9)} \quad \|\nabla \tilde{d}_{\bar{\delta}}(y, t) - \nabla \tilde{d}(y, t)\|_y^* \leq \|\bar{x}_{\bar{\delta}} - x^*\|_{x^*}.$$

c) If $\Delta < 1$ then $\bar{\lambda}_1 \leq \frac{\Delta + \bar{\lambda}}{1 - \Delta}$.

Proof. Since F is standard self-concordant, for any $x \in \text{dom}(F)$ and z such that $\|z - x\|_x < 1$, it follows from [16, Theorem 4.1.6] that:

$$(A.10) \quad (1 - \|z - x\|_x)^2 \nabla^2 F(x) \preceq \nabla^2 F(z) \preceq (1 - \|z - x\|_x)^{-2} \nabla^2 F(x).$$

Since $\nabla^2 F(x)$ is symmetric positive definite, by applying [1, Proposition 8.6.6] to two matrices $(1 - \|z - x\|_x)^{-2} \nabla^2 F(x)$ and $\nabla^2 F(z)$, and then to two matrices $(1 - \|z - x\|_x)^2 \nabla^2 F(x)$ and $\nabla^2 F(z)$ we obtain:

$$(A.11) \quad (1 - \|z - x\|_x)^2 A \nabla^2 F(x)^{-1} A^T \preceq A \nabla^2 F(z)^{-1} A^T \preceq (1 - \|z - x\|_x)^{-2} A \nabla^2 F(x)^{-1} A^T.$$

Using again [1, Proposition 8.6.6] for (A.11) we get:

$$(A.12) \quad \begin{aligned} & (1 - \|z - x\|_x)^2 A^T [A \nabla^2 F(x)^{-1} A^T]^{-1} A \preceq A^T [A \nabla^2 F(z)^{-1} A^T]^{-1} A \\ & \preceq (1 - \|z - x\|_x)^{-2} A^T [A \nabla^2 F(x)^{-1} A^T]^{-1} A. \end{aligned}$$

Now, using (3.10) and (3.11), we have $\nabla^2 \tilde{d}(y, t) = t^{-2} A \nabla^2 F(x^*)^{-1} A^T$. Alternatively, using (4.3) and (4.4), we get $\nabla^2 \tilde{d}_{\bar{\delta}}(y, t) = t^{-2} A \nabla^2 F(\bar{x}_{\bar{\delta}})^{-1} A^T$. Substituting these relations with $x = x_+^*$ and $z = \bar{x}_{\bar{\delta}+}$ into (A.11) and noting that $\delta_+ = \delta(\bar{x}_+, x_+^*)$ defined by (4.6), we obtain (A.8).

Next, we prove b). For any $x \in \text{dom}(F)$, we have $\nabla^2 F(x) \succ 0$. It is not difficult to show that the matrix $M(x) := \begin{bmatrix} \nabla^2 F(x) & A^T \\ A & A \nabla^2 F(x)^{-1} A^T \end{bmatrix}$ is symmetric positive semidefinite. Since A has full-row rank, $A \nabla^2 F(x)^{-1} A^T \succ 0$. By applying Schur's complement to $M(x)$ [1], we obtain:

$$(A.13) \quad A^T [A \nabla^2 F(x)^{-1} A^T]^{-1} A \preceq \nabla^2 F(x).$$

To prove (A.9) we note that $\nabla d_{\bar{\delta}}(y, t) - \nabla d(y, t) = A(\bar{x}_{\bar{\delta}} - x^*)$. Thus $\nabla \tilde{d}_{\bar{\delta}}(y, t) - \nabla \tilde{d}(y, t) = \frac{1}{t} A(\bar{x}_{\bar{\delta}} - x^*)$. This implies:

$$\begin{aligned} \left[\|\nabla \tilde{d}_{\bar{\delta}}(y, t) - \nabla \tilde{d}(y, t)\|_y^* \right]^2 &= t^{-2} (\bar{x}_{\bar{\delta}} - x^*)^T A^T \nabla^2 \tilde{d}(y, t)^{-1} A (\bar{x}_{\bar{\delta}} - x^*) \\ &\stackrel{(3.10), (3.11)}{=} (\bar{x}_{\bar{\delta}} - x^*)^T A^T [A \nabla^2 F(x^*)^{-1} A^T]^{-1} A (\bar{x}_{\bar{\delta}} - x^*) \\ &\stackrel{(A.13)}{\leq} (\bar{x}_{\bar{\delta}} - x^*)^T \nabla^2 F(x^*) (\bar{x}_{\bar{\delta}} - x^*) \\ &= \|\bar{x}_{\bar{\delta}} - x^*\|_{x^*}^2, \end{aligned}$$

which is equivalent to (A.9).

Finally, we prove c). By using the definitions of $\nabla \tilde{d}_{\bar{\delta}}(\cdot, t_+)$ and $\nabla^2 \tilde{d}_{\bar{\delta}}(\cdot, t_+)$ in (4.3), of $\tilde{d}_{\bar{\delta}}(\cdot, t_+)$ in (4.4), for any feasible point \hat{x} of (1.1), it follows from the definition of $\bar{\lambda}_1$ in (4.5) and $A\hat{x} = b$ that:

$$(A.14) \quad \begin{aligned} \bar{\lambda}_1^2 &= \left[\|\nabla \tilde{d}_{\bar{\delta}}(y, t_+)\|_y^* \right]^2 \stackrel{(4.5)}{=} \nabla \tilde{d}_{\bar{\delta}}(y, t_+) \nabla^2 \tilde{d}_{\bar{\delta}}(y, t_+)^{-1} \nabla \tilde{d}_{\bar{\delta}}(y, t_+) \\ &\stackrel{(4.4)}{=} t_+^{-1} \nabla d_{\bar{\delta}}(y, t_+) \nabla^2 d_{\bar{\delta}}(y, t_+)^{-1} \nabla d_{\bar{\delta}}(y, t_+) \\ &\stackrel{(4.3)}{=} (\bar{x}_{\bar{\delta}1} - \hat{x})^T A^T [A \nabla^2 F(\bar{x}_{\bar{\delta}1})^{-1} A^T]^{-1} A (\bar{x}_{\bar{\delta}1} - \hat{x}). \end{aligned}$$

Since $\Delta = \|\bar{x}_{\bar{\delta}1} - \bar{x}_{\bar{\delta}}\|_{\bar{x}_{\bar{\delta}}} < 1$ by assumption, we can apply the right-hand side of (A.12) with $x = \bar{x}_{\bar{\delta}}$ and $z = \bar{x}_{\bar{\delta}1}$ to obtain:

$$(A.15) \quad \bar{\lambda}_1^2 \leq (1 - \Delta)^{-2} (\bar{x}_{\bar{\delta}1} - \hat{x})^T A^T [A \nabla^2 F(\bar{x}_{\bar{\delta}})^{-1} A^T]^{-1} A (\bar{x}_{\bar{\delta}1} - \hat{x}).$$

Now, for any symmetric positive semidefinite matrix Q in $\mathbb{R}^{n \times n}$ and $u, v \in \mathbb{R}^n$, one can easily show that:

$$(A.16) \quad (u+v)^T Q (u+v) \leq [(u^T Q u)^{1/2} + (v^T Q v)^{1/2}]^2.$$

Since $H_{\bar{\delta}} := A^T [A \nabla^2 F(\bar{x}_{\bar{\delta}})^{-1} A^T]^{-1} A \succeq 0$, by applying (A.16) with $Q = H_{\bar{\delta}}$, $u = \bar{x}_{\bar{\delta}1} - \bar{x}_{\bar{\delta}}$ and $v = \bar{x}_{\bar{\delta}} - \hat{x}$, we have:

$$(A.17) \quad \bar{\lambda}_1^2 \leq (1-\Delta)^{-2} \left\{ [(\bar{x}_{\bar{\delta}1} - \bar{x}_{\bar{\delta}})^T H_{\bar{\delta}} (\bar{x}_{\bar{\delta}1} - \bar{x}_{\bar{\delta}})]_{[1]}^{1/2} + [(\bar{x}_{\bar{\delta}} - \hat{x})^T H_{\bar{\delta}} (\bar{x}_{\bar{\delta}} - \hat{x})]_{[2]}^{1/2} \right\}^2.$$

Note that $H_{\bar{\delta}} \preceq \nabla^2 F(\bar{x}_{\bar{\delta}})$ due to (A.13). The first term $[\cdot]_{[1]}$ in (A.17) satisfies:

$$(A.18) \quad [\cdot]_{[1]} \leq (\bar{x}_{\bar{\delta}+} - \bar{x}_{\bar{\delta}})^T \nabla^2 F(\bar{x}_{\bar{\delta}}) (\bar{x}_{\bar{\delta}1} - \bar{x}_{\bar{\delta}}) = \Delta^2.$$

On the other hand, by substituting $\bar{x}_{\bar{\delta}1}$ by $\bar{x}_{\bar{\delta}}$ into (A.14), we get:

$$(A.19) \quad \bar{\lambda}^2 = (\bar{x}_{\bar{\delta}} - \hat{x})^T A^T [A \nabla^2 F(\bar{x}_{\bar{\delta}})^{-1} A^T]^{-1} A (\bar{x}_{\bar{\delta}} - \hat{x}) = (\bar{x}_{\bar{\delta}} - \hat{x})^T H_{\bar{\delta}} (\bar{x}_{\bar{\delta}} - \hat{x}).$$

Combining (A.17), (A.18) and (A.19), we obtain $\bar{\lambda}_1 \leq \frac{(\Delta + \bar{\lambda})}{(1-\Delta)}$ which is indeed c). \square

The proof of Lemma 4.3. Since $\delta_1 + 2\Delta + \bar{\lambda} < 1$, it implies that $\delta_1 < 1$, $\Delta < 1/2$ and $\bar{\lambda} < 1$. The proof of Lemma 4.3 is divided into several steps as follows.

Step 1. First, let $p := y_+ - y$, we prove the following inequality:

$$(A.20) \quad \bar{\lambda}_+ \leq (1 - \delta_+)^{-1} \left\{ \delta_+ + (1 - \|p\|_y)^{-1} \left[\delta_1 + \frac{(2\delta_1 - \delta_1^2)}{(1 - \delta_1)^2} \|p\|_y + \frac{\|p\|_y^2}{1 - \|p\|_y} \right] \right\}.$$

Indeed, it follows from (A.8) that:

$$(A.21) \quad \begin{aligned} \bar{\lambda}_+ &= \|\nabla \tilde{d}_{\bar{\delta}}(y_+, t_+)\|_{y_+}^* = [\nabla \tilde{d}_{\bar{\delta}}(y_+, t_+) \nabla^2 \tilde{d}_{\bar{\delta}}(y_+, t_+)^{-1} \nabla \tilde{d}_{\bar{\delta}}(y_+, t_+)]^{1/2} \\ &\stackrel{(A.8)}{\leq} (1 - \delta_+)^{-1} [\nabla \tilde{d}_{\bar{\delta}}(y_+, t_+) \nabla^2 \tilde{d}(y_+, t_+)^{-1} \nabla \tilde{d}_{\bar{\delta}}(y_+, t_+)]^{1/2} \\ &\leq (1 - \delta_+)^{-1} \|\nabla \tilde{d}_{\bar{\delta}}(y_+, t_+)\|_{y_+}^*. \end{aligned}$$

Furthermore, by using (A.9) we have:

$$(A.22) \quad \begin{aligned} \|\nabla \tilde{d}_{\bar{\delta}}(y_+, t_+)\|_{y_+}^* &\leq \|\nabla \tilde{d}(y_+, t_+)\|_{y_+}^* + \|\nabla \tilde{d}_{\bar{\delta}}(y_+, t_+) - \nabla \tilde{d}(y_+, t_+)\|_{y_+}^* \\ &\stackrel{(A.9)}{\leq} \|\nabla \tilde{d}(y_+, t_+)\|_{y_+}^* + \delta_+. \end{aligned}$$

Since $\tilde{d}(\cdot, t_+)$ is standard self-concordant due to Lemma 3.3, one has:

$$(A.23) \quad \|\nabla \tilde{d}(y_+, t_+)\|_{y_+}^* \leq (1 - \|y_+ - y\|_y)^{-1} \|\nabla \tilde{d}(y_+, t_+)\|_y^* = (1 - \|p\|_y)^{-1} \|\nabla \tilde{d}(y_+, t_+)\|_y^*.$$

Plugging (A.23) and (A.22) into (A.21) we obtain:

$$(A.24) \quad \bar{\lambda}_+ \leq (1 - \delta_+)^{-1} \left[(1 - \|p\|_y)^{-1} \|\nabla \tilde{d}(y_+, t_+)\|_y^* + \delta_+ \right].$$

On the other hand, from (4.12), we have:

$$(A.25) \quad \begin{aligned} \nabla \tilde{d}(y_+, t_+) &\stackrel{(4.12)}{=} \nabla \tilde{d}(y_+, t_+) - [\nabla \tilde{d}_{\bar{\delta}}(y, t_+) + \nabla^2 \tilde{d}_{\bar{\delta}}(y, t_+) (y_+ - y)] \\ &= [\nabla \tilde{d}(y, t_+) - \nabla \tilde{d}_{\bar{\delta}}(y, t_+)]_{[1]} + \{[\nabla^2 \tilde{d}(y, t_+) - \nabla^2 \tilde{d}_{\bar{\delta}}(y, t_+)](y_+ - y)\}_{[2]} \\ &\quad + [\nabla \tilde{d}(y_+, t_+) - \nabla \tilde{d}(y, t_+) - \nabla^2 \tilde{d}(y, t_+) (y_+ - y)]_{[3]}. \end{aligned}$$

By substituting t by t_+ in (A.9), we obtain an estimate for $[\cdot]_{[1]}$ of (A.25) as:

$$(A.26) \quad \|\nabla \tilde{d}(y, t_+) - \nabla \tilde{d}_{\bar{\delta}}(y, t_+)\|_y^* \leq \|\bar{x}_{\bar{\delta}1} - x_1^*\|_{x_1^*} = \delta_1.$$

Next, we consider the second term $[\cdot]_{[2]}$ of (A.25). It follows from (A.8) that:

$$(A.27) \quad [(1-\delta_1)^2-1] \nabla^2 \tilde{d}(y, t_+) \preceq \nabla^2 \tilde{d}_{\bar{\delta}}(y, t_+) - \nabla^2 \tilde{d}(y, t_+) \preceq [(1-\delta_1)^{-2}-1] \nabla^2 \tilde{d}(y, t_+).$$

If we define $G := [\nabla^2 \tilde{d}_{\bar{\delta}}(y, t_+) - \nabla^2 \tilde{d}(y, t_+)]$ and $H := \nabla^2 \tilde{d}(y, t_+)^{-1/2} G \nabla^2 \tilde{d}(y, t_+)^{-1/2}$ then:

$$(A.28) \quad \|[\nabla^2 \tilde{d}(y, t) - \nabla^2 \tilde{d}_{\bar{\delta}}(y, t_+)](y_+ - y)\|_y^* = \|Gp\|_y^* \leq \|H\| \|p\|_y.$$

By virtue of (A.27) and the condition $\delta_1 < 1$, one has:

$$\|H\| \leq \max \{1 - (1 - \delta_1)^2, (1 - \delta_1)^{-2} - 1\} = (1 - \delta_1)^{-2} (2\delta_1 - \delta_1^2).$$

Hence, (A.28) leads to:

$$(A.29) \quad \|[\nabla^2 \tilde{d}(y, t) - \nabla^2 \tilde{d}_{\bar{\delta}}(y, t_+)](y_+ - y)\|_y^* \leq (1 - \delta_1)^{-2} (2\delta_1 - \delta_1^2) \|p\|_y.$$

Furthermore, since $\tilde{d}(\cdot, t)$ is standard self-concordant, similar to the proof of [16, Theorem 4.1.14], the third term $[\cdot]_{[3]}$ of (A.25) is estimated as:

$$(A.30) \quad \|\nabla \tilde{d}(y_+, t_+) - \nabla \tilde{d}(y, t_+) - \nabla^2 \tilde{d}(y, t_+)(y_+ - y)\|_y^* \leq (1 - \|p\|_y)^{-1} \|p\|_y^2.$$

Now, we apply the triangle inequality $\|a + b + c\|_y^* \leq \|a\|_y^* + \|b\|_y^* + \|c\|_y^*$ to (A.25) and then plugging (A.26), (A.29) and (A.30) into the resulting inequality to obtain:

$$\|\nabla \tilde{d}_{\bar{\delta}}(y_+, t_+)\|_y^* \leq \delta_1 + (1 - \delta_1)^{-2} (2\delta_1 - \delta_1^2) \|p\|_y + (1 - \|p\|_y)^{-1} \|p\|_y^2.$$

Finally, by substituting the last inequality into (A.24) we get (A.20).

Step 2. Next, we estimate (A.20) in terms of $\bar{\lambda}_1$ to obtain:

$$(A.31) \quad \bar{\lambda}_+ \leq (1 - \delta_+)^{-1} \left[\left(\frac{\bar{\lambda}_1}{1 - \delta_1 - \bar{\lambda}_1} \right)^2 + \frac{(2\delta_1 - \delta_1^2)}{(1 - \delta_1)^2} \left(\frac{\bar{\lambda}_1}{1 - \delta_1 - \bar{\lambda}_1} \right) + \frac{(1 - \delta_1)\delta_1}{1 - \delta_1 - \bar{\lambda}_1} + \delta_+ \right].$$

Indeed, by using (A.11) with $x = \bar{x}_{\bar{\delta}1}$ and $z = x_1^*$ and then (3.10) we have:

$$(1 - \delta_1)^2 \nabla^2 \tilde{d}_{\bar{\delta}}(y, t_+) \preceq \nabla^2 \tilde{d}(y, t_+) \preceq (1 - \delta_1)^{-2} \nabla^2 \tilde{d}_{\bar{\delta}}(y, t_+).$$

These inequalities together with the definition of $\|\cdot\|_y$ imply:

$$(1 - \delta_1) \|p\|_y \leq \|p\|_y = [p^T \nabla^2 \tilde{d}(y, t_+) p]^{1/2} \leq (1 - \delta_1)^{-1} \|p\|_y.$$

Moreover, since $\|p\|_y = \|\nabla \tilde{d}_{\bar{\delta}}(y, t_+)\|_y^* = \bar{\lambda}_1$ due to (4.12), the last inequality is equivalent to:

$$(A.32) \quad \|p\|_y \leq (1 - \delta_1)^{-1} \bar{\lambda}_1.$$

Note that the right-hand side of (A.20) is nondecreasing w.r.t. $\|p\|_y$ in $[0, 1)$. Substituting (A.32) into (A.20) we finally obtain (A.31).

Step 3. We further estimate (A.31) in terms of Δ and $\bar{\lambda}$. First, we can easily check that the right-hand side of (A.31) is nondecreasing w.r.t. $\bar{\lambda}_1$, δ_1 and δ_+ . Now, by using the definitions of Δ and $\bar{\lambda}$, it follows from Lemma A.1 c) that:

$$\bar{\lambda}_1 \leq (1 - \Delta)^{-1}(\bar{\lambda} + \Delta).$$

Since $\delta_+ < 1$ and $\delta_1 + 2\Delta + \bar{\lambda} < 1$, substituting this inequality into (A.31), we obtain

$$(A.33) \quad \bar{\lambda}_+ \leq (1 - \delta_+)^{-1} \left[\delta_+ + \left(\frac{\bar{\lambda} + \Delta}{1 - \delta_1 - 2\Delta - \bar{\lambda}} \right)^2 + \frac{(2\delta_1 - \delta_1^2)}{(1 - \delta_1)^2} \left(\frac{\bar{\lambda} + \Delta}{1 - \delta_1 - 2\Delta - \bar{\lambda}} \right) + \frac{\delta_1(1 - \delta_1)(1 - \Delta)}{1 - \delta_1 - 2\Delta - \bar{\lambda}} \right].$$

The right-hand side of (A.33) is well-defined and nondecreasing w.r.t. all variables.

Step 4. Finally, we facilitate the right-hand side of (A.33) to obtain (4.15). Since $\bar{\lambda} \geq 0$, we have:

$$\begin{aligned} (1 - \delta_1)(1 - \Delta) &= [1 - \delta_1 - 2\Delta - \bar{\lambda}] + (\bar{\lambda} + \Delta) + \delta_1\Delta \\ &\leq [1 - \delta_1 - 2\Delta - \bar{\lambda}] + (1 + \delta_1)(\bar{\lambda} + \Delta). \end{aligned}$$

The last inequality implies:

$$(A.34) \quad \frac{\delta_1(1 - \delta_1)(1 - \Delta)}{1 - \delta_1 - 2\Delta - \bar{\lambda}} \leq \delta_1 + \delta_1(1 + \delta_1) \left(\frac{\Delta + \bar{\lambda}}{1 - \delta_1 - 2\Delta - \bar{\lambda}} \right).$$

Alternatively, since $0 \leq \delta_1 < 1$, we have $1 + \delta_1 \leq \frac{1}{1 - \delta_1}$. Thus:

$$\begin{aligned} (1 - \delta_1)^{-2}(2\delta_1 - \delta_1^2) + \delta_1(1 + \delta_1) &= \delta_1 [(1 - \delta_1)^{-2} + (1 - \delta_1)^{-1} + (1 + \delta_1)] \\ &\leq \delta_1 [(1 - \delta_1)^{-2} + 2(1 - \delta_1)^{-1}]. \end{aligned}$$

Substituting inequality (A.34) into (A.33) and then using the last inequality and $\xi := \frac{\bar{\lambda} + \Delta}{1 - \delta_1 - 2\Delta - \bar{\lambda}}$, we obtain (4.15).

Step 5. The nondecrease of the right-hand side of (4.15) is obvious. The inequality (4.16) follows directly from (4.15) by noting that $\bar{\lambda} \equiv \lambda$ and $\bar{x}_{\bar{\delta}} \equiv x^*$. \square

A.4. The proof of Theorem 4.4. *Proof.* Let us define $\bar{\xi} := \frac{\Delta + \beta}{1 - \bar{\delta} - \beta - 2\Delta}$ and:

$$\varphi(\beta, \bar{\delta}, \Delta) := (1 - \bar{\delta})^{-1} \{ 2\bar{\delta} + \bar{\xi}^2 + \bar{\delta}[(1 - \bar{\delta})^{-2} + 2(1 - \bar{\delta})^{-1}] \bar{\xi} \},$$

By assumption that $\bar{\lambda} \leq \beta$, it follows from Lemma 4.3 that if $\varphi(\beta, \bar{\delta}, \Delta) \leq \beta$ then $\bar{\lambda}_+ \leq \beta$. This condition holds if a) $0 \leq \bar{\xi} \leq (\sqrt{p^2 + 4q} - p)/2 \equiv \theta$ and b) $0 \leq \bar{\delta} \leq \beta/(\beta + 2)$, where p and q are defined by (4.18). The condition a) is equivalent to $(1 + 2\theta)\Delta \leq \theta(1 - \bar{\delta} - \beta) - \beta$. Because $\Delta > 0$, we need $\theta > (1 - \bar{\delta} - \beta)^{-1}\beta$. This is guaranteed if $\mathcal{P}_{\bar{\delta}}(\beta) > 0$, where $\mathcal{P}_{\bar{\delta}}$ is defined in (4.17). By a well-known characteristic of a cubic polynomial, $\mathcal{P}_{\bar{\delta}}(\beta)$ has three real roots if $18c_0c_1c_2c_3 - 4c_2^3c_0 + c_2^2c_1^2 - 4c_3c_1^3 - 27c_2^2c_0^2 \geq 0$. By numerically checking the last condition, we can show that if $0 \leq \bar{\delta} \leq \bar{\delta}_{\max} := 0.043286$ then the three roots satisfy $0 \leq \beta_* < \beta^* < \beta_3$ and $\mathcal{P}_{\bar{\delta}}(\beta) > 0$ for all $\beta \in (\beta_*, \beta^*)$. With such values of $\bar{\delta}$ and β we have $\theta > (1 - \bar{\delta} - \beta)^{-1}\beta$ and the condition b) is also satisfied. Eventually, if we define $\bar{\Delta} := \frac{\theta(1 - \bar{\delta} - \beta) - \beta}{1 + 2\theta} > 0$ and choose $\bar{\delta}$, β and Δ such that $0 \leq \bar{\delta} \leq \bar{\delta}_{\max}$, $\beta \in (\beta_*, \beta^*)$ and $0 \leq \Delta \leq \bar{\Delta}$ then $\bar{\lambda} \leq \beta$ implies $\bar{\lambda}_+ \leq \beta$. \square

REFERENCES

- [1] Bernstein, D.S.: *Matrix mathematics: Theory, facts and formulas with application to linear systems theory*. Princeton University Press, Princeton and Oxford, (2005).
- [2] Bertsekas, D.P., and Tsitsiklis, J.N.: *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, (1989).
- [3] Boyd, S. and Vandenberghe, L.: *Convex Optimization*. University Press, Cambridge, (2004).
- [4] Chen, G., and Teboulle, M.: A proximal-based decomposition method for convex minimization problems. *Math. Program.*, **64**, 81–101, (1994).
- [5] Connejo, A. J., Mínguez, R., Castillo, E. and García-Bertrand, R.: *Decomposition Techniques in Mathematical Programming: Engineering and Science Applications*. Springer-Verlag, (2006).
- [6] Dolan, E.D. and Moré, J.J.: Benchmarking optimization software with performance profiles. *Math. Program.*, **91**, 201–213, (2002).
- [7] Fukuda, M., Kojima, M. and Shida, M.: Lagrangian dual interior-point methods for semidefinite programs. *SIAM J. Optim.* **12**, 1007–1031, (2002).
- [8] Holmberg, K. and Kiwiel, K.C.: Mean value cross decomposition for nonlinear convex problem. *Optim. Methods and Softw.* **21**(3), 401–417, (2006).
- [9] Kojima, M., Megiddo, N. and Mizuno, S. et al: Horizontal and vertical decomposition in interior point methods for linear programs. Technical Report. Information Sciences, Tokyo Institute of Technology, (1993).
- [10] Komodakis, N., Paragios, N., and Tziritas, G.: MRF Energy Minimization & Beyond via Dual Decomposition. *IEEE Trans Pattern Anal. Mach. Intell.*, **33**(3), 531–552, (2011).
- [11] Mehrotra, S. and Ozevin, M. G.: Decomposition Based Interior Point Methods for Two-Stage Stochastic Convex Quadratic Programs with Recourse. *Operation Research*, **57**(4), 964–974, (2009).
- [12] Neveen, G., Jochen, K.: Faster and simpler algorithms for multi-commodity flow and other fractional packing problems. *SIAM J. Comput.* **37**(2), 630–652, (2007).
- [13] Necoara, I. and Suykens, J.A.K.: Applications of a smoothing technique to decomposition in convex optimization, *IEEE Trans. Automatic control*, **53**(11), 2674–2679, (2008).
- [14] Necoara, I. and J.A.K. Suykens, J.A.K.: Interior-point Lagrangian decomposition method for separable convex optimization. *J. Optim. Theory Appl.*, **143**, 567–588, (2009).
- [15] Nesterov, Y. and Nemerovskii, A.: *Interior point polynomial methods in convex programming: Theory and applications*. SIAM, Philadelphia, (1994).
- [16] Nesterov, Y.: *Introductory Lectures on Convex Optimization*. Kluwer, Boston, (2004).
- [17] Nesterov, Y.: Smooth minimization of nonsmooth functions. *Math. Program.*, **103**(1):127–152, (2005).
- [18] Renegar, J.: *A Mathematical View of Interior-Point Methods in Convex Programming*. Society for Industrial and Applied Mathematics, Philadelphia, (2001).
- [19] Ruszczyński, A.: On convergence of an augmented lagrangian decomposition method for sparse convex optimization. *Math. Oper. Res.*, **20**, 634–656, (1995).
- [20] Samar, S., Boyd, S., and Gorinevsky, D.: Distributed Estimation via Dual Decomposition. *Proceedings European Control Conference (ECC)*, 1511–1516, Kos, Greece, (2007).
- [21] Shida, M.: An interior-point smoothing technique for Lagrangian relaxation in large-scale convex programming. *Optimization*, **57**(1), 183–200, (2008).
- [22] Tran Dinh, Q., Savorgnan, C. and Diehl, M.: Combining Lagrangian Decomposition and Excessive Gap Smoothing Technique for Solving Large-Scale Separable Convex Optimization Problems, (2011) (under revision). <http://arxiv.org/abs/1105.5427>.
- [23] Wei, E, Ozdaglar, A. and Jadbabaie, A.: A Distributed Newton Method for Network Utility Maximization. LIDS report 2832, (2011).
- [24] Venkat, A., Hiskens, I., Rawlings, J., and Wright, S.: Distributed MPC strategies with application to power system automatic generation control. *IEEE Trans. Control Syst. Technol.* **16**(6), 1192–1206, (2008).
- [25] Xiao, L., Johansson, M. and Boyd, S.: Simultaneous routing and resource allocation via dual decomposition. *IEEE Trans. Commun.* **52**(7), 1136–1144, (2004).
- [26] Zhao, G.: Interior point methods with decomposition for solving large-scale linear programs. *J. Optim. Theory Appl.* **102**, 169–192, (1999).
- [27] Zhao, G.: A Log-barrier with Benders decomposition for solving two-stage stochastic programs. *Math. Program.* **90**, 507–536, (2001).
- [28] Zhao, G.: A Lagrangian dual method with self-concordant barriers for multistage stochastic convex programming. *Math. Program.* **102**, 1–24, (2005).