Optimized Choice of Parameters in interior-point methods for linear programming

Luiz-Rafael dos Santos

Joint work with F. Villas-Bôas, A. Oliveira and C. Perin

1st Brazilian Workshop on Interior Point Methods

27 Apr 2015

Outline

1. Motivations
2. Optimized Choice of Parameters Method
3. OCMP convergence analysis
4. Numerical Results

Outline

1 Motivations

2 Optimized Choice of Parameters Method

3 OCMP convergence analysis

4 Numerical Results

Some challenges in IPM

"For interior-point methods, can we give a theoretical explanation for the difference between worst-case bounds and observed practical performance? Can we devise an algorithm whose iteration complexity is better than $\mathcal{O}(\sqrt{n}\ln(1/\varepsilon))$ *to attain* ε*-optimality?" Todd [9]*

Some challenges in IPM

"For interior-point methods, can we give a theoretical explanation for the difference between worst-case bounds and observed practical performance? Can we devise an algorithm whose iteration complexity is better than $\mathcal{O}(\sqrt{n}\ln(1/\varepsilon))$ *to attain* ε*-optimality?" Todd [9]*

- It is possible to implement an IPM that has good practical results and that it has, at the same time, reasonable complexity and convergence proprieties?
- How to combine predictor, corrector or other high order directions to obtain a better direction?
  - Combine directions efficiently.
  - There is no panacea for all problems
- How to keep iterates under "good conditions"?
  - Central path neighborhoods, heuristics, etc.

Background

Colombo and Gondzio [1] and Gondzio [3]: extend Mehrotra primal-dual corrector idea, allowing multiple corrections at the same iterate, to enlarge the step length;

Jarre and Wechs [4]: solve a small LP – using simplex – to combine directions;

Mehrotra and Li [7]: generate predictor and corrector directions using a Krylov subspace search;

Villas-Bôas and Perin [11]: Postpone the barrier parameter and step length choice by solving a polynomial optimization problem, on a self-dual context.

What we have done I

1. **Developed** an Infeasible IPM for LP.
   - Using a merit function that depends on the parameters (α, μ, σ) where: α is the step length; μ defines a central path; σ represents the 2nd order the corrector direction weight.
   - How we choose them?
     - Minimize a predictive polynomial merit function;
     - constrained to a neighborhood of the central path;
     - ensure that the iterate pass the ratio test.
   - Our merit function is assembled using the residuals of both linear and complementarity parts of the LP.
   - We called it Optimized Choice of Parameters Method (OCPM)
2. We **proved** OCPM convergence results
3. We **established** an Assumption that the initial point has to meet in order to assure the convergence.
4. We **implemented** and **tested** OCPM on NETLIB and compared with PCx [2].

Outline

1 Motivations

2 Optimized Choice of Parameters Method

3 OCMP convergence analysis

4 Numerical Results

Problem Formulation

- Linear programming primal and dual problems are defined as

$$\min_{x} \quad c^T x \quad \text{s.t.} \quad \begin{cases} Ax = b \\ x \geq 0 \end{cases} \qquad \text{(Primal)}$$

$$\max_{(y,z)} \quad b^T y \quad \text{s.t.} \quad \begin{cases} A^T y + z = b \\ z \geq 0, \ y \text{ free} \end{cases} \qquad \text{(Dual)}$$

$A \in \mathbb{R}^{m \times n}$, $m \leq n$ is full-rank, $c, x, z \in \mathbb{R}^n$ and $y, b \in \mathbb{R}^m$.

- KKT conditions:

$$\begin{cases} Ax = b, \\ A^T y + z = c, \\ XZe = 0, \\ (x, z) \geq 0, \end{cases} \qquad \text{(KKT)}$$

where $X = \text{diag}(x)$, $Z = \text{diag}(z)$ and $e = (1, \ldots, 1)^T$.

- This formulation is valid for Bounded LP under some transformations, including the implementation.

Approach

Central-path method, to solve a Scaled KKT system, solving

$$\begin{cases} H_P(Ax - b) = 0, \\ H_D(A^T y + z - c) = 0, \\ XZe = \mu e, \\ (x, z) > 0, \end{cases} \qquad \text{(Scaled KKT)}$$

for $\mu \geq 0$.

Approach

Central-path method, to solve a Scaled KKT system, solving

$$\begin{cases} H_P(Ax - b) = 0, \\ H_D(A^T y + z - c) = 0, \\ XZe = \mu e, \\ (x, z) > 0, \end{cases} \qquad \text{(Scaled KKT)}$$

for $\mu \geq 0$.

Search Directions

- Affine-scale or pure-Newton direction: $(\Delta x^{\mathsf{af}}, \Delta y^{\mathsf{af}}, \Delta z^{\mathsf{af}})$.
- Ideal direction: $\Delta w = (\Delta x, \Delta y, \Delta z)$, such that $\hat{w} = (\hat{x}, \hat{y}, \hat{z}) = w + \Delta w$, is the solution of

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

- We set $\Delta w = \Delta w^{\mathsf{af}} + \Delta w^{\mathsf{c}}$, where Δw^{c} is an ideal corrector direction.
- With some simplifications we obtain the nonlinear system

$$\begin{cases} A\Delta x^c = 0 \\ A^T\Delta y^c + \Delta z^c = 0 \\ X\Delta z^c + Z\Delta x^c + \Delta X \Delta z = \mu e \end{cases}$$

  - $\Delta X \Delta z$ is a 2nd order direction similar to the ones used by Gondzio [3] and Mehrotra [6].

Search directions

Generalizing some methods

Intuition: Weight correction

- For $\sigma \geq 0$ bounded, suppose one can use the approximation

$$\Delta X \Delta z \approx \sigma \Delta X^{\text{af}} \Delta z^{\text{af}}.$$

Nonlinear system above transformed onto the linear system

$$\begin{cases} A\Delta x^c = 0 \\ A^T \Delta y^c + \Delta z^c = 0 \\ X\Delta z^c + Z\Delta x^c + \sigma \Delta X^{\text{af}} \Delta z^{\text{af}} = \mu e \end{cases}$$

- If one sets $\sigma = 1$ and $\mu = ((x^{\text{af}})^T (z^{\text{af}})/n)^3/(x^T z/n)$, we have Mehrotra's method.
- In Gondzio's multiple centrality method, $\Delta X \Delta z$ is several times approximated by projections on a central path neighborhood.
- If $\mu = 0$ ad $\sigma = 1$ (feasible point) we have Monteiro, Adler, and Resende's method.

Search directions

- Let Δw^c be divided as

$$\Delta w^{\mathsf{c}} = \mu \Delta w^{\mu} + \sigma \Delta w^{\sigma}.$$

- We write the next point as

$$\hat{x} = x + \alpha(\Delta x^{\mathsf{af}} + \mu \Delta x^{\mu} + \sigma \Delta x^{\sigma})$$
$$\hat{y} = y + \alpha(\Delta y^{\mathsf{af}} + \mu \Delta y^{\mu} + \sigma \Delta y^{\sigma})$$
$$\hat{z} = z + \alpha(\Delta z^{\mathsf{af}} + \mu \Delta z^{\mu} + \sigma \Delta z^{\sigma})$$

(α, μ, σ) yet to be selected

- (α, μ, σ) is considered as a real variable triplet.
- Choose this parameters-variables using at most 3 back-solves.
- Use a merit function that takes into account the KKT.
- Finally, use the above linear combination of directions Δw^{af}, Δw^{μ} and Δw^{σ}, where (α, μ, σ) are the combination constants

Scaled KKT residuals

Definition

Let ρ be the *Scaled KKT residuals vector* for a point (x, y, z), given by

$$\rho(x,y,z) = \begin{cases} \rho_P(x,y,z) = H_P(Ax - b) \\ \rho_D(x,y,z) = H_D(A^T y + z - c) \\ \rho_C(x,y,z) = XZe \end{cases}$$

Definition (Merit function)

We define the *merit function* of a point (x, y, z) as

$$\varphi(x,y,z) = \frac{x^T z}{n} + \frac{1}{m+n} \|\rho_L\|_1 ,$$

where ρ_L e ρ_C are the Scaled KKT residuals at point (x, y, z).

Scaled KKT residuals

Definition

Let ρ be the *Scaled KKT residuals vector* for a point (x, y, z), given by

$$\rho(x,y,z) = \begin{cases} \rho_P(x,y,z) = H_P(Ax - b) \\ \rho_D(x,y,z) = H_D(A^T y + z - c) \\ \rho_C(x,y,z) = XZe \end{cases}$$

Definition (Merit function)

We define the *merit function* of a point (x, y, z) as

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

where ρ_L e ρ_C are the Scaled KKT residuals at point (x, y, z).

Predicting the next merit

- It is possible to predict the merit function for the next iterate $(\hat{x}, \hat{y}, \hat{z})$?

Predicting the next merit

- It is possible to predict the merit function for the next iterate $(\hat{x}, \hat{y}, \hat{z})$?

Definition (Next Merit)

The next merit function value is given by

$$\hat{\varphi}(x^k, y^k, z^k) = \overline{\hat{\rho}_L}(x^k, y^k, z^k) + \overline{\hat{\rho}_C}(x^k, y^k, z^k).$$

It follows from the next residuals definition that

$$\hat{\varphi}(\alpha, \mu, \sigma) = \overline{\hat{\rho}_L}(\alpha, \mu, \sigma) + \overline{\hat{\rho}_C}(\alpha, \mu, \sigma)$$

Predicting the next merit

- It is possible to predict the merit function for the next iterate $(\hat{x}, \hat{y}, \hat{z})$?

Definition (Next Merit)

The next merit function value is given by

$$\hat{\varphi}(x^k, y^k, z^k) = \overline{\hat{\rho}_L}(x^k, y^k, z^k) + \overline{\hat{\rho}_C}(x^k, y^k, z^k).$$

It follows from the next residuals definition that

$$\hat{\varphi}(\alpha, \mu, \sigma) = \overline{\hat{\rho}_L}(\alpha, \mu, \sigma) + \overline{\hat{\rho}_C}(\alpha, \mu, \sigma)$$

Theorem (Predictive Merit Function)

The predictive merit function can be expressed as the following polynomials on variables (α, μ, σ).

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

Predicting the next merit

- It is possible to predict the merit function for the next iterate $(\hat{x}, \hat{y}, \hat{z})$?

Definition (Next Merit)

The next merit function value is given by

$$\hat{\varphi}(x^k, y^k, z^k) = \overline{\hat{\rho}_L}(x^k, y^k, z^k) + \overline{\hat{\rho}_C}(x^k, y^k, z^k).$$

It follows from the next residuals definition that

$$\hat{\varphi}(\alpha, \mu, \sigma) = \overline{\hat{\rho}_L}(\alpha, \mu, \sigma) + \overline{\hat{\rho}_C}(\alpha, \mu, \sigma)$$

Theorem (Predictive Merit Function)

The predictive merit function can be expressed as the following polynomials on variables (α, μ, σ).

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

- Polynomial of total degree 4 in (α, μ, σ):

Central Path Neighborhood as constraints

Given $\gamma \in (0,1)$ and $\beta \geq 1$, the central path infeasible neighborhood from Kojima, Megiddo, and Mizuno [5] is

$$\mathcal{N}_{-\infty}(\gamma,\beta) = \left\{ (x,y,z) \in \mathcal{Q}^{+} : \frac{\|r_L\|}{\|r_L^0\|} \leq \beta \frac{x^T z}{(x^0)^T z^0} \text{ and } x_i z_i \geq \gamma \frac{x^T z}{n}, \forall i = 1,\dots,n \right\}.$$

Central Path Neighborhood as constraints

Given $\gamma \in (0, 1)$ and $\beta \geq 1$, the central path infeasible neighborhood from Kojima, Megiddo, and Mizuno [5] is

$$\mathcal{N}_{-\infty}(\gamma, \beta) = \left\{ (x, y, z) \in \mathcal{Q}^+ : \frac{\|r_L\|}{\|r_L^0\|} \leq \beta \frac{x^T z}{(x^0)^T z^0} \text{ and } x_i z_i \geq \gamma \frac{x^T z}{n}, \forall i = 1, \ldots, n \right\}.$$

Using our notation we get

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

Finding the actual direction

Polynomial Optimization Subproblem

- Find (α, μ, σ) such as the predictive merit function is minimized as long as the next point is constrained to $\mathcal{N}_{-\infty}(\gamma, \beta)$, i.e.,

$$\begin{aligned} &\min_{(\alpha,\mu,\sigma)} \hat{\varphi}(\alpha, \mu, \sigma) \\ &\text{s. a.} \quad (\hat{x}, \hat{y}, \hat{z}) \in \mathcal{N}_{-\infty}(\gamma, \beta) \end{aligned}$$

Finding the actual direction

Polynomial Optimization Subproblem

- Find (α, μ, σ) such as the predictive merit function is minimized as long as the next point is constrained to $\mathcal{N}_{-\infty}(\gamma, \beta)$, i.e.,

$$\min_{(\alpha,\mu,\sigma)} \quad \hat{\varphi}(\alpha,\mu,\sigma)$$
$$\text{s. a.} \quad \begin{cases} g_C^i(\alpha,\mu,\sigma) \geq 0 \quad \forall i = 1,\dots,n \\ g_L(\alpha,\mu,\sigma) \geq 0 \\ 0 \leq (\alpha,\mu,\sigma) \leq u, \end{cases} \tag{SOP}$$

where $u \in \mathbb{R}^3$ is a vector of bounds for (α, μ, σ).

Finding the actual direction

Polynomial Optimization Subproblem

- Find (α, μ, σ) such as the predictive merit function is minimized as long as the next point is constrained to $\mathcal{N}_{-\infty}(\gamma, \beta)$, i.e.,

$$\begin{array}{ll} \min\limits_{(\alpha,\mu,\sigma)} & \hat{\varphi}(\alpha,\mu,\sigma) \\ \text{s. a.} & \begin{cases} g_C^i(\alpha,\mu,\sigma) \geq 0 \quad \forall i = 1,\ldots,n \\ g_L(\alpha,\mu,\sigma) \geq 0 \\ 0 \leq (\alpha,\mu,\sigma) \leq u, \end{cases} \end{array} \tag{SOP}$$

  where $u \in \mathbb{R}^3$ is a vector of bounds for (α, μ, σ).

- Global optimization of a polynomial constrained to a set of $n+1$ polynomials and bounds 0 and u.
- $\hat{\varphi}, g_L$ and g_C^i are polynomials with up to 6 total degree on variables (α, μ, σ).
- Ratio test is performed

Outline

1. Motivations
2. Optimized Choice of Parameters Method
3. OCMP convergence analysis
4. Numerical Results

Convergence and Polynomiality

Analysis showed that:

- OCPM is well defined: One can always find a triplet (α, μ, σ);
- There is a triplet that decreases sufficiently the merit function.

Convergence and Polynomiality

Analysis showed that:

- OCPM is well defined: One can always find a triplet (α, μ, σ);
- There is a triplet that decreases sufficiently the merit function.

Important Consequences

OCPM

- converges to a LP solution with Q-linear rate;
- has at most $\mathcal{O}(n^4)$ iterations – it is polynomial.

Initial Point

- Most common infeasible complexity analysis use the following (or equivalent) assumption:

Assumption

Let

$$\vartheta^* = \min \left\{ \left\| (x^*, z^*) \right\| : (x^*, y^*, z^*) \in \mathcal{F}^* \right\}$$

and

$$\vartheta \geq \left\| (\breve{x}, \breve{z}) \right\|,$$

where $(\breve{x}, \breve{y}, \breve{z})$ *is the least square solution for* $Ax = b$ *and* $A^T y + z = c$*.*
Then

$$\vartheta \geq \vartheta^* / \sqrt{n}.$$

- Under this assumption, authors define

$$(x^0, y^0, z^0) = (\vartheta e, 0, \vartheta e).$$

- This initial point allows polynomial complexity proprieties, however generates poor numerical performance.
- Issue: One need to know *a priori* a bound for an optimal solution.

Out initial point assumption

Assumption

For an interior (x^0, y^0, z^0), *there is a LP optimal solution* (x^*, y^*, z^*) *such that*

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

where $\varsigma \geq 1$ *is given by*

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

- $\varsigma \geq 1$ for any scaled problem.
- Theoretical assumption, not used in OCPM implementation
- All problems in NETLIBsatisfies it, with Mehrotra initial point heuristics

Outline

1 Motivations

2 Optimized Choice of Parameters Method

3 OCMP convergence analysis

4 Numerical Results

Some details

- Implemented in `C++`
- Using PCx framework, Mehrotra's PC method with Gondzio's corrections (PCx-r)
- PCx-OCP is our OCPM implementation.
  - PCx-OCP inherits from PCx-r all linear algebra, initial point and stop criteria routines.
- Same compilation flags
- Source code adapted from Villas-Bôas et al. [10].

Numerical Results

CUTEr-NETLIB-108

- Selected from Neltib (CUTer):
  - 95 feasible LP
  - 12 of 16 Kennington problems
  - Only `qap-8` – from 3 QAP problems
- 4 Kennington LP and 2 QAP LP were kept out because of size (We use Cholesky factorization)
  - Part of NETLIB-108 is used by Colombo and Gondzio [1], Gondzio [3], Jarre and Wechs [4], Mehrotra [6], and Mehrotra and Li [7], as well as by PCx original tests.

Numerical Results

- Robustness of PCx-OCP
  - PCx-r didn't solve 3 LP: `brandy`, `greenbea` e `scfxm2`
  - PCx-OCP didn't solve 5 LP: `bnl1`, `fit1p`, `fit2p`, `greenbea`, `pilot4`.
- Total CPU time and iteration number are comparable

Numerical Results

- Robustness of PCx-OCP
  - PCx-r didn't solve 3 LP: `brandy`, `greenbea` e `scfxm2`
  - PCx-OCP didn't solve 5 LP: `bnl1`, `fit1p`, `fit2p`, `greenbea`, `pilot4`.
- Total CPU time and iteration number are comparable

Numerical Results

- Robustness of PCx-OCP
  - PCx-r didn't solve 3 LP: `brandy`, `greenbea` e `scfxm2`
  - PCx-OCP didn't solve 5 LP: `bnl1`, `fit1p`, `fit2p`, `greenbea`, `pilot4`.
- Total CPU time and iteration number are comparable

PCx-r:	1min 55s
PCx-OCP:	2min 36s

Numerical Results

- Robustness of PCx-OCP
  - PCx-r didn't solve 3 LP: `brandy`, `greenbea` e `scfxm2`
  - PCx-OCP didn't solve 5 LP: `bnl1`, `fit1p`, `fit2p`, `greenbea`, `pilot4`.
- Total CPU time and iteration number are comparable

  PCx-r: 1min 55s

  PCx-OCP: 2min 36s
- CPU time, for us, validates our approach as a proof of concept

Thank you!

lrsantos11@gmail.com
l.r.santos@ufsc.br

References I

M. Colombo and J. Gondzio. "Further development of multiple centrality correctors for interior point methods". In: *Computational Optimization and Applications* 41.3 (2008), pp. 277–305.

J. Czyzyk, S. Mehrotra, M. Wagner, and S. J. Wright. "PCx: an interior-point code for linear programming". In: *Optimization Methods and Software* 11.1 (1999), pp. 397–430.

J. Gondzio. "Multiple centrality corrections in a primal-dual method for linear programming". In: *Computational Optimization and Applications* 6.2 (1996), pp. 137–156.

F. Jarre and M. Wechs. "Extending Mehrotra's corrector for linear programs". In: *Advanced Modeling and Optimization* 1.2 (1999), pp. 38–60.

References II

- M. Kojima, N. Megiddo, and S. Mizuno. “A primal-dual infeasible-interior-point algorithm for linear programming”. In: *Mathematical Programming* 61.3 (1993), pp. 263–280.
- S. Mehrotra. “On the Implementation of a Primal-Dual Interior Point Method”. In: *SIAM Journal on Optimization* 2.4 (1992), pp. 575–601.
- S. Mehrotra and Z. Li. “Convergence Conditions and Krylov Subspace–Based Corrections for Primal-Dual Interior-Point Method”. In: *SIAM Journal on Optimization* 15.3 (2005), pp. 635–653.
- R. D. C. Monteiro, I. Adler, and M. G. C. Resende. “A polynomial-time primal-dual affine scaling algorithm for linear and convex quadratic programming and its power series extension”. In: *Mathematics of Operations Research* 15.2 (1990), pp. 191–214.
- M. J. Todd. “The many facets of linear programming”. In: *Mathematical Programming* 91.3 (2002), pp. 417–436.

References III

F. R. Villas-Bôas, A. R. L. Oliveira, C. Perin, and L.-R. Santos. *Predictive Polynomials in Interior Point Methods*. Tech. rep. IMECC/Unicamp, 2013.

F. R. Villas-Bôas and C. Perin. "Postponing the choice of penalty parameter and step length". In: *Computational Optimization and Applications* 24.1 (2003), pp. 63–81.