Optimized choice of parameters in interior-point methods for linear programming*

LUIZ-RAFAEL SANTOS†

Department of Mathematics, Federal University of Santa Catarina, 89065-300, Blumenau-SC, Brazil.

FERNANDO VILLAS-BÔAS‡, AURELIO R. L. OLIVEIRA§, AND CLOVIS PERIN¶

Department of Applied Mathematics, IMECC, University of Campinas, 13083-859, Campinas-SP, Brazil.

February 6, 2019

In this work, we propose a predictor-corrector interior point method for linear programming in a primal-dual context, where the next iterate is chosen by the minimization of a polynomial merit function of three variables: the first is the steplength, the second defines the central path and the third models the weight of a corrector direction. The merit function minimization is performed by restricting it to constraints defined by a neighborhood of the central path that allows wide steps. In this framework, we combine different directions, such as the predictor, the corrector and the centering directions, with the aim of producing a better one. The proposed method generalizes most of predictor-corrector interior point methods, depending on the choice of the variables described above. Convergence analysis of the method is carried out, considering an initial point that has a good practical performance, which results in Q-linear convergence of the iterates with polynomial complexity.

*This research was sponsored by Brazilian Agencies *Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP)* (grants 08/09685-8 and 10/06822-4) and *Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq)*.

†Email: l.r.santos@ufsc.br
‡Email: fernandovillasboas@gmail.com
§Email: aurelio@ime.unicamp.br
¶Email: clovisperin@gmail.com

Numerical experiments using the Netlib test set are made, which show that this approach is competitive when compared to well established solvers, such as PCx.

1 Introduction

Since the early nineties, when primal-dual interior point methods research began to spread in the optimization community and when Mehrotra's [10] predictor-corrector method arose, many researchers tried to extend or improve it. Some of the main issues nowadays in IPMs include how to combine different directions, such as predictor, corrector or any high order correctors, so that a better direction is generated at each step [5]. A direction can be seen as a solution of a Newton system, derived from the KKT conditions of a Linear Programming (LP) problem. The directions that are used in each iteration are computed through a solution of linear systems whose coefficient matrices are the same but the right-hand side vectors are not.

Gondzio [4], for instance, combines Mehrotra's predictor-corrector main idea with multiple centrality corrections at the same iterate, with the objective of increasing the steplength. Colombo and Gondzio [2] extend this work, choosing predictor and corrector directions in a different manner, using a weight for the corrector direction and defining, according to the problem, the number of corrections.

On the other hand, Jarre and Wechs [7] solve, at each iteration, a subproblem to decrease the LP residuals. Such subproblem is, *per se*, an LP with very small dimension when compared to the original problem and its variables are the weight that each direction component will have in the final direction. They solve the subproblem using the Simplex method. In addition, Mehrotra and Li [11] search for a direction that is a combination of predictor and corrector directions through a small LP. In their work, multiple corrector directions are found using information generated by a suitable Krylov subspace. In [1], Berti et al. project the search direction in order to extend the step size in a multiple centrality corrections framework.

It is well known that path following interior-point methods, such as the ones previously cited, generate a sequence of points inside a neighborhood of the central path, which prevents iterates to get close to the feasibility bound too soon. For this, it is mandatory to impose predefined conditions that guarantee that the next point is in this neighborhood, to ensure a good performance of the method.

In this work, we will tackle some of these issues and other correlated ones. The main goal of this paper is to build and implement an infeasible path-following interior-point method for LP, using real polynomials in variables (α, μ, ω), where α is the steplength, μ is the parameter that defines the central path and ω models the weight of the corrector direction; indeed, this is a predictor-corrector method.

We will consider these parameters as variables and we will postpone their choice, which will be made through an optimization subproblem. This subproblem consists of a minimization of a predictive merit function that is a polynomial of maximum degree two on the variables (α, μ, ω). We call the merit function *predictive* because we will build it in such a way that it is possible to preview a measure of both linear and complementarity residuals for the next iterate. We call it Optimized Choice of Parameters Method (OCPM). A similar approach was first proposed by Villas-Bôas and Perin [13], on an auto-dual context.

We derive a convergence analysis of OCPM, using the tools proposed by Zhang [17] and correlated works for interior-point methods that are of the type of Mehrotra's predictor-corrector. Among those tools, it is necessary to choose a suitable initial point as well as a measure of the distance from the optimal solution to this point. As pointed out by Wright [15], several theoretical analysis of interior-point methods use initial points that depend on the norm of an optimal solution. However, these methods cannot be implemented since such norm cannot be obtained until one reaches this optimal solution.

To overcome this issue, we established Assumption 1, which any initial point must satisfy in order to guarantee convergence for our algorithm. Such assumption takes into account a measure for the LP data size and a distance between the initial point and an optimal solution. However it is not necessary to know *a priori*, an optimal solution. Moreover, this Assumption has only theoretical purposes and we do not need to use it during implementation.

The OCPM implementation was tested using the NETLIB test set. In addition, we compare our numerical results with PCx [3], a mature implementation of Mehrotra's predictor-corrector method that uses Gondzio's corrections.

Throughout this paper, we will use Greek letters for scalars and Latin letters for vectors and matrices. In a vector, a superscript denotes iteration count and a subscript denotes indexes of vector component. For scalars, the superscript will denote a power, while subscript will denote the iteration count. We will use t as a local variable, which could be a vector or a scalar. According to context, if k is the present iteration, we may omit the iteration index of a vector or scalar t. In this case, we will use the caret symbol ˆ over a vector or scalar to represent its value in the next iteration, i.e., $\hat{t} = t^{k+1}$. Moreover, $\|\cdot\|$ will denote the 2-norm. We shall use the following notations: for vectors u, v, $U = \text{diag}(u)$, $\min(u) = \min_{i=1,\dots,n}\{u_i\}$, and $uv = UVe = Uv$, where e is vector whose components are all equal to one with the appropriate dimension. If $v \in \mathbb{R}^p$, we define the *average of* v as $\overline{v} := \dfrac{1}{p}\sum_{i=1}^{p} v_i$. The scalar $\bar{v}$ is the arithmetic mean of the components of a vector v.

This work is organized in the following way: in Section 2 we present the linear programming problem and we introduce the Optimized Choice of Parameters Method (OCPM); in Section 3 we establish our main convergence and complexity results by specifying a minimum steplength that guarantees OCPM convergence; in Section 4 we present the computational tests and we discuss their results; we conclude with final remarks.

2 Optimized Choice of Parameters Method framework

Let the primal-dual linear programming problem, in its standard form, be

$$\begin{array}{llcll} \min\limits_{x} & c^T x & & \max\limits_{(y,z)} & b^T y \\ \text{s.t.} & \begin{cases} Ax = b \\ x \geq 0 \end{cases} & \text{and} & \text{s.t.} & \begin{cases} A^T y + z = c \\ z \geq 0,\ y \text{ free} \end{cases} \end{array}, \tag{P-D}$$

where $A \in \mathbb{R}^{m \times n}$ is full rank and vectors x, y, z, b, c have appropriate dimension.

For convenience, we define the following notation. We say that $w := (x, y, z)$ is primal-dual feasible if $w \in \mathcal{F} := \{w \mid Ax = b,\ A^T y + z = c,\ (x, z) \geq 0\}$. We also define the set $\mathcal{Q}^+ := \{w \mid (x, z) > 0\}$. Due to those definitions, we call $\mathcal{F}^+ := \mathcal{F} \cap \mathcal{Q}^+$ the *feasible interior point set of* (P-D). A point w satisfies the Karush-Kuhn-Tucker (KKT) conditions for problem (P-D) if such point is in $\mathcal{F}_S := \{w \mid w \in \mathcal{F},\ xz = 0\}$. Since (P-D) is a convex problem, $\mathcal{F}_S$ is its set of optimal solutions.

As usual in IPMs, we assume that $\mathcal{F} \neq \varnothing$ — *i.e.*, a feasible point exists.

We begin the description of our method by showing its simplified scheme in Algorithm 1, which resumes the Optimized Choice of Parameters Method (OCPM) framework. In fact, each step of Algorithm 1 — except steps 3, 7 and 10 —, will be explained in this section, where we define OCPM. The initial point choice and the termination criteria will be explained in the next sections.

Algorithm 1 Optimized Choice of Parameters Method.

1: **procedure** OCPM(A, b, c)
2: $k \leftarrow 0$;
3: $w^0 = (x^0, y^0, z^0) \leftarrow$ InitialPoint(A, b, c); ▷ Ensure $(x^0, z^0) > 0$;
4: **repeat**
5: Compute $(\Delta w^{\text{af}})^k$, $(\Delta w^{\mu})^k$ and $(\Delta w^{\omega})^k$
6: Compute the coefficients of $\hat{\varphi}(\alpha, \mu, \omega)$, $h_C(\alpha, \mu, \omega)$ and $g_C^i(\alpha, \mu, \omega)$;
7: Solve the global optimization problem Eq. (19) and find $(\alpha_k, \mu_k, \omega_k)$.
8: Ensure $\alpha_k \in (0, 1)$ such that $(x^{k+1}, z^{k+1}) > 0$, using the ratio test Eq. (1) and do

$$w^{k+1} = w^k + \alpha_k((\Delta w^{\text{af}})^k + \mu_k(\Delta w^{\mu})^k + \omega_k(\Delta w^{\omega})^k);$$

9: $k \leftarrow k + 1$
10: **until** Termination criteria is achieved.
11: **end procedure**

Note that, as in usual IPM, we will perform the ratio test, as mentioned in step 8 above, to ensure that each (x^k, z^k) remains on $\mathcal{Q}^+$. For that, we will choose the step length α_k as

$$\alpha_k \leq \alpha_k^R := \frac{-1}{\min\left\{(X^k)^{-1}\Delta x^k, (Z^k)^{-1}\Delta z^k, -1\right\}}. \tag{1}$$

To guarantee the well-definedness of OCPM, some transformations are mandatory, and we proceed now with their explanations. These transformations will allow us to write each parameter (α, μ, ω) as free-variables, and will let us set a framework to choose them in an optimized manner.

For any given $w = (x, y, z)$, we can define the residual vectors r_P, r_D and r_C as

$$r_P := Ax - b, \quad r_D := A^T y + z - c, \quad r_C := xz. \tag{2}$$

Let (x^0, y^0, z^0) be any initial point of our method, whose components related to x and z are positive. Then the initial residual can be written as

$$r_P^0 = Ax^0 - b, \quad r_D^0 = A^T y^0 + z^0 - c, \quad r_C^0 = x^0 z^0 > 0.$$

In our method, we need to guarantee that the primal and dual residuals are non-negative in each step. The reason for that will be explained later. Choosing (x^0, y^0, z^0) such that the initial residual is non-negative can be a hard task. Thus, we will adopt the following strategy.

Let (x^0, y^0, z^0) be any initial point and S_P and S_D be diagonal matrices such that every component of the diagonal is chosen using the following rule: for $J = P, D$, let $(S_J)_{ii} = 1$, in case $(r_J^0)_i \geq 0$, and $(S_J)_{ii} = -1$, in case $(r_J^0)_i < 0$. Thus, we guarantee that $S_P(r_P^0) \geq 0$ and $S_D(r_D^0) \geq 0$.

Moreover,

$$\mathcal{S} := \{w \mid S_P(Ax - b) = 0,\ S_D(A^T y + z - c) = 0,\ xz = 0 \text{ and } (x, z) \geq 0\} \tag{3}$$

is equal to $\mathcal{F}_S$, as S_D and S_P are nonsingular matrices. Note that we are only multiplying each row of the original KKT system by 1 or -1. We will call the equations that define $\mathcal{S}$ as *the sign-adjusted KKT system*. In this view, we propose an homotopy IPM scheme to find, at each iteration k, an approximate solution of

$$\begin{cases} S_P(Ax - b) = 0, \\ S_D(A^T y + z - c) = 0, \\ xz = \mu e, \\ (x, z) > 0, \end{cases}$$

where $\mu > 0$. As k increases, if $\mu \to 0$, then this solution is an approximation of a solution of (P-D).

Such an approach is necessary to our method, since we have to guarantee that, at each iteration, the residuals of each KKT row do not change sign from the previous iteration. The reason for that will be explained in the following, particularly at Remark 1.

2.1 Search directions

Since our method is an IPM of the Mehrotra type, we will use the affine-scaling direction Δw^{af} in our final direction. Such direction is the solution of

$$\begin{cases} A\Delta x^{\text{af}} + r_P = 0, \\ A^T \Delta y^{\text{af}} + \Delta z^{\text{af}} + r_D = 0, \\ z\Delta x^{\text{af}} + x\Delta z^{\text{af}} + r_C = 0. \end{cases} \tag{4}$$

As usual in IPMs, to solve this linear system, we use the Cholesky factorization of ADA^T — where $D \coloneqq XZ^{-1}$ — and one backsolve.

2.1.1 An ideal direction

Let $\Delta w \coloneqq (\Delta x, \Delta y, \Delta z)$ be a single *ideal* (in a utopian sense) direction such that, for w and a parameter $\mu \geq 0$ given, $\hat{w} = w + \Delta w$ would be a solution of

$$\begin{cases} A\hat{x} - b = 0, \\ A^T\hat{y} + \hat{z} - c = 0, \\ \hat{x}\hat{z} = \mu e, \\ \hat{x}, \hat{z} \geq 0. \end{cases}$$

We will try to find the ideal direction Δw as a combination of Δw^{af}, the affine-scaling direction, and Δw^{c}, the *ideal* corrector direction, such that, $\Delta w = \Delta w^{\mathrm{af}} + \Delta w^{\mathrm{c}}$. Using the definition of $\hat{w}$, we obtain the nonlinear system

$$\begin{cases} A\Delta x^c = 0, \\ A^T\Delta y^c + \Delta z^c = 0, \\ x\Delta z^c + z\Delta x^c + \Delta x\Delta z = \mu e. \end{cases} \tag{5}$$

Vector $\Delta x\Delta z$ is a second order correction, used in several IPMs [4, 10]. In this work, we will somehow generalize second order corrections. To achieve this, we suppose that for a scalar $\omega \in [\omega_{\min}, \omega_{\max}]$ and a given scalar ε_ω we have $\left\|\Delta x\Delta z - \omega\Delta x^{\mathrm{af}}\Delta z^{\mathrm{af}}\right\| < \varepsilon_\omega$.

That leads us to use $\omega\Delta x^{\mathrm{af}}\Delta z^{\mathrm{af}}$ as a reasonable approximation of $\Delta x\Delta z$. In particular, note that in Mehrotra's work [10], $\omega = 1$. On the other hand, in Gondzio's[4], $\Delta x\Delta z$ is approximated — several times, according to some predetermined settings — by directions that are component-wise projections of the complementarity term in a symmetric neighborhood of the central path. If one chooses $\mu = 0$ and $\omega = 1$ and sets a feasible initial point, the method described by Monteiro, Adler and Resende arises [12].

In fact, the choice of μ and the manner the high order corrections is chosen and used determines the different IPMs that are used nowadays [5]. In this view, our method generalizes some of the most known and used IPMs because those methods use a particular fixed choice of μ or ω while OCPM treats those parameters as variables and optimizes such choices. The utilization of the optimization subproblem solution allows us to argue that at each iteration the choices of μ and ω are *better* than the fixed ones by our counterparts.

2.1.2 How to combine directions

It is common to use weights to combine different directions in order to obtain a better direction. Colombo and Gondzio [2] generalize and extend Gondzio's [4] work by allowing a combination of multiple directions with some weights which are chosen by a linear search. Jarre and Wechs [7] solve a linear subproblem whose solution determines the weights of the high order directions they use. Villas-Bôas and Perin [13] postpone the choice of the penalty barrier parameter and the steplength. They show that the next iterate can be expressed as a quadratic function of the barrier parameter, and they prove that the parameterization is useful to guarantee both the non negativity of the next iterate and the proximity of these iterates to the central path.

One of the main contributions of this work is to postpone the choice of μ — which is the parameter that determines the central path — and ω — which is the weight of the high order correction — by using an extension of Villas-Bôas and Perin [13] ideas. These two parameters and the steplength α will be seen here as *free variables*. To select them, we will use a merit function, whose use will be justified next.

First, if we use the approximation $\omega\Delta x^{\text{af}}\Delta z^{\text{af}}$, the nonlinear system Eq. (5) becomes the *linear system*

$$\begin{bmatrix} A & 0 & 0 \\ 0 & A^T & I \\ Z & 0 & X \end{bmatrix} \begin{bmatrix} \Delta x^{\text{c}} \\ \Delta y^{\text{c}} \\ \Delta z^{\text{c}} \end{bmatrix} = \mu \begin{bmatrix} 0 \\ 0 \\ e \end{bmatrix} + \omega \begin{bmatrix} 0 \\ 0 \\ -\Delta x^{\text{af}}\Delta z^{\text{af}} \end{bmatrix}. \tag{6}$$

Let $J(w)$ be the left-hand side matrix of the above system — such matrix is the same one used in Eq. (4). For any $(\mu, \omega) \geq 0$ — yet to be chosen —, let Δw^{c} be split as

$$\Delta w^{\text{c}} = \mu \Delta w^{\mu} + \omega \Delta w^{\omega}. \tag{7}$$

In order to find Δw^{μ} and Δw^{ω}, we solve two linear systems: $J(w)\Delta w^{\mu} = (0, 0, e)$ and $J(w)\Delta w^{\omega} = (0, 0, -\Delta x^{\text{af}}\Delta z^{\text{af}})$, using the same Cholesky factorization of ADA^T and two backsolves.

The next iterate is given by

$$\begin{aligned} \hat{w} &= w + \alpha(\Delta w^{\text{af}} + \Delta w^{\text{c}}) \\ &= w + \alpha(\Delta w^{\text{af}} + \mu\Delta w^{\mu} + \omega\Delta w^{\omega}), \end{aligned} \tag{8}$$

and the computational effort, by iteration, to compute it is dominated by the computation of one Cholesky factorization followed by at most three backsolves.

Up to now, the tuple (α, μ, ω) is yet to be chosen. To take the step as written above, we will treat α, μ, ω algebraically as real free variables of a merit function and solve an optimization subproblem, at each iteration, that sets a value for each of these parameters. The next section describes such subproblem.

2.2 A polynomial optimization subproblem

In general, merit functions used in IPMs are considered as tools to measure how close the iterate is to an optimal solution. Nevertheless, it is not very common to use them as a termination criteria for a given algorithm. Zhang [17], for instance, defines a merit function such that the complementarity and both primal and dual infeasibilities are reduced. Nevertheless, such function is only used in a theoretical manner to prove the algorithm convergence and complexity. In addition, PCx [3], an implementation of

Mehrotra's IPM, uses a merit function to detect infeasibility and to verify if the solution is unknown or suboptimal.

In this work, we will define a merit function that serves both as a measure of complementarity and feasibility — quite like Zhang's [17] — as well as a guide to the choice of the next point. In fact, the merit function is a core part of our algorithm as a tool to select α, μ, ω. We will formulate our merit function using a prediction of the residuals of the KKT system Eq. (3) for the next iterate.

Definition 1. The *vector of residuals* of system Eq. (3), for a given point (x, y, z) is given by

$$\rho(x,y,z) := \begin{cases} \rho_P(x,y,z) = S_P(Ax-b) \\ \rho_D(x,y,z) = S_D(A^T y + z - c) \\ \rho_C(x,y,z) = xz \end{cases} .$$

The linear part of ρ is denoted by $\rho_L := (\rho_P, \rho_D) \in \mathbb{R}^{m+n}$.

Using Definition 1, given any point (x, y, z), we can predict the sign-adjusted residual for the next iterate depending on a choice of (α, μ, ω), that is $(\hat{x}, \hat{y}, \hat{z}) = \hat{w}(\alpha, \mu, \omega)$, using Theorem 1.

Theorem 1. *The residual of the next iterate for system* (3)*,* $\hat{\rho}$*, depends on the choice of* (α, μ, ω) *and can be written as*

$$\hat{\rho}(\alpha,\mu,\omega) = \begin{bmatrix} (1-\alpha)\rho_P \\ (1-\alpha)\rho_D \\ (1-\alpha)\rho_C + \alpha\mu e + \alpha(\alpha-\omega)L_{0,0} + \alpha^2\Lambda(\mu,\omega) \end{bmatrix}, \tag{9}$$

where

$$\Lambda(\mu,\omega) := \mu^2 L_{2,0} + \mu L_{1,0} + \mu\omega L_{1,1} + \omega^2 L_{0,2} + \omega L_{0,1}, \tag{10}$$

and

$$\begin{aligned} L_{0,0} &:= \Delta x^{af}\Delta z^{af}, & L_{1,1} &:= \Delta x^{\mu}\Delta z^{\omega} + \Delta x^{\omega}\Delta z^{\mu}, \\ L_{1,0} &:= \Delta x^{af}\Delta z^{\mu} + \Delta z^{af}\Delta x^{\mu}, & L_{0,1} &:= \Delta x^{af}\Delta z^{\omega} + \Delta z^{af}\Delta x^{\omega}, \\ L_{2,0} &:= \Delta x^{\mu}\Delta z^{\mu}, & L_{0,2} &:= \Delta x^{\omega}\Delta z^{\omega}. \end{aligned} \tag{11}$$

Proof. For the primal linear residual $\hat{\rho}_P$, using Eq. (8) in Eq. (3) we obtain

$$\hat{\rho}_P = S_P(A(x + \alpha\Delta x^{\text{af}}) - b) + \alpha S_P A \Delta x^{\text{c}}.$$

From Eq. (4) we derive $A\Delta x^{\text{af}} = b - Ax$ and thus the following holds:

$$S_P(A(x+\alpha\Delta x^{\text{af}}) - b) = S_P(\alpha-1)A\Delta x^{\text{af}} = (1-\alpha)S_P(Ax-b) = (1-\alpha)\rho_P.$$

If we use equation Eq. (5), we have $A\Delta x^{c} = 0$ so $\hat{\rho}_P = (1-\alpha)\rho_P$. The proof for the dual linear part of $\hat{\rho}$ is similar, we will omit it.

For the complementarity part of $\hat{\rho}$, $\hat{\rho}_C = \hat{x}\hat{z}$. Then, from Eq. (8), we have

$$\hat{x}\hat{z} = \alpha^2(\Delta x^{\text{af}} + \Delta x^{\text{c}})(\Delta z^{\text{af}} + \Delta z^{\text{c}}) + \alpha\left[(x\Delta z^{\text{af}} + z\Delta x^{\text{af}}) + (x\Delta z^{\text{c}} + z\Delta x^{\text{c}})\right] + xz.$$

Since $(x\Delta z^{\text{af}}+z\Delta x^{\text{af}}) = -xz$ — by Eq. (4) — equation Eq. (7) gives us $x\Delta z^{\mu}+z\Delta x^{\mu} = e$ and $x\Delta z^{\omega} + z\Delta x^{\omega} = -\Delta x^{\text{af}}\Delta z^{\text{af}}$. Then, from Eq. (6) follows $(x\Delta z^{c} + z\Delta x^{c}) = \mu e - \omega\Delta x^{\text{af}}\Delta z^{\text{af}}$. Thus, we have $\hat{x}\hat{z} = t_1 + t_2$, where

$$t_1 = (1-\alpha)xz+\alpha\mu e+\alpha(\alpha-\omega)\Delta x^{\text{af}}\Delta z^{\text{af}} \text{ and } t_2 = \alpha^2\left(\Delta x^{\text{af}}\Delta z^{c} + \Delta z^{\text{af}}\Delta x^{c} + \Delta x^{c}\Delta z^{c}\right)$$

are local variables.

Since $\Delta w^{\text{c}} = \mu\Delta w^{\mu} + \omega\Delta w^{\omega}$, then

$$\begin{aligned}
\Delta x^{\text{af}}\Delta z^{c} &= \mu\Delta x^{\text{af}}\Delta z^{\mu} + \omega\Delta x^{\text{af}}\Delta z^{\omega},\\
\Delta z^{\text{af}}\Delta x^{c} &= \mu\Delta z^{\text{af}}\Delta x^{\mu} + \omega\Delta z^{\text{af}}\Delta x^{\omega},\\
\Delta x^{c}\Delta z^{c} &= \mu^2\Delta x^{\mu}\Delta z^{\mu} + \mu\omega\left(\Delta x^{\mu}\Delta z^{\omega} + \Delta x^{\omega}\Delta z^{\mu}\right) + \omega^2\Delta x^{\omega}\Delta z^{\omega}.
\end{aligned}$$

This way, t_2 is now

$$\begin{aligned}
t_2 = \alpha^2 \Big(&\mu^2\Delta x^{\mu}\Delta z^{\mu} + \mu\omega\left(\Delta x^{\mu}\Delta z^{\omega} + \Delta x^{\omega}\Delta z^{\mu}\right) + \\
&+\mu\left(\Delta x^{\text{af}}\Delta z^{\mu} + \Delta z^{\text{af}}\Delta x^{\mu}\right) + \omega\left(\Delta x^{\text{af}}\Delta z^{\omega} + \Delta z^{\text{af}}\Delta x^{\omega}\right) + \omega^2\Delta x^{\omega}\Delta z^{\omega}\Big).
\end{aligned}$$

This expansion of t_2 gives us the aimed $\hat{x}\hat{z}$.

If we define vectors $\Lambda(\mu,\omega)$ as stated in Eq. (10) and $L_{i,j}$ as stated in Eq. (11) and performing the necessary substitutions, we can find the expression for $\hat{\rho}_C$. □ □

The vector $\hat{\rho}(\alpha,\mu,\omega) \in \mathbb{R}^{m+2n}$ has the same number of rows of Eq. (3). In addition, a consequence of Theorem 1 is that all residuals remain non-negative if the initial point is interior.

Corollary 1. *In each iteration k of OCPM, for $\alpha \in (0,1]$ and $(\mu,\omega) \geq 0$, if $\rho^0 \geq 0$, the inequality $\rho^k(\alpha,\mu,\omega) \geq 0$ holds.*

Proof. For the linear part of Eq. (3), by Theorem 1, it follows that $\hat{\rho}_L(\alpha,\mu,\omega) = (1-\alpha)\rho_L$. Since $\rho^0 \geq 0$ and $0 < \alpha \leq 1$, by induction, the linear part is nonnegative.

For the complementarity part, observe that step 8 of Algorithm 1 ensures

$$(x^{k+1}, z^{k+1}) > 0,$$

provided $(x^k, z^k) \in \mathcal{Q}^+$. Since $\rho^0 \geq 0$, we have $\rho_C^k > 0$ for all k, and the result holds. □ □

We can now define our merit function, using the results above.

Definition 2 (Merit Function)**.** The *merit function* for any given point (x,y,z) is

$$\varphi(x,y,z) := \frac{1}{m+n}\left\|\rho_L\right\|_1 + \frac{x^T z}{n}. \tag{12}$$

Remark 1. Although it is possible to estimate how far a point (x,y,z) is from an optimal solution in many ways, our choice — the merit function φ above — has some properties that we will exploit and explain in the following:

(i) The tuple of parameters (α, μ, ω) will be treated as variables (with constraints), thus we can postpone their choice until it is utterly needed, as in [13];

(ii) Since we are using matrices S_P and S_D and Corollary 1, we ensured that (x, y, z) computed by our method is in $\mathcal{Q}^+$ and the *sign* of each row of ρ_L remains the same in all iterations — in fact, we make sure that such rows are non-negative. In this case, we are sure that for all i $\rho_L(x, y, z)_i \geq 0$. Therefore, we can discard the norm-1 modulus in the definition of ρ_L, that is,

$$\frac{1}{m+n}\|\rho_L\|_1 = \frac{1}{m+n}\sum_{\ell=1}^{m+n}(\rho_L)_\ell.$$

(iii) Due to Definition 1, we can rewrite the complementarity gap average as $x^T z/n = 1/n\sum_{i=1}^n x_i z_i = 1/n\sum_{i=1}^n(\rho_C)_i$;

(iv) From Corollary 1, given $(x, y, z) \in \mathcal{Q}^+$, we have that $\varphi(x, y, z) \geq 0$;

(v) Finally, if (x^*, y^*, z^*) is an optimal solution of Eq. (3), then $\varphi(x^*, y^*, z^*) = 0$.

Using Items (ii) and (iii) above, we can rewrite Eq. (12) as

$$\varphi(x, y, z) = \frac{1}{m+n}\sum_{\ell=1}^{m+n}(\rho_L)_\ell + \frac{1}{n}\sum_{i=1}^{n}(\rho_C)_i.$$

Using the average of a vector notation, we can represent the merit function as $\varphi(x, y, z) = \overline{\rho_L} + \overline{\rho_C}$.

If we use Definition 2 together with Theorem 1, we can find a *prediction* of the merit function for the next iterate $(\hat{x}, \hat{y}, \hat{z})$. Such prediction will be denoted as $\hat{\varphi}$, and the next result gives us the algebraic expression of $\hat{\varphi}$, a polynomial on variables (α, μ, ω).

Theorem 2. *The predictive merit function for the point $(\hat{x}, \hat{y}, \hat{z})$ on the next iteration of Algorithm 1 can be written as the real polynomial $\hat{\varphi}$, which depends on the variables (α, μ, ω), with following expression*

$$\hat{\varphi}(\alpha, \mu, \omega) = (1-\alpha)(\overline{\rho_L} + \overline{\rho_C}) + \alpha\mu + \alpha(\alpha - \omega)\overline{L_{0,0}} + \alpha^2\overline{\Lambda(\mu, \omega)}. \tag{13}$$

Proof. Using Eq. (9), we know the expression of the next residual, depending on a choice of (α, μ, ω). In this case, the predictive merit function will be

$$\hat{\varphi}(\alpha, \mu, \omega) = \overline{\hat{\rho}_L}(\alpha, \mu, \omega) + \overline{\hat{\rho}_C}(\alpha, \mu, \omega). \tag{14}$$

If we apply Theorem 1 on Eq. (14) and if we use the average of a vector notation, we have Eq. (13). □ □

Remark 2. Since Δx^μ and Δz^μ are orthogonal, as well as Δx^ω and Δz^ω, one can see that $\overline{L_{2,0}} = \frac{(\Delta x^\mu)^T(\Delta z^\mu)}{n} = 0$ and $\overline{L_{0,2}} = \frac{(\Delta x^\omega)^T(\Delta z^\omega)}{n} = 0$.

We regard $\hat{\varphi}$ defined above as complying with all items listed on Remark 1. Therefore, we built a merit function that has good mathematical properties — a polynomial in variables (α, μ, ω) with degree up to two — and that would be a reliable measure of the iterate quality.

Our method relies on finding at each iteration the global minimizer $(\alpha^*, \mu^*, \omega^*)$ of $\hat{\varphi}(\alpha, \mu, \omega)$. Since $\hat{\varphi}$ can predict the arithmetic mean of the next residual, this optimal choice allows us to take a step in a direction that minimizes ρ, at least in average.

Many good IPMs use some neighborhood to ensure that the iterates are within a reasonable distance from the central path. We show now that if an iterate is on a feasible set generated by polynomial constraints on variables (α, μ, ω), built from the *wide* neighborhood $\mathcal{N}_{-\infty}$ [15], then this iterate is in this neighborhood as well.

Our objective here is to guarantee, using such polynomials as constraints in the global optimization of merit function $\hat{\varphi}$, that the next iterate not only reduces the value of the merit function, but also the next point has the good convergence properties of the $\mathcal{N}_{-\infty}$ neighborhood.

Let us now establish the relationship between the linear part of residual at iteration k and the same residuals at iteration $k+1$, given by Theorem 1.

Proposition 1. *Let* $\{(x^k, y^k, z^k)\}$ *be a sequence of iterates generated by Algorithm 1. Then for* $k \geq 0$*,*

$$r_P^{k+1} = (1-\alpha_k) r_P^k = \nu_{k+1} r_P^0, \quad r_D^{k+1} = (1-\alpha_k) r_D^k = \nu_{k+1} r_D^0,$$

where $\nu_0 \coloneqq 1$ *and* $\nu_{k+1} \coloneqq \prod_{j=0}^{k} (1-\alpha_j) \geq 0$*. Moreover* $\rho_L^{k+1} = \nu_{k+1} \rho_L^0$ *and* $\nu_k = \overline{\rho_L^k} / \overline{\rho_L^0}$*.*

Proof. We present here the proof for the primal part ρ_L^{k+1}. The others proofs are similar. By the definition of the residuals given in Eq. (2) and by Definition 1, we have $\rho_P^k = S_P r_P^k$ and $\rho_D^k = S_D r_D^k$ for all k. Since S_P and S_D are nonsingular and since $\rho_L^k = (\rho_P^k, \rho_D^k)$, the first line of Eq. (9) in Theorem 1, one can see that $\rho_P^k = (1-\alpha_{k-1}) \rho_P^{k-1}$ and then $S_P r_P^k = (1-\alpha_{k-1}) S_P r_P^{k-1}$. We conclude then that $S_P^{-1} S_P r_P^k = (1-\alpha_{k-1}) r_P^{k-1}$, that is, $r_P^k = (1-\alpha_{k-1}) r_P^{k-1}$.

Therefore, it follows by induction that $r_P^k = (1-\alpha_{k-1}) r_P^{k-1} = \nu_k r_P^0$. The proof for the dual part is similar, so it is omitted. Therefore,

$$\rho_L^k = (S_P r_P^k, S_D r_D^k) = \nu_k (S_P r_P^0, S_D r_D^0) = \nu_k (\rho_P^0, \rho_D^0) = \nu_k \rho_L^0.$$

Finally, using the average operator definition, we have $\nu_k = \overline{\rho_L^k} / \overline{\rho_L^0}$. □ □

Remark 3. If $\rho_L^0 = 0$, then the initial point is a feasible point and Algorithm 1 solves the LP like a feasible method does. In practice, however, to find a feasible initial point is unusual.

In our method, we use the wide neighborhood of the central path [8], denoted by $\mathcal{N}_{-\infty}(\gamma, \beta)$, to make use of the good properties that points that lie in it have. Since

$\rho_C = xz$ and $(\rho_C)_i = x_i z_i$ using our notation with Definition 1 and Proposition 1 for the sign-adjusted KKT system, it is possible to rewrite such neighborhood as

$$\mathcal{N}_{-\infty}(\gamma,\beta) \coloneqq \left\{(x,y,z) \in \mathcal{Q}^+ : \frac{\overline{\rho_L}}{\overline{\rho_L^0}} \leq \beta \frac{\overline{\rho_C}}{\overline{\rho_C^0}}, (\rho_C)_i \geq \gamma \overline{\rho_C}, \forall i = 1,\ldots,n\right\}, \tag{15}$$

where $\gamma \in (0,1)$ and $\beta \geq 1$. Using Proposition 1 we have $\nu_k \leq \beta \overline{\rho_C^k}/\overline{\rho_C^0}$, for any point $(x^k, y^k, z^k) \in \mathcal{N}_{-\infty}(\gamma,\beta)$.

Our method solves the sign-adjusted KKT system Eq. (3) and uses the merit function $\hat{\varphi}$ given by Eq. (13) as a guide to choose the parameters (α,μ,ω) — seen here as variables — and the next iterate. This way, we have to assure that the next iterate is in $\mathcal{N}_{-\infty}(\gamma,\beta)$, which is equivalent to guarantee that this point is within an appropriate distance from the central path.

For this, we will establish functions, on variables (α,μ,ω), that can be used as constraints of the global minimization subproblem of merit function $\hat{\varphi}$ and that represent $\mathcal{N}_{-\infty}(\gamma,\beta)$. In fact, due to Proposition 1 a point $(x,y,z) \in \mathcal{Q}^+$ generated by Algorithm 1 fulfills the conditions given by $\mathcal{N}_{-\infty}(\gamma,\beta)$ if the following inequalities hold:

$$g_C^i(x,y,z) \coloneqq (\rho_C)_i(x,y,z) - \gamma \overline{\rho_C}(x,y,z) \geq 0, \quad \forall i = 1,\ldots,n, \tag{16a}$$

$$h_C(x,y,z) \coloneqq \overline{\rho_C}(x,y,z) - \frac{\nu_k}{\beta}\overline{\rho_C^0} \geq 0. \tag{16b}$$

We need explicit formulae for this inequalities for the next point $(\hat{x},\hat{y},\hat{z})$. To accomplish that, we use Theorems 1 and 2 and the fact that $\hat{\nu} = (1-\alpha)\nu_k$, and we obtain (again changing the variables to α,μ,ω):

$$\begin{aligned} g_C^i(\alpha,\mu,\omega) = (1-\alpha)\left((\rho_C)_i - \gamma\overline{\rho_C}\right) + \alpha\mu(1-\gamma) + \\ + \alpha(\alpha-\omega)\left((L_{0,0})_i - \gamma\overline{L_{0,0}}\right) + \alpha^2\left(\Lambda(\mu,\omega)_i - \gamma\overline{\Lambda(\mu,\omega)}\right). \end{aligned} \tag{17}$$

Using the same argument and defining $\beta_L = \dfrac{\overline{\rho_C^0}}{\beta}$ we also have

$$h_C(\alpha,\mu,\omega) = (1-\alpha)\left(\overline{\rho_C} - \beta_L\nu_k\right) + \alpha\mu + \alpha(\alpha-\omega)\overline{L_{0,0}} + \alpha^2\overline{\Lambda(\mu,\omega)}. \tag{18}$$

In this view, by ensuring that inequalities $g_C^i(\alpha,\mu,\omega) \geq 0 \quad \forall i = 1,\ldots,n$ and $h_C(\alpha,\mu,\omega) \geq 0$ hold for a tuple (α,μ,ω), we guarantee that the next point — which will be taken as in Eq. (8) — is in $\mathcal{N}_{-\infty}(\gamma,\beta)$. These inequalities become the main part of the constraints we will use in the optimization subproblem.

Briefly, the global optimization subproblem can be written as

$$\begin{aligned} \min_{(\alpha,\mu,\omega)} \quad & \hat{\varphi}(\alpha,\mu,\omega) \\ \text{s.t.} \quad & \begin{cases} g_C^i(\alpha,\mu,\omega) \geq 0,\ i = 1,\ldots,n, \\ h_C(\alpha,\mu,\omega) \geq 0, \\ \alpha \in [0,1],\ \mu \geq 0,\ \omega \geq 0. \end{cases} \end{aligned} \tag{19}$$

Since φ, g_C^i, for $i = 1, \ldots, n$, and h_C are all polynomials, subproblem (19) is a Polynomial Optimization Problem (POP) and it is, in general, hard to solve. For some instances, it can be a NP-Hard problem [9].

In Section 2.3 we describe our method to solve (19).

2.3 Solving the polynomial optimization subproblem

We now outline the method that we are using to solve the subproblem (19), *i.e.*, how we take step 7 of Algorithm 1. Here, we treat the integers i, j and k as local variables.

For a fixed $\mu \geq 0$, subproblem (19) becomes

$$\begin{array}{ll} \min\limits_{\alpha,\omega} & \hat{\varphi}(\alpha,\omega) \\ \text{s.t.} & \begin{cases} c_i(\alpha,\omega) \geq 0, \ i = 1,\ldots,n+1 \\ \alpha \in [0,1], \ \omega \geq 0 \end{cases} , \end{array} \tag{20}$$

where any of the constraints g_C^i and h_C of (19) are represented by c_j.

Let Ξ be the set of all pairs (α,ω) that satisfy all constraints $c_i(\alpha,\omega) \geq 0$, $i = 1,\ldots,n+1$, $\alpha \in [0,1]$, $\omega \geq 0$, so that the problem (20) above is equivalent to

$$\begin{array}{ll} \min\limits_{\alpha,\omega} & \varphi(\alpha,\omega) \\ \text{s.t.} & (\alpha,\omega) \in \Xi \end{array} .$$

We shall focus on studying and describing this set Ξ aiming at simplifying the solution of the problem above.

For that, we need the following notation: each *constraint* $c_i(\alpha,\omega) \geq 0$ defines an *equation* $c_i(\alpha,\omega) = 0$, which in turn implicitly defines a function $\alpha = f_{c_i}(\omega)$ (possibly more than one) for which $\alpha \leq f_{c_i}(\omega)$ implies $c_i(\alpha,\omega) \geq 0$, that is, f_{c_i} is part of the trace of the surface $c_i(\alpha,\omega) = 0$ on the plane $(\omega\alpha)$.

2.3.1 A road map

We will *build* p regions Ω_k such that $\bigcup_{k=1}^{p} \Omega_k = \Xi$ where

$$\Omega_k := \{(\alpha,\omega) : \alpha \in [0,1], \ \omega \in [\omega_k, \omega_{k+1}], \ \alpha \leq f_{R_k}(\omega)\},$$

and $R_k(\alpha,\omega) \geq 0$ is a *single constraint* uniquely associated to Ω_k such that $R_k(\alpha,\omega) = c_i(\alpha,\omega)$ for some $i = 1,\ldots,n+1$, so that problem (20) becomes equivalent to

$$\begin{array}{ll} \min\limits_{\alpha,\omega} & \varphi(\alpha,\omega) \\ \text{s.t.} & (\alpha,\omega) \in \Omega_k, \ k = 1,\ldots,p \end{array} .$$

If we manage to build such regions Ω_k, each sub-subproblem for a fixed k given by

$$\begin{array}{ll} \min\limits_{\alpha,\omega} & \varphi(\alpha,\omega) \\ \text{s.t.} & (\alpha,\omega) \in \Omega_k \end{array} \tag{P_k}$$

is *easy to solve* with a *closed-form expression*, since the first-order optimality conditions for each problem (P_k) are equivalent to the intersection of two conic sections.

Further elaborating on the ideas above, a second degree polynomial equation in two variables is a conic section, and there are up to 4 points of intersection between any such two conic sections. Finding these intersections is equivalent to finding all real roots of a fourth degree real-valued polynomial of a single real variable, after some algebraic manipulations, hence the closed-form expression. In our implementation, we used the closed-form routine in `C`-language by Herbison-Evans [6].

Note that optimality of the subproblem is guaranteed provided we manage to build such regions, as they are obtained by construction, unless such construction fails at some step, in which case we would have $\Xi = \varnothing$. Furthermore, Ω_k is closed and bounded, but not necessarily convex. In addition, the intersection of any consecutive regions Ω_k and Ω_{k+1} is *nonempty* since it is the line segment joining $(0, \omega_{k+1})$ and $(\alpha_{k+1}, \omega_{k+1})$. A possible illustration of Ω_k is in Figure 1.

Definition 3. We call the constraint $R_k(\alpha, \omega)$ as above the *relevant constraint* in Ω_k.

We now outline how to solve exactly the subproblem by construction in two steps, first building each region Ω_k and then solving the subproblem.

2.3.2 Building the partitions – an algorithmic outline

Each of the intervals $[\omega_k, \omega_{k+1}]$ and its relevant constraint R_k are found by construction.

We start by setting $k \leftarrow 0$, $\omega_0 \leftarrow 0$.

For each k, set:

$R_k \leftarrow c_1(\alpha, \omega)$,

$\alpha_k \leftarrow f_{R_k}(\omega_k)$, and

ω_{k+1} as the smallest ω such that $f_{R_k}(\omega)$ is continuous on $[\omega_k, \omega_{k+1}]$ and such that $\alpha \leq f_{R_k}(\omega)$ implies $R_k(\alpha, \omega) \geq 0$. This is trivial since $R_k(\alpha, \omega) = 0$ is a second degree polynomial, thus requiring $\mathcal{O}(1)$ operations to solve this inequality.

We proceed by computing the intersections of R_k with the other constraints $c_i(\alpha, \omega)$, $i = 2, \ldots, n+1$.

For each $i > 1$

For each $j = 1, \ldots, 4$ we compute the intersections (α_j, ω_j) of R_k with $c_i(\alpha, \omega)$, discarding intersections where:

- $\alpha_j < 0$
- $\omega_j \leq \omega_k$
- There is some discontinuity or any asymptote of $c_i(\alpha, \omega)$ in $[\omega_k, \omega_j]$.

· The derivative of $f_{R_k}(\omega)$ at ω_j is greater than the derivative of $f_{c_i}(\omega)$ at ω_j. This means that R_k remains the relevant constraint in (α_j, ω_j) when compared with $c_i(\alpha, \omega)$. Note that concavity is not important in this case, only the analysis of the derivatives at ω_j.

If none of the above conditions happens, the derivative of $f_{c_i}(\omega)$ at ω_j is *smaller* than the derivative of $f_{R_k}(\omega)$ at ω_j, which implies that $f_{c_i}(\omega) \leq f_{R_k}(\omega)$ in $[\omega_k, \omega_j]$, so we accept c_i as the new relevant constraint in this partition, updating $R_k \leftarrow c_i$, $\omega_{k+1} \leftarrow \omega_j$, $\alpha_{k+1} \leftarrow \alpha_j$, $i \leftarrow i+1$, stopping when $i = n+1$.

After inspecting these $n+1$ intersections we end up with the relevant constraint R_k, $[\omega_k, \omega_{k+1}]$ and $[\alpha_k, \alpha_{k+1}]$, *i.e.*, Ω_k.

This step is well-defined — if there are no acceptable intersections then $R_k \leftarrow c_1(\alpha, \omega)$ is the relevant constraint we seek and the next partition $[\omega_k, \omega_{k+1}]$ is the whole domain of ω.

We set $k \leftarrow k+1$ and proceed.

Remark 4. Each accepted intersection pair (α_k, ω_k) is simply stored for later use at the optimization step; α_k plays no bounding role.

Noting that there are up to $2n(n+1)$ intersections of two different constraint-associated conics, we stop when $k = 2n(n+1)$ or $\alpha_{k+1} = 0$ or R_k does not intercept any other constraint c_i for any $\omega > \omega_k$. This means that in the worst case $p = \mathcal{O}(n^2)$, therefore solving the subproblem involves $\mathcal{O}(n^3)$ operations, which does not affect the overall number of operations. Fortunately, for all iterations $p \simeq 10$ in all problems of our test set.

As illustration, these partitions and the traces of their relevant constraints R_k may have two aspects: (i) where the last constraint ends in an asymptote as in Fig. 1a; or (ii) where the last constraint ends in a zero of the equation associated with the constraint as in Fig. 1b.

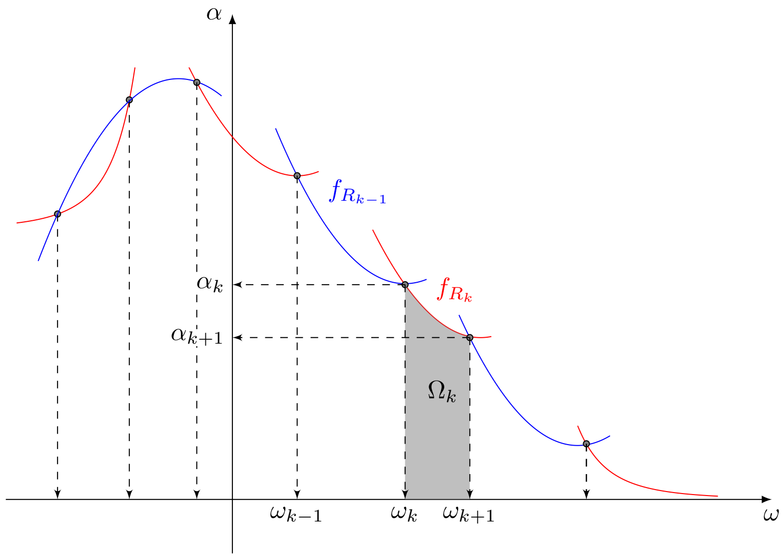
There are implementation details, for instance in the choice of the initial relevant constraint at each interval (*e.g.* by choosing a better initial constraint as candidate R_k), which accelerate the method, but they are not important for the description above.

2.3.3 Solving the subproblem

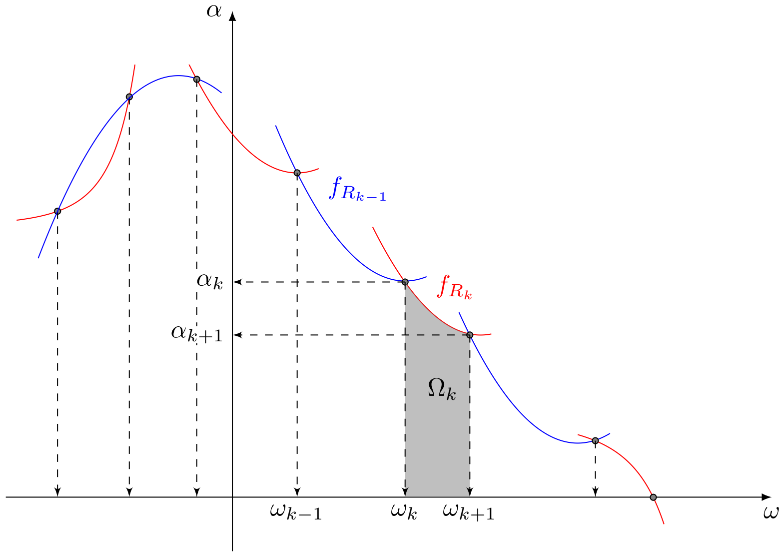
Now we solve p subproblems

$$\begin{array}{ll} \min\limits_{\alpha,\omega} & \varphi(\alpha,\omega) \\ & \alpha \leq f_{R_k}(\omega) \\ \text{s.t.} & \omega \in [\omega_k, \omega_{k+1}] \\ & k = 1, \ldots, p \end{array},$$

but here we have computed all regions Ω_k, *marked in gray* in the graphs on Figure 1, with the traces of the respective relevant constraints R_k.



(a) Last relevant constraint ends in an asymptote.



(b) Last relevant constraint ends in a zero of the equation associated with the constraint.

Figure 1: Aspects of possible traces of relevant constraints R_k for each interval $[\omega_k, \omega_{k+1}]$.

Finally, the optimum is obtained by picking the best of all local optima of problems (P_k), for $k = 1, \dots, p$.

3 Convergence results

We now are ready to establish the major convergence results for OCPM. Once more, it is important to stress that our approach is to treat OCPM on both the theoretical and the practical points of view, ensuring that the analyses are performed on the algorithm that in fact is being implemented.

Algorithm 1 is well-defined since $\alpha = 0$ is always a solution. However, this is not enough for convergence, as with $\alpha = 0$ the sequence of points generated by the algorithm is fixed. What we have to prove is that there is a fixed $\delta > 0$ such that, at each iteration, there is a solution $(\tilde{\alpha}, \tilde{\mu}, \tilde{\omega})$, with $\tilde{\alpha} > \delta$, which in turn guarantees a suitable improvement in the merit function. This will be shown in Theorem 3.

It is inconsequential how arbitrary $(\tilde{\alpha}, \tilde{\mu}, \tilde{\omega})$ is, for its existence is sufficient to guarantee that there is a global solution $(\alpha^*, \mu^*, \omega^*)$ of problem (19), with $\hat{\varphi}(\alpha^*, \mu^*, \omega^*) \leq \hat{\varphi}(\tilde{\alpha}, \tilde{\mu}, \tilde{\omega})$. Unfortunately, to our knowledge, there is no closed form solution of (19) and thus we cannot use it directly in our analysis.

Moreover, we will show that Algorithm 1 has a Q-linear rate of convergence and is polynomial. For that, the analysis relies and is adapted to OCPM from the seminal texts of Wright [15, ch. 6], of Zhang [17] and of Zhang and Zhang [18, 19] that analyze infeasible IPMs as well as Mehrotra's IPM.

3.1 Technical results

We begin the analysis by proving some technical results that are keystones of our results. For that, we discuss a condition that an initial point should meet to guarantee that OCPM converges.

An assumption about the initial point. When analyzing convergence and polynomiality of infeasible IPMs, in particular of a Mehrotra type like our OCPM, most works make use of an assumption that establishes a condition over the distance between the initial point and an optimal solution (see [14, 16, 17, 18, 19]). Such assumptions allow for the establishment of a bound for sequence $\left\{\|(x^k, z^k)\|_p\right\}$, for some p-norm, which is vital for those analyses.

The assumptions used in the works cited above, though allow their authors to prove a polynomial order of convergence, are impracticable on the implementation side, since they use initial points that lead to bad numerical behavior of the method and, at the same time, they rely on the knowledge *a priori* of a bound for an optimal solution, which is not available. Instead, we will use the following assumption, which avoids such issues.

Assumption 1. *For an initial point* (x^0, y^0, z^0)*, where* $(x^0, z^0) > 0$*, there is an optimal solution* (x^*, y^*, z^*) *of problems* (P-D) *that satisfies*

$$\frac{2(x^0)^T z^0 + (x^0)^T z^* + (x^*)^T z^0}{(x^0)^T z^0 \min_i\{x_i^0, z_i^0\}} \left\|(x^0, z^0) - (x^*, z^*)\right\| < \varsigma^4, \tag{21}$$

where ς *is*

$$\varsigma := \max\left\{\left|A_{ij}\right|, \left|b_i\right|, \left|c_j\right|, \text{ for } 1 \leq i \leq m \text{ and } 1 \leq j \leq n\right\}, \tag{22}$$

which is the maximum absolute value of all input data for problems (P-D)*.*

The positive constant ς represents the LP data size measure. Moreover and without loss of generality, one can suppose that $\varsigma \geq 1$. For this to be true, it is enough to use a trivial scaling of data. However, the data scaling heuristics used in IPMs solvers generally are much more sophisticated.

Assumption 1 is only a theoretical hypothesis that is not necessary in the implementation of OCPM, that is, Algorithm 1 does not rely on it in practice. This assumption, as usual, establishes a condition on the distance between the initial point and an optimal solution, however, the knowledge of a bound to the value of the distance between the initial point and an optimal solution is not needed beforehand.

Furthermore, we state that the restrictiveness of Assumption 1 *is not sharp* for it was verified to be true for all problems in our test set (see Section 4), when the initial point (x^0, y^0, z^0) is given by Mehrotra's [10] heuristic and the Curtis-Reid scaling is used. The role of this approach is to guarantee that the analysis is performed on the algorithm that is *de facto* implemented.

Nevertheless the order of converge of OCPM given in what follows is higher than the ones provided by our peers — yet still polynomial —, Assumption 1 allowed us to use Mehrotra initial point, whose good practical performance is well established. Future works can determine if other initial points comply with this assumption and, in this case, such points can be used on OCPM with guaranteed convergence.

The next remark will be useful in several parts of our analysis.

Remark 5. Let $(\Delta x, \Delta y, \Delta z)$ be any triplet of directions. If

$$A\Delta x = 0 \text{ and } A^T\Delta y + \Delta z = 0, \tag{23}$$

then

$$\Delta x^T \Delta z = -\Delta x^T(A^T\Delta y) = -(A\Delta x)^T\Delta y = 0. \tag{24}$$

Let (x^k, y^k, z^k) be the point generated at iteration k. The next lemma shows an upper bound for $\nu_k \left\|(x^k, z^k)\right\|_1$.

Lemma 1. *Suppose that Assumption 1 holds for problem* (P-D)*, where* (x^0, y^0, z^0) *is the initial point given by Mehrotra's heuristic [10] and* (x^*, y^*, z^*) *is an optimal solution. Then inequality*

$$\nu_k \left\|(x^k, z^k)\right\|_1 \left\|(x^0, z^0) - (x^*, z^*)\right\| \leq \varsigma^4 n \overline{\rho_C^k}$$

holds for any point $(x^k, y^k, z^k) \in \mathcal{N}_{-\infty}(\gamma, \beta)$*, with* $\beta = 1$*.*

Proof. This lemma is an adaptation of Lemma 6.3 in [15, p. 113]. Following that, let

$$(\tilde{x}, \tilde{y}, \tilde{z}) := \nu_k(x^0, y^0, z^0) + (1 - \nu_k)(x^*, y^*, z^*) - (x^k, y^k, z^k)$$

be an auxiliary point. By hypothesis, (x^0, y^0, z^0) is the initial point given by Mehrotra's heuristic [10], (x^*, y^*, z^*) is an optimal solution of (P-D).

A straightforward calculation shows that $A\tilde{x} = 0$ and $A^T\tilde{y} + \tilde{z} = 0$. Therefore, because of Eq. (23), Eq. (24) holds, that is, $(\tilde{x})^T\tilde{z} = 0$. Thus, we have

$$\begin{aligned} 0 = \tilde{x}^T\tilde{z} &= (\nu_k x^0 + (1-\nu_k)x^* - x^k)^T(\nu_k z^0 + (1-\nu_k)z^* - z^k) \\ &= \nu_k^2(x^0)^T z^0 + (1-\nu_k)^2(x^*)^T z^* + \nu_k(1-\nu_k)((x^0)^T z^* + (x^*)^T z^0) \\ &\quad + (x^k)^T z^k - \nu_k((x^k)^T z^0 + (x^0)^T z^k) - (1-\nu_k)((x^k)^T z^* + (x^*)^T z^k). \end{aligned} \tag{25}$$

Since $(x^k, z^k) > 0$ and $(x^*, z^*) \geq 0$, $((x^k)^T z^* + (x^*)^T z^k) \geq 0$ holds. Moreover, (x^*, y^*, z^*) is an optimal solution, and so $(x^*)^T z^* = 0$. Using these remarks and taking into account that $\nu_k \in (0, 1)$, we can derive from Eq. (25) the following inequality

$$\nu_k((x^k)^T z^0 + (x^0)^T z^k) \leq \nu_k^2(x^0)^T z^0 + (x^k)^T z^k + \nu_k(1-\nu_k)((x^0)^T z^* + (x^*)^T z^0). \tag{26}$$

Since $(x^k, z^k) > 0$, one can see that

$$\begin{aligned} \min_i\{x_i^0, z_i^0\} \left\|(x^k, z^k)\right\|_1 &= \min_i\{x_i^0, z_i^0\} \left(\sum_{i=1}^n x_i^k + \sum_{i=1}^n z_i^k\right) \\ &\leq \min_i(z_i^0) \left\|x^k\right\|_1 + \min_i(x_i^0) \left\|z^k\right\|_1 \leq (x^k)^T z^0 + (x^0)^T z^k, \end{aligned}$$

which implies

$$\left\|(x^k, z^k)\right\|_1 \leq \min_i\{x_i^0, z_i^0\}^{-1}((x^k)^T z^0 + (x^0)^T z^k).$$

Comparing the right side of this inequality with Eq. (26) we obtain

$$\nu_k \left\|(x^k, z^k)\right\|_1 \leq \frac{\left[\nu_k^2(x^0)^T z^0 + (x^k)^T z^k + \nu_k(1-\nu_k)((x^0)^T z^* + (x^*)^T z^0)\right]}{\min_i\{x_i^0, z_i^0\}}.$$

Since $\nu_k \in (0, 1)$, $\nu_k^2 < \nu_k$ and $\nu_k(1-\nu_k) < \nu_k$; furthermore, for any k, $(x^k)^T z^k = n\overline{\rho_C^k}$. Using these arguments on the previous inequality we obtain

$$\nu_k \left\|(x^k, z^k)\right\|_1 \leq \min_i\{x_i^0, z_i^0\}^{-1} \left[n\nu_k\overline{\rho_C^0} + n\overline{\rho_C^k} + \nu_k((x^0)^T z^* + (x^*)^T z^0)\right],$$

which is similar to Eq. (6.22) in [15, p. 114].

Since $\beta = 1$ and $(x^k, y^k, z^k) \in \mathcal{N}_{-\infty}(\gamma, \beta)$ we have $\nu_k \leq \overline{\rho_C^k}/\overline{\rho_C^0}$. Thus we obtain

$$\begin{aligned} \nu_k \left\|(x^k, z^k)\right\|_1 &\leq \min_i\{x_i^0, z_i^0\}^{-1} \left[2n\overline{\rho_C^k} + \frac{n\overline{\rho_C^k}}{(x^0)^T z^0}((x^0)^T z^* + (x^*)^T z^0)\right] \\ &= \min_i\{x_i^0, z_i^0\}^{-1} \left[2 + \frac{((x^0)^T z^* + (x^*)^T z^0)}{(x^0)^T z^0}\right] n\overline{\rho_C^k}. \end{aligned}$$

Multiplying both sides of this last inequality by $\|(x^0, z^0) - (x^*, z^*)\|$ and using Assumption 1, finally we have

$$\nu_k \left\|(x^k, z^k)\right\|_1 \left\|(x^0, z^0) - (x^*, z^*)\right\| \leq$$
$$\leq \frac{2(x^0)^T z^0 + ((x^0)^T z^* + (x^*)^T z^0)}{(x^0)^T z^0 \min_i \{x_i^0, z_i^0\}} \left\|(x^0, z^0) - (x^*, z^*)\right\| n\overline{\rho_C^k} < \varsigma^4 n\overline{\rho_C^k}. \qquad \square$$

$\square$

The following proposition will be used in the next lemmas. The last inequality is easily derived from the arithmetic-geometric mean relation.

Proposition 2. *Let D be a nonsingular diagonal matrix of order n and let u and v be any real vectors of dimension n. Then*

$$\|uv\| \leq \|Du\| \left\|D^{-1}v\right\| \leq \frac{1}{2}\left(\|Du\|^2 + \left\|D^{-1}v\right\|^2\right). \tag{27}$$

Henceforth, given (x^k, y^k, z^k) generated by OCPM, we define matrix D^k as $D^k = (X^k)^{-1/2}(Z^k)^{1/2}$.

The following lemmas show the existence of a bound for the x and z components of the directions we use in OCPM. The first one is an adaptation of Lemmas 6.5 and 6.6 of [15, pp. 115-118].

Lemma 2. *Suppose that Assumption 1 holds. Then, for a solution of* (4) *we have*

$$\left\|(\Delta x^{af})^k (\Delta z^{af})^k\right\| \leq \frac{(1+2\varsigma^4)^2}{\gamma} n^2 \overline{\rho_C^k}. \tag{28}$$

Proof. Following the steps of the proof of Lemma 6.5 in Wright [15] let

$$(\Delta x, \Delta y, \Delta z) := ((\Delta x^{\mathrm{af}})^k, (\Delta y^{\mathrm{af}})^k, (\Delta z^{\mathrm{af}})^k) + \nu_k(x^0 - x^*, y^0 - y^*, z^0 - z^*),$$

be an auxiliary direction, where (x^0, y^0, z^0) is the initial point given by Mehrotra's heuristic and (x^*, y^*, z^*) is an optimal solution, and assume that for both points Inequality (21) holds.

Hence,

$$A\Delta x = A((\Delta x^{\mathrm{af}})^k + \nu_k(x^0 - x^*)) = -r_P^k + \nu_k r_P^0 = -r_P^k + r_P^k = 0,$$

and

$$\begin{aligned} A^T\Delta y + \Delta z &= A^T((\Delta y^{\mathrm{af}})^k + \nu_k(y^0 - y^*)) + ((\Delta z^{\mathrm{af}})^k + \nu_k(z^0 - z^*)) \\ &= -r_D^k + \nu_k r_D^0 = -r_D^k + r_D^k = 0. \end{aligned}$$

Therefore, $(\Delta x, \Delta y, \Delta z)$ satisfies Eq. (23) and hence Eq. (24) holds, *i.e.*, $\Delta x^T \Delta z = 0$. Thus,

$$0 = \Delta x^T \Delta z = \left((\Delta x^{\mathrm{af}})^k + \nu_k(x^0 - x^*)\right)^T \left((\Delta z^{\mathrm{af}})^k + \nu_k(z^0 - z^*)\right). \tag{29}$$

Using this auxiliary direction and the third line of (4), we obtain

$$Z^k \left((\Delta x^{\text{af}})^k + \nu_k(x^0 - x^*)\right) + X^k \left((\Delta z^{\text{af}})^k + \nu_k(z^0 - z^*)\right)$$
$$= -x^k z^k + \nu_k Z^k(x^0 - x^*) + \nu_k X^k(z^0 - z^*).$$

Multiplying all this expression by $(X^k Z^k)^{-1/2}$, and since $D^k = (X^k Z^k)^{-1/2} Z^k$ and $(D^k)^{-1} = (X^k Z^k)^{-1/2} X^k$, the following holds:

$$D^k \left((\Delta x^{\text{af}})^k + \nu_k(x^0 - x^*)\right) + (D^k)^{-1} \left((\Delta z^{\text{af}})^k + \nu_k(z^0 - z^*)\right) =$$
$$- (x^k z^k)^{1/2} + \nu_k D^k(x^0 - x^*) + \nu_k (D^k)^{-1}(z^0 - z^*). \quad (30)$$

Since Eq. (29) holds, if we take the squared norm of the left-hand side of Eq. (30) and use the Pythagorean Theorem we obtain

$$\left\|D^k \left((\Delta x^{\text{af}})^k + \nu_k(x^0 - x^*)\right) + (D^k)^{-1} \left((\Delta z^{\text{af}})^k + \nu_k(z^0 - z^*)\right)\right\|^2 =$$
$$\left\|D^k \left((\Delta x^{\text{af}})^k + \nu_k(x^0 - x^*)\right)\right\|^2 + \left\|(D^k)^{-1} \left((\Delta z^{\text{af}})^k + \nu_k(z^0 - z^*)\right)\right\|^2.$$

Using the triangular inequality in Eq. (30), we obtain

$$\left\|D^k \left((\Delta x^{\text{af}})^k + \nu_k(x^0 - x^*)\right)\right\|^2 + \left\|(D^k)^{-1} \left((\Delta z^{\text{af}})^k + \nu_k(z^0 - z^*)\right)\right\|^2 \leq$$
$$\left\{\left\|(x^k z^k)^{1/2}\right\| + \nu_k \left\|D^k(x^0 - x^*)\right\| + \nu_k \left\|(D^k)^{-1}(z^0 - z^*)\right\|\right\}^2.$$

Since the second term of this inequality is nonnegative, we have

$$\left\|D^k \left((\Delta x^{\text{af}})^k + \nu_k(x^0 - x^*)\right)\right\| \leq$$
$$\left\|(x^k z^k)^{1/2}\right\| + \nu_k \left\|D^k(x^0 - x^*)\right\| + \nu_k \left\|(D^k)^{-1}(z^0 - z^*)\right\|.$$

A simple application of the triangular inequality and the addition of an extra term $\nu_k \left\|(D^k)^{-1}(z^0 - z^*)\right\|$ to the right-hand side gives

$$\left\|D^k(\Delta x^{\text{af}})^k\right\| = \left\|D^k \left((\Delta x^{\text{af}})^k + \nu_k(x^0 - x^*) - \nu_k(x^0 - x^*)\right)\right\|$$
$$\leq \left\|(x^k z^k)^{1/2}\right\| + 2\nu_k \left\|D^k(x^0 - x^*)\right\| + 2\nu_k \left\|(D^k)^{-1}(z^0 - z^*)\right\|. \quad (31)$$

Now, for the first term of the above inequality, we have

$$\left\|(x^k z^k)^{1/2}\right\| = \left(\sum_{i=1}^{n} x_i^k z_i^k\right)^{1/2} = ((x^k)^T z^k)^{1/2} = n^{1/2} \overline{\rho_C^k}^{1/2} \leq \frac{n}{\gamma^{1/2}} \overline{\rho_C^k}^{1/2}, \quad (32)$$

since $\gamma \in (0, 1)$ and $\sqrt{n} \leq n$, for all $n \in \mathbb{N}$, while the norm-2 of matrix D^k is

$$\left\|D^k\right\| = \max_{i=1,\ldots,n} \left|D_{ii}^k\right| = \left\|D^k e\right\|_\infty = \left\|(X^k Z^k)^{-1/2} z^k\right\|_\infty \leq \left\|(X^k Z^k)^{-1/2}\right\| \left\|z^k\right\|_1,$$

and similarly

$$\left\|(D^k)^{-1}\right\| \le \left\|(X^kZ^k)^{-1/2}\right\| \left\|x^k\right\|_1 .$$

Furthermore, since $(x^k, y^k, z^k) \in \mathcal{N}_{-\infty}(\gamma, \beta)$, from Eq. (15), we have

$$\left\|(X^kZ^k)^{-1/2}\right\| = \max_{i=1,\dots,n} \frac{1}{(x_i^k z_i^k)^{1/2}} \le \frac{1}{\left(\gamma\overline{\rho_C^k}\right)^{1/2}}. \tag{33}$$

Using those inequalities for D^k and $(D^k)^{-1}$, Inequality (6.34) in [15] is, in our case,

$$\begin{aligned}
\nu_k \left(\left\|D^k(x^0 - x^*)\right\| + \left\|(D^k)^{-1}(z^0 - z^*)\right\|\right) \le & \\
\le \nu_k \left(\left\|D^k\right\| + \left\|(D^k)^{-1}\right\|\right) \left\|(x^0, z^0) - (x^*, z^*)\right\| & \\
\le \nu_k \left\|(X^kZ^k)^{-1/2}\right\| \left(\left\|x^k\right\|_1 + \left\|z^k\right\|_1\right) \left\|(x^0, z^0) - (x^*, z^*)\right\| & \\
\le \nu_k \left\|(x^k, z^k)\right\|_1 \left\|(x^0, z^0) - (x^*, z^*)\right\| \left\|(X^kZ^k)^{-1/2}\right\| . &
\end{aligned}$$

From Lemma 1 and following the same arguments and substitutions in the proof of Wright's Lemma 6.5 [15] in the inequality above, we have

$$\nu_k \left\|D^k(x^0 - x^*)\right\| + \nu_k \left\|(D^k)^{-1}(z^0 - z^*)\right\| \le \varsigma^4 n\overline{\rho_C^k} \frac{1}{\gamma^{1/2}\overline{\rho_C^k}^{1/2}} = \frac{\varsigma^4}{\gamma^{1/2}} n\overline{\rho_C^k}^{1/2}. \tag{34}$$

If we use Eqs. (32) and (34), and compare them with Eq. (31), we have

$$\left\|D^k(\Delta x^{\mathrm{af}})^k\right\| \le \frac{n}{\gamma^{1/2}}\overline{\rho_C^k}^{1/2} + \frac{2\varsigma^4}{\gamma^{1/2}} n\overline{\rho_C^k}^{1/2} = \frac{1 + 2\varsigma^4}{\gamma^{1/2}} n\overline{\rho_C^k}^{1/2},$$

where $\dfrac{1 + 2\varsigma^4}{\gamma^{1/2}} > 1$, since $\gamma \in (0, 1)$. Therefore, we obtain

$$\left\|D^k(\Delta x^{\mathrm{af}})^k\right\| \le \left(\frac{1 + 2\varsigma^4}{\gamma^{1/2}}\right) n\overline{\rho_C^k}^{1/2} \quad \text{and} \quad \left\|(D^k)^{-1}(\Delta z^{\mathrm{af}})^k\right\| \le \left(\frac{1 + 2\varsigma^4}{\gamma^{1/2}}\right) n\overline{\rho_C^k}^{1/2}. \tag{35}$$

Using Proposition 2 and the above equation, we finally have the result. □ □

It is well established in IPMs to set the centralization parameter μ as a fraction of the average of the complementarity gap. Thus, we set $\mu := \eta\dfrac{x^Tz}{n} = \eta\overline{\rho_C}$, for some $\eta \in (\eta_{\min}, \eta_{\max})$. Using this value for μ, we state and prove the next lemma, which is a modification of Lemma 5.2 of Zhang and Zhang [19]. We will follow their proof, with the appropriate adaptations.

Lemma 3. *Suppose that Assumption 1 holds and set* $\mu := \eta\overline{\rho_C}$*. If* $\omega > \dfrac{\gamma(\sqrt{2\gamma} - \eta)}{(1 + 2\varsigma^4)^2}$*, and* $\eta \in [0, 1]$ *then, inequality*

$$\left\|(\Delta x^c)^k(\Delta z^c)^k\right\| \le \frac{\left[\eta\gamma + \omega(1 + 2\varsigma^4)^2\right]^2}{2\gamma^3} n^4\overline{\rho_C^k} \tag{36}$$

holds for any solution of (6)*; in addition* $\frac{\left[\eta\gamma+\omega(1+2\varsigma^4)^2\right]^2}{2\gamma^3} > 1$ *also holds.*

Proof. Multiplying the third line of Eq. (6) by $(X^k Z^k)^{-1/2}$ we obtain

$$D^k(\Delta x^{\mathrm{c}})^k + (D^k)^{-1}(\Delta z^{\mathrm{c}})^k = (X^k Z^k)^{-1/2}(\mu e - \omega(\Delta x^{\mathrm{af}})^k(\Delta z^{\mathrm{af}})^k).$$

In addition, $((\Delta x^{\mathrm{c}})^k, (\Delta y^{\mathrm{c}})^k, (\Delta z^{\mathrm{c}})^k)$ satisfies Eq. (23) and then Eq. (24) holds for it. Thus, we have

$$\left\|D^k(\Delta x^{\mathrm{c}})^k + (D^k)^{-1}(\Delta z^{\mathrm{c}})^k\right\|^2 = \left\|D^k(\Delta x^{\mathrm{c}})^k\right\|^2 + \left\|(D^k)^{-1}(\Delta z^{\mathrm{c}})^k\right\|^2.$$

and then

$$\left\|D^k(\Delta x^{\mathrm{c}})^k\right\|^2 + \left\|(D^k)^{-1}(\Delta z^{\mathrm{c}})^k\right\|^2 \le \left\|(X^k Z^k)^{-1/2}\right\|^2 \left\|\mu e - \omega(\Delta x^{\mathrm{af}})^k(\Delta z^{\mathrm{af}})^k\right\|^2.$$

Now, using Eq. (33) and the triangle inequality we obtain

$$\left\|D^k(\Delta x^{\mathrm{c}})^k\right\|^2 + \left\|(D^k)^{-1}(\Delta z^{\mathrm{c}})^k\right\|^2 \le \left(\gamma\overline{\rho_C^k}\right)^{-1}\left(\mu\sqrt{n} + \omega\left\|(\Delta x^{\mathrm{af}})^k(\Delta z^{\mathrm{af}})^k\right\|\right)^2. \tag{37}$$

To use Lemma 2, let $t \coloneqq 1 + 2\varsigma^4$ be a local variable, such that Eq. (28) becomes

$$\left\|(\Delta x^{\mathrm{af}})^k(\Delta z^{\mathrm{af}})^k\right\| \le \frac{t^2}{\gamma} n^2 \overline{\rho_C^k}.$$

Assume that μ is set as $\eta\overline{\rho_C^k}$ and that ω is yet an undetermined parameter, Inequality (37) can be transformed into

$$\begin{aligned}\left\|D^k(\Delta x^{\mathrm{c}})^k\right\|^2 + \left\|(D^k)^{-1}(\Delta z^{\mathrm{c}})^k\right\|^2 &\le (\gamma\overline{\rho_C^k})^{-1}\left(\eta\overline{\rho_C^k}\sqrt{n} + \omega\frac{t^2}{\gamma}n^2\overline{\rho_C^k}\right)^2 \\ &\le \frac{(\eta+\omega t^2)^2}{\gamma^3} n^4 \overline{\rho_C^k}.\end{aligned} \tag{38}$$

Due to the inequalities given by Proposition 2 and from Inequality (38), we derive

$$\left\|(\Delta x^{\mathrm{c}})^k(\Delta z^{\mathrm{c}})^k\right\| \le \frac{1}{2}\left(\left\|D^k(\Delta x^{\mathrm{c}})^k\right\|^2 + \left\|(D^k)^{-1}(\Delta z^{\mathrm{c}})^k\right\|^2\right) \le \frac{\left[\eta\gamma+\omega t^2\right]^2}{2\gamma^3} n^4 \overline{\rho_C^k}.$$

Finally, since we assumed that $\omega > \frac{\gamma(\sqrt{2\gamma}-\eta)}{t^2}$, and $\eta \in [0,1]$ we have

$$\frac{\left[\eta\gamma+\omega t^2\right]^2}{2\gamma^3} > \frac{\left[\eta\gamma + \frac{\gamma(\sqrt{2\gamma}-\eta)}{t^2}t^2\right]^2}{2\gamma^3} = \frac{\left[\eta\gamma+\sqrt{2\gamma^3}-\eta\gamma\right]^2}{2\gamma^3} = 1. \quad \square$$

□

Remark 6. Note that the requirement that $\omega > \dfrac{\gamma(\sqrt{2\gamma}-\eta)}{t^2}$ used on Lemma 3 gives us a relation between ω and η, and thus, between ω and μ. This relation is required to guarantee that the corrector direction decreases sufficiently. In fact, since it seems natural that one cannot hope to achieve good behavior of an IPM if there is no centralization or correction, that is, if $\omega = \eta = 0$, this inequality guarantees that this will never happen. For instance, if $\omega = 0$, one must have $\eta > \sqrt{2\gamma}$. On the other hand, if $\eta = 0$, then $\omega > \sqrt{2}\gamma^{\frac{3}{2}}/t^2$, where $t = (1 + 2\varsigma^2)$. Also, this relation could give one a path to explore IPMs where no centralization parameter μ is used, but with a optimized choice of ω.

The following result is a consequence of Lemmas 2 and 3, similar to Corollary 5.1 in [19, pg. 310] and will be used to prove the main theorem.

Lemma 4. *Suppose the same hypothesis of Lemmas 2 and 3. Then inequality*

$$\left\|(\Delta x^{af})^k(\Delta z^{c})^k + (\Delta x^{c})^k(\Delta z^{af})^k\right\| \leq \frac{\sqrt{2}(1+2\varsigma^4)\left[\eta\gamma + \omega(1+2\varsigma^4)^2\right]}{\gamma^2}\, n^3\overline{\rho^k_C} \tag{39}$$

holds.

Proof. Let $t_1 := \dfrac{1+2\varsigma^4}{\gamma^{1/2}}$ and $t_2 := \dfrac{\left[\eta\gamma + \omega(1+2\varsigma^4)^2\right]^2}{2\gamma^3}$ be local variables. From Eqs. (27), (35) and (36) it follows that

$$\begin{aligned}\left\|(\Delta x^{\mathrm{af}})^k(\Delta z^{\mathrm{c}})^k\right\| &\leq \left\|D^k(\Delta x^{\mathrm{af}})^k\right\| \left\|(D^k)^{-1}(\Delta z^{\mathrm{c}})^k\right\| \\ &\leq t_1 n\overline{\rho^k_C}^{1/2}(t_2 n^4\overline{\rho^k_C})^{1/2} = (t_1 t_2^{1/2})n^3\overline{\rho^k_C}.\end{aligned}$$

Similarly, $\left\|(\Delta x^{\mathrm{c}})^k(\Delta z^{\mathrm{af}})^k\right\| \leq (t_1 t_2^{1/2})n^3\overline{\rho^k_C}$. Thus, by the triangle inequality we have

$$\left\|(\Delta x^{\mathrm{af}})^k(\Delta z^{\mathrm{c}})^k + (\Delta x^{\mathrm{c}})^k(\Delta z^{\mathrm{af}})^k\right\| \leq 2(t_1 t_2^{1/2})n^3\overline{\rho^k_C}.$$

Finally, we obtain

$$2(t_1 t_2^{1/2}) = 2\left(\frac{1+2\varsigma^4}{\gamma^{1/2}}\right)\cdot\left(\frac{\eta\gamma + \omega(1+2\varsigma^4)^2}{\sqrt{2}\gamma^{3/2}}\right) = \frac{\sqrt{2}(1+2\varsigma^4)\left[\eta\gamma + \omega(1+2\varsigma^4)^2\right]}{\gamma^2}. \qquad \square$$

$\square$

3.2 OCPM complexity and convergence results

As noted before, Algorithm 1 is well-defined if problem (19) is feasible for some γ and β. We claim that if we set, as we did in Lemmas 3 and 4, the centralization parameter μ as $\tilde{\mu} := \eta\overline{\rho^k_C}$, and we fix ω as $\tilde{\omega} := 0$ there is a steplength $\tilde{\alpha} > 0$ such that $(\tilde{\alpha}, \eta\overline{\rho^k_C}, 0)$ is the desired tuple, *i.e.*, $(\tilde{\alpha}, \eta\overline{\rho^k_C}, 0)$ lies in $\mathcal{N}_{-\infty}(\gamma,\beta)$ and is a point of the feasible set of Algorithm 1. Moreover, this $\tilde{\alpha}$ is large enough to make the merit function $\hat{\varphi}$ decrease polynomially towards zero with Q-linear rate of convergence.

Before showing that, we will rewrite our results for the fixed choices of μ and ω. From now on we assume that those two parameters are fixed according to Corollary 2.

Corollary 2. *Suppose we set* $\mu := \eta\overline{\rho_C^k}$ *and* $\omega := 0$. *Then*

(i) Inequality (36) *of Lemma 3 is rewritten as*

$$\left\|(\Delta x^c)^k(\Delta z^c)^k\right\| \le \frac{\eta^2}{2\gamma}\, n^4\overline{\rho_C^k}, \tag{40}$$

and also $\eta > \sqrt{2\gamma}$;

(ii) Inequality (39) *of Lemma 4 is rewritten as*

$$\left\|(\Delta x^{af})^k(\Delta z^c)^k + (\Delta x^c)^k(\Delta z^{af})^k\right\| \le \frac{\sqrt{2}(1+2\varsigma^4)\eta}{\gamma}\, n^3\overline{\rho_C^k}; \tag{41}$$

(iii) The merit function for the next point given in Theorem 2 depends only on a choice of α *and is given as*

$$\hat{\varphi}(\alpha) = (1-\alpha)(\nu_k\overline{\rho_L^0} + \overline{\rho_C^k}) + \alpha\eta\overline{\rho_C^k} + \alpha^2\left(\overline{L_{0,0}} + \eta\overline{\rho_C^k}\,\overline{L_{1,0}}\right). \tag{42}$$

Now, let us define an auxiliary function $\theta(\alpha)$ as

$$\theta(\alpha) := \frac{\alpha\left[\nu_k\overline{\rho_L^0} + (1-\eta)\overline{\rho_C^k} - \alpha\left(\overline{L_{0,0}} + \eta\overline{\rho_C^k}\,\overline{L_{1,0}}\right)\right]}{\nu_k\overline{\rho_L^0} + \overline{\rho_C^k}}, \tag{43}$$

so we can derive the following relationship between the present value of the merit function, φ, and the value of the merit function at the next iteration, $\hat{\varphi}$:

$$\hat{\varphi} = (1-\theta(\alpha))\varphi. \tag{44}$$

We have to assure that $\hat{\varphi}$ is nonnegative. For this, it is sufficient to choose a steplength α_k at each iteration k such that $\theta_k := \theta(\alpha_k) < 1$. Moreover, if there is a constant $\tilde{\theta} > 0$ such that $\tilde{\theta} = \liminf(\theta_k)$, then for sufficiently large k

$$\frac{\varphi_{k+1}}{\varphi_k} < (1-\tilde{\theta}) < 1, \tag{45}$$

and the sequence $\{\varphi_k\}$ generated by Algorithm 1 indeed converges to zero Q-linearly.

We will also guarantee that there is an $\tilde{\alpha} > 0$ such that for some $\alpha_k \in (0, \tilde{\alpha}]$ in every iteration k, Eq. (44) holds and every point generated by OCPM is in neighborhood $\mathcal{N}_{-\infty}(\gamma, \beta)$.

For this, let $\tilde{\alpha}$ be chosen as

$$\begin{cases} \tilde{\alpha} := \min\{\tilde{\alpha}_C, \tilde{\alpha}_L\}, \\ \tilde{\alpha}_C := \min\limits_{1\le i\le n}\{\tilde{\alpha}_C^i\}, \\ \tilde{\alpha}_C^i := \max\limits_{\alpha\in(0,1]}\{\alpha : g_C^i(\upsilon, \tilde{\mu}, \tilde{\omega}) \ge 0 \text{ for all } 0 \le \upsilon \le \alpha\}, \quad i = 1, \ldots, n, \\ \tilde{\alpha}_L := \max\limits_{\alpha\in(0,1]}\{\alpha : h_C(\upsilon, \tilde{\mu}, \tilde{\omega}) \ge 0 \text{ for all } 0 \le \upsilon \le \alpha\}, \end{cases} \tag{46}$$

and g_C^i and h_C are given by Eqs. (17) and (18).

Furthermore, we will show that this particular choice ensures that Algorithm 1 generates a sequence $\{\varphi_k\}$ that decreases sufficiently at each step to ensure that $\varphi_k \to 0$ when $k \to \infty$. Therefore, point (x^k, y^k, z^k) generated by OCPM inherits the good IPM convergence properties given by neighborhood $\mathcal{N}_{-\infty}(\gamma, \beta)$, which makes the sequence $\left\{(x^k, y^k, z^k)\right\}$ converge to an optimal solution (x^*, y^*, z^*).

Let us choose $\gamma \in (0,1)$ and $\beta \geq 1$ in a suitable manner in order to well define neighborhood $\mathcal{N}_{-\infty}(\gamma, \beta)$. Note that for the first parameter, Colombo and Gondzio [2] used $\gamma = 1/10$ and and Y. Zhang [17] and Y. Zang and D. Zang [19] used $\gamma \in (0,1)$ that satisfies $\gamma \leq \frac{\min\limits_i (x_i^0 z_i^0)}{(x^0)^T z^0/n}$. By merging this two ideas, in our turn, we will select γ such that

$$\gamma \leq \min \left\{ \frac{\min\limits_i (x_i^0 z_i^0)}{(x^0)^T z^0/n}, \frac{1}{10} \right\},$$

which guarantees that the initial point satisfies inequality (16a).

Arguing in a similar manner, for any $\beta \geq 1$, an initial point always satisfies Eq. (16b). Smaller values of β speed up the reduction of the average of the complementarity residuals, given by $\overline{\rho_C^k}$, when compared with the reduction of the average of the linear residuals, given by $\overline{\rho_L^k}$. Thus, let us set $\beta \coloneqq 1$.

We now prove that $\tilde{\alpha}_C$ and $\tilde{\alpha}_L$ exist.

Lemma 5. *The real number $\tilde{\alpha}_C$ given in Eq.* (46) *is well defined and inequality*

$$\tilde{\alpha}_C \geq \delta_1 / n^4$$

holds, where $\delta_1 = \dfrac{2\eta\gamma(1-\gamma)}{\eta^2 + 4\sqrt{2}\eta(1+2\varsigma^4) + 4(1+2\varsigma^4)^2}$ *is a constant independent of* n.

Proof. First, assume that we set $\mu \coloneqq \eta\overline{\rho_C^k}$ and $\omega \coloneqq 0$, as in Corollary 2. This allows us to rewrite function g_C^i, for $i = 1, \ldots, n$, given in Eq. (17), as depending only on a choice of α, that is,

$$\begin{aligned} g_C^i(\alpha) = (1-\alpha)(\rho_C^k)_i + \alpha\eta\overline{\rho_C^k} + \alpha^2 \left[(L_{0,0}^k)_i + \eta\overline{\rho_C^k}(L_{1,0}^k)_i + (\eta\overline{\rho_C^k})^2 (L_{2,0}^k)_i \right] + \\ - \gamma \left[(1-\alpha)\overline{\rho_C^k} + \alpha\eta\overline{\rho_C^k} + \alpha^2 \left(\overline{L_{0,0}^k} + \eta\overline{\rho_C^k L_{1,0}^k} \right) \right], \end{aligned}$$

in the same manner of what was done in Eq. (42).

Let

$$t_0 \coloneqq (L_{0,0}^k)_i - \gamma\overline{L_{0,0}^k}, \quad t_1 \coloneqq \eta\overline{\rho_C^k}\left((L_{1,0}^k)_i - \gamma\overline{L_{1,0}^k} \right), \text{ and } t_2 \coloneqq (\eta\overline{\rho_C^k})^2 (L_{2,0}^k)_i, \tag{47}$$

be temporary variables. Using the fact that the present point is in neighborhood

$\mathcal{N}_{-\infty}(\gamma,\beta)$ and from the definitions of Eq. (47), we rewrite $g_C^i(\alpha)$ as

$$\begin{aligned} g_C^i(\alpha) &= \underbrace{(1-\alpha)((\rho_C^k)_i - \gamma\overline{\rho_C^k})}_{\geq 0} + (1-\gamma)\eta\overline{\rho_C^k}\alpha + (t_0+t_1+t_2)\alpha^2 \\ &\geq (1-\gamma)\eta\overline{\rho_C^k}\alpha + (t_0+t_1+t_2)\alpha^2 \\ &\geq (1-\gamma)\eta\overline{\rho_C^k}\alpha - (|t_0|+|t_1|+|t_2|)\alpha^2 \\ &= \alpha\left[(1-\gamma)\eta\overline{\rho_C^k} - (|t_0|+|t_1|+|t_2|)\alpha\right] = f^i(\alpha), \end{aligned}$$

where f^i is a concave quadratic depending on α with positive and unique nonzero root. In fact, this positive root is

$$\alpha_C^i \coloneqq \frac{(1-\gamma)\eta\overline{\rho_C^k}}{|t_0|+|t_1|+|t_2|}, \tag{48}$$

and thus, for all $i=1,\ldots,n$, $g_C^i(\alpha) \geq f^i(\alpha) \geq 0$, whenever $\alpha \in [0,\alpha_C^i]$.

Let us obtain bounds for $|t_0|$, $|t_1|$ and $|t_2|$. The definitions of vectors $L_{\ell,j}^k$ given by Eq. (11), by average operator definition and by the vector norm equivalences, give us the following inequalities

$$\left|(L_{\ell,j}^k)_i\right| \leq \left\|L_{\ell,j}^k\right\| \quad \text{and} \quad \left|\overline{L_{\ell,j}^k}\right| \leq \frac{1}{n}\left\|L_{\ell,j}^k\right\|_1 \leq \frac{1}{n^{1/2}}\left\|L_{\ell,j}^k\right\|. \tag{49}$$

Since $\gamma/n^{1/2} < 1$, we obtain

$$\left|(L_{\ell,j}^k)_i - \gamma\overline{L_{\ell,j}^k}\right| \leq \left|(L_{\ell,j}^k)_i\right| + \left|\gamma\overline{L_{\ell,j}^k}\right| \leq \left\|L_{\ell,j}^k\right\| + \frac{\gamma}{n^{1/2}}\left\|L_{\ell,j}^k\right\| < 2\left\|L_{\ell,j}^k\right\|. \tag{50}$$

From Eqs. (11), (47) and (50), we have the following bound for $|t_0|$:

$$|t_0| < 2\left\|L_{0,0}^k\right\| \leq 2\left\|(\Delta x^{\mathrm{af}})^k(\Delta z^{\mathrm{af}})^k\right\|. \tag{51}$$

Since $\omega = 0$ and $\mu = \eta\overline{\rho_C^k}$, it follows from Eq. (7) that $\Delta w^{\mathrm{c}} = \eta\overline{\rho_C^k}\Delta w^{\mu}$. Hence

$$\eta\overline{\rho_C^k}(\Delta x^{\mu})^k = (\Delta x^{\mathrm{c}})^k \quad \text{and} \quad \eta\overline{\rho_C^k}(\Delta z^{\mu})^k = (\Delta z^{\mathrm{c}})^k.$$

Thus,

$$\begin{aligned} \eta\overline{\rho_C^k}L_{1,0}^k = (\Delta x^{\mathrm{af}})^k(\eta\overline{\rho_C^k}\Delta z^{\mu})^k + (\Delta z^{\mathrm{af}})^k(\eta\overline{\rho_C^k}\Delta x^{\mu})^k& \\ = (\Delta x^{\mathrm{af}})^k(\Delta z^{\mathrm{c}})^k + (\Delta z^{\mathrm{af}})^k(\Delta x^{\mathrm{c}})^k&, \end{aligned}$$

and, again, using Eqs. (11) and (50), we obtain a bound for $|t_1|$, that is,

$$|t_1| = \left|\eta\overline{\rho_C^k}\left((L_{1,0}^k)_i - \gamma\overline{L_{1,0}^k}\right)\right| \leq 2\left\|(\Delta x^{\mathrm{af}})^k(\Delta z^{\mathrm{c}})^k + (\Delta x^{\mathrm{c}})^k(\Delta z^{\mathrm{af}})^k\right\|. \tag{52}$$

Finally, if we use Eq. (7) and the values set for μ and ω, we have

$$(\Delta x^{\mathrm{c}})^k(\Delta z^{\mathrm{c}})^k = (\eta\overline{\rho_C^k}\Delta x^{\mu})(\eta\overline{\rho_C^k}\Delta z^{\mu}) = (\eta\overline{\rho_C^k})^2(L_{2,0}^k).$$

Therefore, a bound for $|t_2|$ is given by

$$|t_2| = \left|(\eta\overline{\rho_C^k})^2(L_{2,0}^k)_i\right| = \left|((\Delta x^{\rm c})^k(\Delta z^{\rm c})^k)_i\right| \leq \left\|(\Delta x^{\rm c})^k(\Delta z^{\rm c})^k\right\|. \tag{53}$$

Now, using Eqs. (51) to (53) it is possible to find an upper bound for the denominator of Eq. (48). First, note that

$$|t_0| + |t_1| + |t_2| < \\ 2\left\|(\Delta x^{\rm af})^k(\Delta z^{\rm af})^k\right\| + 2\left\|(\Delta x^{\rm af})^k(\Delta z^{\rm c})^k + (\Delta x^{\rm c})^k(\Delta z^{\rm af})^k\right\| + \left\|(\Delta x^{\rm c})^k(\Delta z^{\rm c})^k\right\|.$$

From Lemmas 2 to 4 and Corollary 2 the above inequality is transformed in

$$|t_0| + |t_1| + |t_2| < \frac{2(1+2\varsigma^4)^2}{\gamma} n^2\overline{\rho_C^k} + \frac{2\sqrt{2}(1+2\varsigma^4)\eta}{\gamma} n^3\overline{\rho_C^k} + \frac{\eta^2}{2\gamma} n^4\overline{\rho_C^k}.$$

Defining the local variable $t = (1+2\varsigma^4)$ we have

$$\begin{aligned} |t_0| + |t_1| + |t_2| &< \left[\frac{2t^2}{\gamma} + \frac{2\sqrt{2}\eta t}{\gamma} + \frac{\eta^2}{2\gamma}\right] n^4\overline{\rho_C^k} \\ &= \frac{1}{2\gamma}\left[4t^2 + 4\sqrt{2}\eta t + \eta^2\right] n^4\overline{\rho_C^k}. \end{aligned} \tag{54}$$

If we consider that $\tilde{\alpha}_C^i$ given by Eq. (46) is the maximum value in $(0, 1]$ such that $g_C^i(\alpha) \geq 0$ for $\alpha \leq \tilde{\alpha}_C^i$, and moreover, that for all α, $g_C^i(\alpha) \geq f^i(\alpha)$ holds, then $\tilde{\alpha}_C^i \geq \alpha_C^i$ for all $i = 1, \ldots, n$.

Therefore, from Eqs. (48) and (54) we have

$$\begin{aligned} \tilde{\alpha}_C^i \geq \alpha_C^i &> \frac{(1-\gamma)\eta\overline{\rho_C^k}}{\frac{1}{2\gamma}\left(4t^2 + 4\sqrt{2}\eta t + \eta^2\right)\overline{\rho_C^k} n^4} \\ &= \frac{2\gamma(1-\gamma)\eta}{4t^2 + 4\sqrt{2}\eta t + \eta^2}\,\frac{1}{n^4} = \frac{\delta_1}{n^4} > 0. \end{aligned}$$

The definition of t gives us δ_1 as required and the proof is complete considering that for some $j \in \{1, \ldots, n\}$, $\alpha_C^j = \tilde{\alpha}_C$. □ □

Lemma 6. *The real number $\tilde{\alpha}_L$ given in Eq. (46) is well defined and inequality $\tilde{\alpha}_L \geq \delta_2/n^{5/2}$ holds, where*

$$\delta_2 = \frac{\gamma\eta}{(1+2\varsigma^4)^2 + 2(1+2\varsigma^4)\eta} \tag{55}$$

is a constant independent of n.

Proof. Assume that $\mu := \eta\overline{\rho_C^k}$ and $\omega := 0$. Thus, function h_C given by Eq. (18) becomes a function that depends only on the choice of α, that is,

$$h_C(\alpha) = (1-\alpha)\left(\overline{\rho_C^k} - \beta_L\nu_k\right) + \alpha\eta\overline{\rho_C^k} + \alpha^2\left(\overline{L_{0,0}^k} + \eta\overline{\rho_C^k}\,\overline{L_{1,0}^k}\right).$$

Since the present point is in $\mathcal{N}_{-\infty}(\gamma, \beta)$, we have $\overline{\rho_C^k} - \beta_L \nu_k \geq 0$ and

$$h_C(\alpha) \geq \alpha \left[\eta\overline{\rho_C^k} - \alpha \left(\left|\overline{L_{0,0}^k}\right| + \left|\eta\overline{\rho_C^k}\,\overline{L_{1,0}^k}\right|\right)\right].$$

Note that $\tilde{\alpha}_L \in (0, 1]$ — see Eq. (46) — is the maximum value such that $h_C(\alpha) \geq 0$, for $\alpha \in [0, \tilde{\alpha}_L]$. Due to the inequality above, its existence is guaranteed and we obtain inequality

$$\tilde{\alpha}_L \geq \frac{\eta\overline{\rho_C^k}}{\left|\overline{L_{0,0}^k}\right| + \left|\eta\overline{\rho_C^k}\,\overline{L_{1,0}^k}\right|}. \tag{56}$$

As in the proof of Lemma 5, we will use Eq. (49) to obtain

$$\left|\overline{L_{0,0}^k}\right| = \frac{1}{n^{1/2}} \left\|L_{0,0}^k\right\| \leq \frac{1}{n^{1/2}} \left\|(\Delta x^{\mathrm{af}})^k (\Delta z^{\mathrm{af}})^k\right\|. \tag{57}$$

Moreover, since $\omega = 0$, $\eta\overline{\rho_C^k}(\Delta x^\mu)^k = (\Delta x^{\mathrm{c}})^k$ and $\eta\overline{\rho_C^k}(\Delta z^\mu)^k = (\Delta z^{\mathrm{c}})^k$. Hence,

$$\begin{aligned}\left|\eta\overline{\rho_C^k}\,\overline{L_{1,0}^k}\right| &= \frac{1}{n^{1/2}} \left\|\eta\overline{\rho_C^k} L_{1,0}^k\right\| \\ &= \frac{1}{n^{1/2}} \left\|(\Delta x^{\mathrm{af}})^k (\Delta z^{\mathrm{c}})^k + (\Delta x^{\mathrm{c}})^k (\Delta z^{\mathrm{af}})^k\right\|.\end{aligned} \tag{58}$$

From Lemmas 2 and 4, together with Eqs. (56) to (58) we derive

$$\begin{aligned}\tilde{\alpha}_L &\geq \frac{\eta\overline{\rho_C^k}}{\frac{1}{n^{1/2}}\left(\|(\Delta x^{\mathrm{af}})^k(\Delta z^{\mathrm{af}})^k\| + \|(\Delta x^{\mathrm{af}})^k \Delta z^{\mathrm{c}} \Delta x^{\mathrm{c}} (\Delta z^{\mathrm{af}})^k\|\right)} \\ &\geq \left(\frac{\gamma\eta}{(1+2\varsigma^4)^2 + 2(1+2\varsigma^4)\eta}\right)\frac{1}{n^{5/2}} = \frac{\delta_2}{n^{5/2}},\end{aligned}$$

where δ_2 is exactly the same given in Eq. (55). □ □

Now we can finally state and proof the Theorem that ensures that the sequence $\{\varphi_k\}$ generated by OCPM is convergent.

Theorem 3 (Convergence of Algorithm 1)**.** *Sequence $\{\varphi_k\}$ generated by Algorithm 1 is such that*

$$\varphi_{k+1} \leq \left(1 - \frac{\hat{\delta}}{n^4}\right)\varphi_k, \tag{59}$$

for all k, where $0 < \hat{\delta} < 1$ and $\hat{\delta}$ is independent of n.

Proof. Using Lemmas 5 and 6 together with Eq. (46), we established that $\tilde{\alpha}$ exists and we can state that $\tilde{\alpha} \geq \delta_1/n^4$, since $(1-\gamma) < 1$, $(2t^2 + 2\sqrt{2}\eta t + \eta^2/2) > (t^2 + 2\eta t)$ for $t = (1 + 2\varsigma^4)$, $\eta > 0$ and $n^4 > n^{5/2}$, which implies $\delta_1/n^4 < \delta_2/n^{5/2}$.

We prove now that for function $\theta(\alpha)$ given by Eq. (43), in the worst case we have $\theta(\tilde{\alpha}) = \mathcal{O}(1/n^4)$. In fact, let $t := (1+2\varsigma^4)$ and $t_1 = \dfrac{t}{\gamma}$ be a local variables. Using Eqs. (11), (57) and (58), Lemmas 2 and 4, and Corollary 2 we derive

$$-\overline{L^k_{0,0}} \geq -\frac{1}{n^{1/2}} \left\|(\Delta x^{\text{af}})^k (\Delta z^{\text{af}})^k\right\| \geq -\frac{(1+2\varsigma^4)^2}{\gamma} n^{3/2}\overline{\rho^k_C} \geq -\gamma t_1^2 n^{3/2}\overline{\rho^k_C}$$

and

$$\begin{aligned} -\eta\overline{\rho^k_C}\,\overline{L^k_{1,0}} \geq -\frac{1}{n^{1/2}} &\left\|(\Delta x^{\text{af}})^k(\Delta z^{\text{c}})^k + (\Delta x^{\text{c}})^k(\Delta z^{\text{af}})^k\right\| \\ &\geq -\frac{2(1+2\varsigma^4)\eta}{\gamma}\, n^2\overline{\rho^k_C} \geq -2\eta t_1 n^{5/2}\overline{\rho^k_C}. \end{aligned}$$

Therefore, since $-\dfrac{\overline{\rho^k_C}}{\nu_k\overline{\rho_L}_0 + \overline{\rho^k_C}} \geq -1$, we have

$$\begin{aligned} \theta(\tilde{\alpha}) &= \frac{1}{\nu_k\overline{\rho_L}_0 + \overline{\rho^k_C}}\left[(\nu_k\overline{\rho_L}_0 + \overline{\rho^k_C})\tilde{\alpha} - \eta\overline{\rho^k_C}\tilde{\alpha} - \tilde{\alpha}^2\left(\overline{L^k_{0,0}} + \eta\overline{\rho^k_C}\,\overline{L^k_{1,0}}\right)\right] \\ &\geq \tilde{\alpha}\left[1 - \eta - \tilde{\alpha}t_1\left(\gamma t_1 n^{3/2} + 2\eta n^{5/2}\right)\right]. \end{aligned}$$

Using the fact that $\tilde{\alpha} \geq \delta_1/n^4$ in the last inequality we have

$$\theta(\tilde{\alpha}) \geq \left[(1-\eta)\delta_1 - \delta_1^2 t_1\left(\gamma t_1 + 2\eta\right)\right]\frac{1}{n^4} = \frac{\hat{\delta}}{n^4}, \tag{60}$$

where

$$\hat{\delta} = (1-\eta)\delta_1 - \delta_1^2 t_1\left(\gamma t_1 + 2\eta\right) = (1-\eta)\delta_1 - \frac{\delta_1^2}{\gamma}\left(t^2 + 2\eta t\right). \tag{61}$$

We need to prove that $0 < \hat{\delta} < 1$. In effect, using Lemma 5 definition of δ_1 in Eq. (61), we derive an expression for $\hat{\delta}$ depending on a choice of γ and η, that is,

$$\hat{\delta} = 2(\gamma\eta)(1-\gamma)\,\frac{\left(4t^2 + \left(4\sqrt{2}t - (6-2\gamma)t^2\right)\eta + \left(1 - (12-4\gamma)t\right)\eta^2 - \eta^3\right)}{\left(4t^2 + 4\sqrt{2}t\eta + \eta^2\right)^2}$$

Note that $0 < (2\gamma\eta)(1-\gamma) < 1$. In addition, the denominator above is greater than the numerator for any value of γ and η. We show now that its numerator is positive. In effect, such numerator is the polynomial

$$p(\eta,\gamma) = 4t^2 + \left(4\sqrt{2}t - (6-2\gamma)t^2\right)\eta + \left(1 - (12-4\gamma)t\right)\eta^2 - \eta^3.$$

Now, $t \geq 3$, since $\varsigma \geq 1$, and $\sqrt{\gamma} < \eta < 1$ since $0 < \gamma \leq 1/10$. Therefore,

$$p(\sqrt{1/10},\gamma) > 0 \text{ and } p(1,\gamma) < 0.$$

Under these conditions, once γ is chosen, p is a cubic polynomial in η. By Bolzano's Intermediate Value Theorem, there is a least one root of p in the open interval $(\sqrt{1/10}, 1)$.

Let η_1 be the smallest of these roots. In this case, by continuity of p, for any $\eta < \eta_1$ and γ chosen as above, $p(\eta, \gamma) > 0$. We conclude that for $0 < \gamma \le 1/10$ and $\sqrt{\gamma} < \eta < \eta_1$, $0 < \hat{\delta} < 1$ whenever $\eta < \eta_1 < 1$.

Finally, using Eq. (60) and from the relationship between φ_{k+1} and φ_k given by Eq. (44) we obtain for this $\tilde{\alpha}$,

$$\varphi_{k+1} = (1 - \theta(\tilde{\alpha}))\, \varphi_k \le \left(1 - \frac{\hat{\delta}}{n^4}\right) \varphi_k. \qquad \square$$

$\square$

Theorem 3 states that relation (45) holds. Therefore, the convergence rate for the sequence $\{\varphi_k\}$ is Q-linear. Using this fact, we can now state and prove the Theorem that gives the order of complexity of the number of iterations.

Theorem 4 (Complexity of Algorithm 1)**.** *Let $0 < \varepsilon < 1$ be given. Suppose that sequence $\{\varphi_k\}$ is generated by Algorithm 1. Then there is an index*

$$K_\varepsilon = \mathcal{O}\left(n^4 \ln \frac{1}{\varepsilon}\right),$$

such that $\varphi_k \le \varepsilon$ for all $k \ge K_\varepsilon$.

Proof. Applying the natural logarithm to both sides of the Inequality (59) given by Theorem 3, we obtain $\ln \varphi_{k+1} \le \ln\left(1 - \frac{\hat{\delta}}{n^4}\right) + \ln \varphi_k$. Using an inductive argument, we have $\ln \varphi_k \le k \ln\left(1 - \frac{\hat{\delta}}{n^4}\right) + \ln \varphi_0$

For $\kappa \ge \dfrac{\ln(1/\varphi_0)}{\ln(\varepsilon)}$ it follows that $\varphi_0 \le \dfrac{1}{\varepsilon^\kappa}$. Thus we obtain

$$\ln \varphi_k \le k \ln\left(1 - \frac{\hat{\delta}}{n^4}\right) + \kappa \ln \frac{1}{\varepsilon}.$$

Since $\ln(1 + r) \le r$ whenever $r > -1$ (see [15, Lemma 4.1, p. 68]), we have $\ln \varphi_k \le k\left(-\frac{\hat{\delta}}{n^4}\right) + \kappa \ln \frac{1}{\varepsilon}$.

In order to prove that the convergence criteria $\varphi_k \le \varepsilon$ holds, we have to guarantee that $k\left(-\frac{\hat{\delta}}{n^4}\right) + \kappa \ln \frac{1}{\varepsilon} \le \ln \varepsilon$. Finally, we have that such inequality holds as long as

$$k \ge \frac{n^4}{\hat{\delta}} (1 + \kappa) \ln \frac{1}{\varepsilon}. \qquad \square$$

$\square$

Theorem 4 implies that, for a sufficiently large k, φ_k is as small as needed. However, such result does not imply that any criteria commonly used in IPM is satisfied. Nevertheless, it is possible to guarantee that for a particular choice of ε, the termination criteria we use in our implementation — see Eq. (64) — is satisfied. In fact, the next Corollary shows that, and it proves that the sequence $\left\{\left\|(r_P^k, r_D^k)\right\|\right\}$ is convergent with an R-linear rate, since this sequence has as upper bound the sequence $\{\varphi_k\}$, which is Q-linear.

Corollary 3. *Suppose we choose a tolerance $\tau > 0$ and a scalar $\varepsilon > 0$ such that*

$$\varepsilon < \min\left\{\frac{\tau(1+\|b\|)}{m+n}, \frac{\tau(1+\|c\|)}{m+n}, \frac{\tau}{n}\right\}. \tag{62}$$

If for a sufficient large k, the sequence $\{\varphi_k\}$ generated by Algorithm 1 is such that $\varphi_k \leq \varepsilon$, then the termination criteria given by Eq. (64) *is satisfied.*

Proof. By Definitions 1 and 2, we have

$$\begin{aligned}\varphi_k = \frac{\left\|\rho_L^k\right\|_1}{m+n} + \frac{(x^k)^T z^k}{n} = \\ = \frac{\left\|S_P(Ax^k - b)\right\|_1}{m+n} + \frac{\left\|S_D(A^T y^k + z^k - c)\right\|_1}{m+n} + \frac{(x^k)^T z^k}{n}.\end{aligned} \tag{63}$$

Moreover, by the equivalence of norms and by the definitions of matrices S_P and S_D, $\left\|Ax^k - b\right\| \leq \left\|Ax^k - b\right\|_1 = \left\|S_P(Ax^k - b)\right\|_1$ and $\left\|A^T y^k + z^k - c\right\| \leq \left\|A^T y^k + z^k - c\right\|_1 = \left\|S_D(A^T y^k + z^k - c)\right\|_1$.

Furthermore, if the sequence $\{\varphi_k\}$ generated by Algorithm 1 converges, there is a sufficiently large k such that $\varphi_k < \varepsilon$. Hence, each of the terms on the last part of Eq. (63) is less than ε.

Thus, from Eq. (62) we obtain

$$\frac{\left\|Ax^k - b\right\|}{1+\|b\|} \leq \frac{\left\|S_P(Ax^k - b)\right\|_1}{1+\|b\|} < \frac{\varepsilon(m+n)}{1+\|b\|} < \frac{\tau(1+\|b\|)}{(m+n)}\frac{(m+n)}{(1+\|b\|)} = \tau,$$

and

$$\frac{\left\|A^T y^k + z^k - c\right\|}{1+\|c\|} \leq \frac{\left\|S_D(A^T y^k + z^k - c)\right\|_1}{1+\|c\|} < \frac{\tau(1+\|c\|)}{(m+n)}\frac{(m+n)}{(1+\|c\|)} = \tau.$$

Moreover, since $1 + |c^T x^k| \geq 1$, we have $\frac{(x^k)^T z^k}{1+|c^T x^k|} < \frac{n \cdot \varepsilon}{1+|c^T x^k|} \leq n \cdot \varepsilon < \tau$. □ □

4 Numerical results

We start this section by showing the framework in which we performed our numerical tests. Then we present how we solve Equation (19) and finally the numerical results are presented.

4.1 Framework to numerical experiments

Algorithm 1 was implemented in C++ and PCx [3], an implementation of Mehrotra's predictor-corrector method, was used as framework and comparison for OCPM. PCx was set to perform Gondzio's [4] multiple corrections and we allowed up to 2 corrections, as well as a solution refinement by conjugate gradient method used on PCx original implementation. We call this reference implementation as PCx-r.

Our implementation of Algorithm 1 is thereby called PCx-OCP. In fact, PCx-OCP and PCx-r share the same data processing, linear algebra routines and compilation flags.

The initial point used in our tests is based on Mehrotra's heuristic [10]. The termination criteria that we will use is the one used in PCx [3], that is

$$\frac{\|Ax-b\|}{1+\|b\|} \leq 10^{-8}, \quad \frac{\left\|A^Ty+z-c\right\|}{1+\|c\|} \leq 10^{-8}, \quad \frac{x^Tz}{1+|c^Tx|} \leq 10^{-8}. \tag{64}$$

In fact, two minor changes on the termination criteria of PCx-r routines were made and we shall stress them. Some IPM solvers, when declare optimality for a point $(\overline{x}, \overline{y}, \overline{z})$, have a finite termination routine that recomputes the dual variable $\overline{z}$, projecting it onto the boundary of the feasible set. Also, PCx-r ensures that $\overline{z}$ has its value recomputed by the dual problem definition and not by the algorithm but only after it is declared optimal. Instead, we implemented this idea at each iteration, when checking the stopping criteria, which means that we project every point generated by PCx onto the boundary of the feasible set and we test if this projection is optimal. If it is optimal, we stop. If it is not optimal, we continue the method with the original iterate. This projection is inexpensive since all the necessary matrix vector operations to perform it have been computed already.

Moreover, we realized that though PCx-r declares optimality in some problems, the reported primal or dual infeasibility for such problems are greater than the tolerance. This happens because the termination criteria is performed on the scaled problem. On the other hand, the report is presented for the unscaled data. Therefore, we computed primal and dual infeasibility — only — for the unscaled data. If, for the unscaled data, the termination criteria are satisfied, then we stop. If they are not, we continue the algorithm with the scaled data. That is also an inexpensive procedure.

In addition, OCPM uses the same steplength for both primal and dual directions, which is implemented in our code. The rationale of this lies on the algebraic complexity that our approach would lead if we used different steplengths. In fact, the subproblem we solve in OCPM, given by g_C^i e h_C in Eq. (19), would depend now on variables α_P and α_D, as well as μ and ω. For any μ and ω fixed, each of $\mathcal{O}(n)$ inequalities is a region on the $\mathbb{R}^2$ plane given by one hyperbole with axis parallels to axis α_P and α_D. To satisfy these inequalities, we would need to compute $\mathcal{O}(n)$ intersections of these regions. This would lead us to a nonconvex region, which could have up to $\mathcal{O}(n)$ vertexes. From both algebraic and numerical point of view, finding these intersections would be a huge task.

On the other hand, Mehrotra predictor-corrector method implemented in PCx-r — as well as most of predictor-corrector IPM — uses different steplengths for primal and dual directions. Thus, we also compare our algorithm with a version of PCx that uses the same steplength for both primal and dual directions, which we call PCx-s.

The test set that we use is taken from the NETLIB. We chose all 95 feasible LP problems from it. Moreover, from the same repository, we chose 12 of 16 Kennington problems. Also, we chose 1 — `qap-8` — of 3 QAP problems. Such set is called here NETLIB-108. The 4 Kennington problems and the 2 QAP problems that we did not choose have sizes and structures incompatible with an implementation that uses Cholesky factorization to solve the normal equations, like ours.

Computational tests were performed on an Intel i7 4790 3877MHz with 32GB of RAM in a Windows 10 64bits and compiled on Microsoft Visual Studio C++ 2015 with Intel C++ Parallel Studio 2015.

4.2 Results

Table 1 shows the total number of iterations and CPU time. In PCx-OCP, CPU time includes the time spent on solving subproblem (19), with the method explained in Section 2.3. An $*$ symbol after the number of iterations of the problem indicates that optimality was not reached for the respective method.

Note that PCx-r does not solve 3 problems: `brandy`, `greenbea` and `scfxm2` while PCx-s does no solve 1 problem: greenbea. Meanwhile PCx-OCP does not solve 5 of them: `bnl1`, `fit1p`, `fit2p`, `greenbea` and `pilot4`. Only problem `greenbea` was not solved by the three implementations. Thus, we achieve a very good robustness for PCx-OCP, solving 95% of the problems from the test set. The fact that PCx-s, using the same steplength for both primal and dual directions is more robust than PCx-r is a very interesting result.

The results show that PCx-OCP solved the NETLIB-108 in 1m 41s (101.451 s) while PCx-r solved the same set in 56s (55.520 s) and PCx-s used 1m 12s (71.634 s), with comparable number of iterations. In terms of CPU time, the implementations have the same order of magnitude. This is the most important parameter for us, because CPU time validates our approach as a *proof-of-concept*. Moreover, PCx-OCP is the implementation of an algorithm that has proof of convergence, and, to our knowledge, this is not valid for PCx.

One of the reasons of the numerical results is that OCPM is a *greedy* type of method, in the sense that it looks for the minimization of the residuals average, taking into account the next iteration. Note, however, that any information from forward iterations depends, on each iteration, on one Cholesky factorization, which means that it is computationally expensive/costly.

Table 1: Comparison between PCx-r and PCx-OCP.

	PCx-r		**PCx-s**		**PCx-OCP**	
Problem	**It.**	**CPU (s)**	**It.**	**CPU (s)**	**It.**	**CPU (s)**
`25fv47`	24	0.200	26	0.244	24	0.458
`80bau3b`	39	0.447	43	0.810	38	0.920

Continues on the next page.

Table 1: Comparison between PCx-r and PCx-OCP (continuation).

	PCx-r		PCx-s		PCx-OCP	
Problem	**It.**	**CPU (s)**	**It.**	**CPU (s)**	**It.**	**CPU (s)**
adlittle	10	0.072	11	0.093	14	0.324
afiro	6	0.047	6	0.050	8	0.252
agg	17	0.123	18	0.140	21	0.450
agg2	20	0.152	21	0.182	24	0.547
agg3	19	0.152	21	0.191	21	0.458
bandm	15	0.105	16	0.127	20	0.435
beaconfd	9	0.064	9	0.071	13	0.322
blend	8	0.060	8	0.065	10	0.214
bnl1	33	0.252	32	0.271	54*	1.180
bnl2	33	0.405	35	0.577	29	0.701
boeing1	18	0.141	19	0.174	23	0.559
boeing2	12	0.083	14	0.122	19	0.368
bore3d	14	0.103	15	0.131	22	0.438
brandy	17*	0.148	16	0.129	20	0.391
capri	17	0.120	19	0.174	23	0.515
cycle	22	0.218	25	0.292	20	0.486
czprob	25	0.180	26	0.220	28	0.597
d2q06c	27	0.422	29	0.502	29	0.921
d6cube	20	0.257	19	0.263	21	0.667
degen2	10	0.113	11	0.133	11	0.245
degen3	23	0.469	14	0.311	22	0.693
dfl001	55	17.146	52	21.204	48	16.133
e226	17	0.128	17	0.153	21	0.418
etamacro	27	0.186	27	0.228	31	0.613
fffff800	29	0.212	29	0.257	35	0.779
finnis	22	0.153	23	0.197	28	0.558
fit1d	16	0.127	18	0.151	20	0.434
fit1p	16	0.210	16	0.201	19*	0.500
fit2d	22	0.287	23	0.362	32	0.980
fit2p	21	0.428	22	0.456	24*	0.897
forplan	20	0.142	22	0.194	25	0.482
ganges	15	0.123	16	0.154	16	0.360
gfrd-pnc	15	0.106	17	0.153	17	0.410
greenbea	49*	0.504	63*	0.712	31*	0.716
greenbeb	37	0.401	39	0.497	35	0.793
grow7	15	0.102	16	0.136	18	0.393
grow15	18	0.131	19	0.167	20	0.447
grow22	20	0.152	22	0.187	25	0.521
israel	19	0.135	20	0.167	24	0.477
kb2	10	0.079	12	0.110	19	0.387
lotfi	12	0.081	13	0.124	17	0.405
maros-r7	16	0.741	18	0.927	19	1.150
maros	17	0.131	19	0.169	21	0.442
modszk1	19	0.155	23	0.227	22	0.459
nesm	28	0.216	30	0.265	33	0.675
perold	32	0.238	39	0.345	40	0.865
pilot.ja	31	0.283	42	0.434	41	0.989

Continues on the next page.

Table 1: Comparison between PCx-r and PCx-OCP (continuation).

	PCx-r		PCx-s		PCx-OCP	
Problem	**It.**	**CPU (s)**	**It.**	**CPU (s)**	**It.**	**CPU (s)**
pilot.we	44	0.331	62	0.527	49	0.952
pilot	37	0.809	43	1.065	34	1.269
pilot4	49	0.362	57	0.496	41*	0.964
pilot87	33	1.973	36	2.180	37	2.512
pilotnov	16	0.147	16	0.165	17	0.483
recipe	7	0.052	8	0.076	10	0.268
sc50a	6	0.049	6	0.058	9	0.198
sc50b	5	0.041	5	0.052	8	0.192
sc105	8	0.060	9	0.080	10	0.218
sc205	9	0.072	10	0.082	15	0.311
scagr7	12	0.081	14	0.112	15	0.282
scagr25	16	0.124	17	0.146	22	0.418
scfxm1	16	0.119	17	0.144	18	0.379
scfxm2	19*	0.166	19	0.156	20	0.401
scfxm3	18	0.137	20	0.177	23	0.495
scorpion	10	0.123	10	0.140	12	0.307
scrs8	21	0.155	23	0.187	27	0.572
scsd1	8	0.068	8	0.068	9	0.213
scsd6	10	0.080	13	0.105	14	0.323
scsd8	10	0.073	11	0.105	14	0.277
sctap1	13	0.091	14	0.115	21	0.446
sctap2	12	0.104	13	0.119	17	0.375
sctap3	13	0.101	14	0.126	17	0.367
seba	12	0.128	13	0.174	14	0.386
share1b	17	0.116	18	0.145	28	0.545
share2b	16	0.120	17	0.140	22	0.424
shell	20	0.179	21	0.215	24	0.506
ship04l	11	0.081	12	0.095	16	0.335
ship04s	11	0.090	13	0.106	20	0.410
ship08l	13	0.097	15	0.134	17	0.371
ship08s	10	0.081	11	0.089	14	0.311
ship12l	14	0.106	17	0.154	20	0.434
ship12s	12	0.100	12	0.116	13	0.318
sierra	18	0.208	19	0.253	21	0.521
stair	13	0.102	14	0.132	16	0.357
standata	12	0.087	12	0.098	13	0.291
standgub	12	0.093	12	0.094	13	0.280
standmps	22	0.154	27	0.218	29	0.593
stocfor1	10	0.072	11	0.101	13	0.258
stocfor2	19	0.149	20	0.193	24	0.542
stocfor3	30	0.481	33	0.587	35	1.104
truss	17	0.159	20	0.223	20	0.465
tuff	17	0.126	19	0.157	24	0.471
vtp-base	9	0.064	9	0.078	12	0.276
wood1p	22	0.301	24	0.311	22	0.490
woodw	29	0.248	33	0.334	33	0.714
cre-a	25	0.320	27	0.390	24	0.617

Continues on the next page.

Table 1: Comparison between PCx-r and PCx-OCP (continuation).

	PCx-r		PCx-s		PCx-OCP	
Problem	**It.**	**CPU (s)**	**It.**	**CPU (s)**	**It.**	**CPU (s)**
cre-b	43	1.315	46	1.615	42	2.057
cre-c	25	0.318	27	0.393	25	0.629
cre-d	41	1.107	48	1.497	40	1.767
ken-07	13	0.154	14	0.205	16	0.445
ken-11	20	0.381	23	0.478	20	0.762
osa-07	21	0.296	29	0.430	50	1.399
osa-14	24	0.561	34	0.922	55	2.363
osa-30	24	0.942	32	1.446	73	4.997
pds-02	24	0.316	30	0.461	31	0.830
pds-06	35	2.302	41	3.302	41	3.855
pds-10	44	11.714	51	16.182	50	15.268
qap8	7	0.204	7	0.236	10	0.414
Totals	2140	55.520	2366	71.634	2594	101.451

5 Final Remarks

In this work we tried to answer some of the actual issues in IPMs on both theoretical and numerical views by proposing the Optimal Choice Parameters Method (OCPM) that relies on an optimal choice, at each iteration, of parameters (α, μ, ω), which allows us to take the steps. We have shown that OCPM is polynomial with $\mathcal{O}(n^4)$ iterations and that it has Q-linear convergence rate.

Although, there are algorithms with better convergence complexity for infeasible IPMs [16, 17], to our knowledge these methods were not implemented. In our case, the analysis is performed on the same algorithm that was actually implemented.

This way, we have a reasonable theoretical result with an algorithm that works well in practice. In fact, in our view, the above numerical results show that OCPM is robust enough and has CPU time and iteration count comparable with PCx, a mature and tested implementation of Mehrotra's algorithm.

References

[1] Berti, L.F., Oliveira, A.R.L., Ghidini, C.T.L.S.: A variation on the interior point method for linear programming using the continued iteration. Math. Methods Oper. Res. **85**(1), 1–15 (2016)

[2] Colombo, M., Gondzio, J.: Further development of multiple centrality correctors for interior point methods. Comput. Optim. Appl. **41**(3), 277–305 (2008)

[3] Czyzyk, J., Mehrotra, S., Wagner, M., Wright, S.J.: PCx: an interior-point code for linear programming. Optim. Methods Softw. **11**(1), 397–430 (1999)

[4] Gondzio, J.: Multiple centrality corrections in a primal-dual method for linear programming. Comput. Optim. Appl. **6**(2), 137–156 (1996)

[5] Gondzio, J.: Interior point methods 25 years later. European J. Oper. Res. **218**(3), 587–601 (2012)

[6] Herbison-Evans, D.: Solving Quartics and Cubics for Graphics. In: A.W. Paeth (ed.) Graphics Gems V, pp. 3–15. Academic Press, Boston (1995)

[7] Jarre, F., Wechs, M.: Extending Mehrotra's corrector for linear programs. Adv. Model. Optim. **1**(2), 38–60 (1999)

[8] Kojima, M., Mizuno, S.: A primal-dual infeasible-interior-point algorithm for linear programming. Math. Program. **61**(3), 263–280 (1993)

[9] Lasserre, J.B.: Global Optimization with Polynomials and the Problem of Moments. SIAM J. Optim. **11**(3), 796–817 (2001)

[10] Mehrotra, S.: On the Implementation of a Primal-Dual Interior Point Method. SIAM J. Optim. **2**(4), 575–601 (1992)

[11] Mehrotra, S., Li, Z.: Convergence Conditions and Krylov Subspace–Based Corrections for Primal-Dual Interior-Point Method. SIAM J. Optim. **15**(3), 635–653 (2005)

[12] Monteiro, R.D.C., Adler, I., Resende, M.G.C.: A polynomial-time primal-dual affine scaling algorithm for linear and convex quadratic programming and its power series extension. Math. Oper. Res. **15**(2), 191–214 (1990)

[13] Villas-Bôas, F.R., Perin, C.: Postponing the choice of penalty parameter and step length. Comput. Optim. Appl. **24**(1), 63–81 (2003)

[14] Wright, S.J.: An infeasible-interior-point algorithm for linear complementarity problems. Math. Program. **67**(1-3), 29–51 (1994)

[15] Wright, S.J.: Primal-dual interior point methods, 1 edn. SIAM, Philadelphia, PA (1997)

[16] Wright, S.J., Zhang, Y.: A superquadratic infeasible-interior-point method for linear complementarity problems. Math. Program. **73**(3), 269–289 (1996)

[17] Zhang, Y.: On the convergence of a class of infeasible interior-point methods for the horizontal linear complementarity problem. SIAM J. Optim. **4**(1), 208–227 (1994)

[18] Zhang, Y., Zhang, D.: Superlinear convergence of infeasible-interior-point methods for linear programming. Math. Program. **66**(1-3), 361–377 (1994)

[19] Zhang, Y., Zhang, D.: On polynomiality of the Mehrotra-type predictor-corrector interior-point algorithms. Math. Program. **68**(3), 303–318 (1995)