2nd International Conference on Engineering Optimization
September 6-9, 2010, Lisbon, Portugal

On the asymptotic order in path following interior point methods

Jean-Pierre Dussault

June 4, 2010

Abstract

Interior point methods are related to Newton's method applied to a parametric KKT system. The parameter is driven to zero thereby ensuring convergence of the approximate solutions of the parametric system. Path following variants explicitly define and follow the parametric solutions, the so called central path. In this talk, we address two generalizations of the usual interior point framework from an asymptotic analysis point of view.

First, we address inexact solutions of the linear systems arising within Newton's method. Depending on the inexactness allowed, asymptotic convergence order reduces; for primal variants, this reduction is more severe than for primal-dual formulations.

Second, we investigate the use of higher order path following strategies in those methods. The usual high order path following variants (Mehrotra's formulation) are reminiscent of Shamanskii's generalization of Newton's method. We also consider the approach based on a high order expansion of the so-called central path, somewhat reminiscent of Chebyshev's third order method and its generalizations. For primal variants, the use of higher order representation of the path yields spectacular improvement in the convergence order while for primal-dual formulations, the benefits are important, but not surprising.

Introduction

We consider non linear programs (NLP) of the form

$$\begin{aligned} &\min_{x\in\mathbb{R}^n} \ f(x) \\ &\text{subject to } g(x) \leq 0 \end{aligned} \tag{1}$$

with $g : \mathbb{R}^n \to \mathbb{R}^m$ which may be reformulated as

$$\begin{aligned} &\min_{x,s\in\mathbb{R}^n\times\mathbb{R}^m} \ f(x) \\ &\text{subject to } g(x) \ + \quad s = 0 \\ &\qquad\qquad\qquad\quad\ \ s \geq 0 \end{aligned} \tag{2}$$

We will address the formulation Eq.(1) using the classical log barrier method in the SUMT spirit while the formulation Eq.(2) will be developed using a primal-dual infeasible algorithm in the spirit of interior point methods for linear programming.

Fiacco and McCormick pioneered the study of the log barrier, and obtained the important result that close to a solution, the barrier sub-problems induce differential trajectories, and proposed the use of extrapolation to follow the trajectories. In linear programming, Mehrotra popularized the use of so called predictor-corrector algorithms, intimately related to the extrapolations of the log barrier trajectory. For linear programming, many variant were studied, involving higher order predictor steps as well as multiple corrector steps.

We present asymptotic results related to inexact variants of both approaches, looking for super-linear convergence of order lower than 2, as well as high order extrapolations to aim for faster asymptotic convergence.

1 Path following methods

We now recall two path following approaches, allowing to settle our notation and present some basic properties. Path following methods are related to the so-called central path, and involves steps named as predictor, corrector, centrality corrector, higher order predictor. We will detail below the terms we will use.

1.1 Primal – SUMT

The log barrier approach to solve Eq.(1) consists in solving a sequence of sub-problems of the form

$$\min_{g(x)<0} \phi(x, \rho_k) = f(x) - \rho_k \sum \log(-g_i(x)) \tag{3}$$

in the interior of the feasible set $E = \{x : g(x) \leq 0\}$. Writing out the optimality conditions

$$\nabla_x \phi(x, \rho_k) = \nabla f(x) - \rho_k \sum \frac{1}{g_i(x)} \nabla g_i(x) = 0 \tag{4}$$

and by making the substitution $y_i = \frac{\rho_k}{g_i(x)}$ and introducing the residual Φ_k, one arrives at the primal-dual equations

$$\begin{aligned} \nabla f(x) - \sum y_i \nabla g_i(x) &= \Phi_k \\ y_i g_i(x) &= \rho_k \end{aligned} \tag{5}$$

Under suitable assumptions, this last system of equations implicitly defines differentiable trajectories $x(\rho, \Phi)$ and $y(\rho, \Phi)$ close to $\rho = 0$.

Theorem 1.1 *[4]Let x^* be a regular point of the constraints $g_{I^*}(x) = 0$ which satisfies to the second order sufficient optimality conditions for (1) as well as to the strict complementarity condition $y_{I^*} > 0$. If the functions f and g are $\mathcal{C}^p(\mathbb{R}^n)$, then there exists differentiable trajectories $x(\rho, \Phi)$ and $y(\rho, \Phi)$ of class $\mathcal{C}^{p-1}(\mathbb{R}^n)$ such that*

1. *$x(0, 0) = x^*$ and $y(0, 0) = y^*$;*
2. *if ρ and $\|\Phi\|$ are small enough, $x(\rho, 0)$ satisfies to the second order sufficient optimality conditions for the penalized sub-problems $\min f(x) - \rho \sum_{i=1}^m \log(-g_i(x))$, where $x(r, \Phi)$, $y(r, \Phi)$ are solutions of the following equations:*

$$\nabla f(x) + \nabla g(x)^t y = \Phi \tag{6}$$

$$g_i(x) y_i + \rho = 0 \quad , i = 1, 2, \ldots, m. \tag{7}$$

Moreover, the following bounds hold asymptotically:

(a) $\|x(\rho, \Phi) - x^*\| \sim \mathcal{O}(\max(\rho, \|\Phi\|))$;
(b) $\|y(\rho, \Phi) - y^*\| \sim \mathcal{O}(\max(\rho, \|\Phi\|))$;
(c) $\|g_{I^*}(x(\rho, \Phi))\| \sim \mathcal{O}(\rho)$.

Remark 1.1 *Although we use primal-dual equations, in this SUMT variant, the dual variables y are dependent on the primal x, so that global convergence in infered from the fact that the log barrier is minimized (using globally convergent algorithms), allowing to prove that cluster points of the generated sequence are indeed stationary. The price to pay for such strong global convergence property is two-fold:*

- *the need for an interior starting point*
- *poorer asymptotic convergence property due to the explicit dependence of y on x.*

We will denote $G(x) = diag(g_i(x))$ and

$$\Phi(x,\rho) = \nabla_x \phi(x,\rho) = \nabla f(x) - \sum \frac{\rho}{g_i(x)} \nabla g_i(x) = \nabla f(x) - \rho \nabla g^t(x) G(x)^{-1} e. \tag{8}$$

ϕ is closely related to the Lagrangian $l(x,\lambda) = f(x) + g(x)\lambda$, and by defining $\lambda = -\rho G(x)^{-1} e$, i.e. $\lambda_i = -\frac{\rho}{g_i(x)}$, $\nabla_x l(x,\lambda) = L(x,\lambda) = \nabla f(x) + \nabla g^t(x)\lambda = \Phi(x,\rho)$.

We are concerned with approximate solutions $x(\rho, r)$ which satisfy $\Phi(x(\rho,r),\rho) = r$. In the sequel $\|r\| \sim \rho$.

The basic predictor-corrector path following approach consists then in having an estimate $x(\rho, r)$ which satisfy $\Phi(x(\rho,r),\rho) = r$ and then iterate the following two steps:

---

predictor extrapolate $\hat{x}^1 = x + \frac{\partial x}{\partial \rho}(\rho^+ - \rho) + \frac{\partial x}{\partial r}(-r)$

corrector perform Newton corrections from $\hat{x}^1$ on the problem Eq.(3) for ρ^+ until $\|\Phi(x,\rho^+)\| \leq \rho^+$.

For this basic scheme, a single Newton correction asymptotically yields $x(\rho^+, r^+)$ with $\|r^+\| \leq \rho^+$ provided that $\frac{\rho^+}{\rho^{\frac{4}{3}}} \to 0$ yielding a two-steps super-linear convergence of limiting order $\frac{4}{3}$. If one is prepared to perform two Newton corrections, then the limiting order is improved to $\frac{\rho^+}{\rho^{\frac{8}{5}}} \to 0$ [5] yielding a three-steps super-linear convergence of limiting order $\frac{8}{5}$. Using a measure similar to Ostrovski efficiency, the optimal strategy for this family of algorithms is to aim for two Newton corrections following an extrapolation [5].

1.2 Primal-dual

The primal-dual algorithm is written out directly from the KKT conditions for Eq.(2):

$$\begin{aligned} \nabla f(x) + \nabla g(x) y &= 0 \\ e + z &= 0 \\ g(x) + s &= 0 \\ ZSe &= 0 \end{aligned} \tag{9}$$

by formulating

$$H(x,y,s,z) = \begin{pmatrix} F(x,y,s,z) \\ ZSe \end{pmatrix} = 0. \tag{10}$$

Then, denoting $w = (x,y,s,z)$, parametrized sub-problems

$$H(w) = \rho \begin{pmatrix} 0 \\ e \end{pmatrix} \tag{11}$$

define a primal-dual central path $w(\rho)$ corresponding to so-called differentiable trajectories in the SUMT notation.

The usual approach is to use Newton steps

$$\nabla H(w_k)(w_k^+ - w_k) + H(w_k) - \rho_{k+1}\tilde{e} = 0 \tag{12}$$

and provide conditions on the sequence ρ to ensure convergence.

Remark 1.2 *Here, primal and dual variables are all treated "equal", allowing enhanced asymptotic convergence order when compared to the SUMT approach. Global convergence is trickier, though, and for instance, the method can get trapped near an infeasible point for the non linear constraint $g(x) + s = 0$.*

Most (all?) linear programming (LP) (interior point) codes now use variants of the so called infeasible predictor-corrector method. The method, as proposed by Mehrotra, consists in solving first $\nabla H(w_k)(\tilde{w}_k - w_k) + H(w_k) = 0$, the so called affine scaling step, used as a predictor, then fix ρ using a clever heuristic, and then solve $\nabla H(\tau_k \tilde{w}_k + (1-\tau_k)w_k)(w_k^+ - w_k) + H(\tilde{w}_k) - \rho_{k+1}\tilde{e} = 0$ as a corrector step for suitable τ_k, also known as a centering step.

For non linear problems, [6] obtain a predictor-corrector algorithm that attains almost quadratic convergence without (asymptotically) requiring any corrector step, thus a one-step super-linear convergence of order strictly lower than two.

For NLPs, an elegant result by Armand and Benoist [1] ensures that in a suitable neighborhood, centering steps are unnecessary as long as $\frac{\rho^+}{\rho^2}$ approaches 0. In other words, if the sequence ρ_k converges strictly slower than quadratic, the iterates will approach the central path naturally.

1.3 Terminology

First order extrapolations are usually named "predictor" steps. Higher order terms are sometimes named "corrections" to the predictor, but we will stick to the terminology "higher order". Once a predictor (of arbitrary order) is computed, sometimes it is necessary to perform Newton iterations, refered to as a "corrector" steps. This is sometimes named "centrality correction" steps.

Predictor steps aim at changing the trajectory parameter ρ to a smaller value while corrector steps aim at improving the parametric solution for a given ρ-value.

2 Inexact variants

Since extrapolations (predictor steps) and Newton corrections are related to Newton steps, one may devise strategies to approximately compute the steps. We address in this section the asymptotic convergence order of such variants where both (predictors and correctors) steps are computed approximately.

2.1 Inexact SUMT

We now address inexact extrapolations and corrections. By solving approximately the equations defining the first order extrapolation $\hat{x}^1$, we get an extrapolate, denoted $\hat{x}$ with $\|\hat{x} - x(\rho^+, 0)\| \sim \rho^{a+1}$. If $a = 1$, we get as good a prediction as $\hat{x}^1$ while if $\hat{x}$ is computed cheaply, we insist to at least obtain an order $a + 1$ for some $a > 0$.

Assume that the Newton's direction is computed approximately such that $\nabla_x \Phi(\hat{x}, \rho^+) d_N + \Phi(\hat{x}, \rho^+) = R$ with $\|R\| \leq \rho^{1+c} = \gamma$.

Lemma 2.1 *Let $\hat{x}$ such that $\|\hat{x} - x(\rho^+, 0)\| \sim \rho^{a+1}$. Then, $d_N \sim \mathcal{O}(\gamma + \rho^{a+1})$.*

Proof. The primal-dual Newton's direction is written

$$\begin{pmatrix} \nabla_x L(\hat{x}, \hat{\lambda}) & \nabla \hat{g} \\ \hat{\Lambda} \nabla \hat{g}^t & \hat{G} \end{pmatrix} \begin{pmatrix} d_x \\ d_y \end{pmatrix} = \begin{pmatrix} -L(\hat{x}, \hat{\lambda}) + R \\ 0 = \hat{G}\hat{\lambda} - \rho^+ e \end{pmatrix}, \tag{13}$$

where $\hat{\lambda}_i = \frac{\rho^+}{\hat{g}_i}$. Define also $\lambda_i = \frac{\rho^+}{g_i}$, where $g_i = g_i(x(\rho^+, 0))$. By defining $\bar{d}_y = d_y - \lambda + \hat{\lambda}$, we may rewrite Eq.(13) as

$$\begin{pmatrix} \nabla_x L(\hat{x}, \hat{\lambda}) & \nabla \hat{g} \\ \hat{\Lambda} \nabla \hat{g}^t & \hat{G} \end{pmatrix} \begin{pmatrix} d_x \\ \bar{d}_y \end{pmatrix} = \begin{pmatrix} -L(\hat{x}, \lambda) + R \\ \hat{G}(\lambda - \hat{\lambda}) \end{pmatrix}, \tag{14}$$

We now observe that $L(\hat{x}, \lambda) = \mathcal{O}(\rho^{a+1})$ and $\hat{G}(\lambda - \hat{\lambda}) = \rho^+ G^{-1}(\hat{g} - g) = \mathcal{O}(\rho^{a+1})$ and $\|R\| \sim \mathcal{O}(\rho^{a+1})$ which concludes the proof. □

Now consider the effect of an inexact Newton correction.

Lemma 2.2 *Let $\hat{x}$ such that $\|\hat{x} - x(\rho^+, 0)\| \sim \rho^{a+1}$. Then, $\Phi(\hat{x} + d_N, \rho^+) \sim \mathcal{O}\left(\frac{\left(\gamma + \rho^{(a+1)}\right)^2}{\rho^{+2}}\right)$.*

Proof. We write

$$\Phi(\hat{x} + d_N, \rho^+) = \Phi(\hat{x}, \rho^+) + \nabla_x \Phi(\hat{x}, \rho^+) d_N + \mathcal{O}\left(\frac{\|d_N\|^2}{(\rho^+)^2}\right), \tag{15}$$

noting the last denominator $(\rho^+)^2$ which comes from derivating twice $\frac{\rho^+}{g_i(\hat{x})}$. The first two terms are bounded by γ and using the bound on d_N from the lemma 2.1 we get $\frac{\left(\gamma + \rho^{(a+1)}\right)^2}{\rho^{+2}}$. □

Consider $\rho^+ = \rho^b$. If we want to be as cheap as possible while ensuring super-linear convergence ($b > 1$), we deduce that $b < \frac{2(1+c)}{3}$ while $c \leq a$. Therefore, to get super-linear convergence, one has to pick $c = a > 0.5$.

2.2 Inexact primal-dual

Primal-dual variants were mainly studied from a complexity (polynomial time) point of view and recently, from a practical implemented efficiency. Few results actually address the asymptotic convergence order of inexact variants.

Bellavia [3] studies inexact interior point methods for non linear programs, and obtains super-linear convergence results, without characterizing the actual order.

Armand, Benoist and Dussault [2] extend the result in [1] to inexact variants using various forms of forcing sequences. If a forcing sequence which would provide the inexact Newton method super-linear convergence order of ξ is used, then asymptotically, no centrality correction is required as long as $\frac{\rho^\xi}{\rho^+} \to 0$. Therefore, one gets super-linear convergence with limiting order ξ.

3 High order variants

Instead of solving approximately the predictor and-or corrector steps, we investigate here the effect of using higher order Taylor expressions of the central path.

3.1 High order SUMT

We are now concerned with higher order extrapolations ($a > 1$). Postponing the actual computations of such a $\hat{x}^a$ for $a \neq 1$, we already may obtain the following.

Lemma 3.1 *Let $\hat{x}$ such that $\|\hat{x} - x(\rho^+, 0)\| \sim \rho^{a+1}$ with $\frac{\rho^{a+1}}{\rho^+} < \infty$. Then, $\nabla\phi(\hat{x}, \rho^+) \sim \mathcal{O}(\frac{\rho^{a+1}}{\rho^+})$.*

Proof. Denote $x = x(\rho^+, 0)$; then $\Phi(x, \rho^+) = 0$ and write:

$$\begin{aligned}
\Phi(\hat{x}, \rho^+) &= \Phi(\hat{x}, \rho^+) - \Phi(x, \rho^+) && (16)\\
&= \nabla\hat{f} - \nabla f + \sum \frac{\rho^+}{\hat{g}_i}\nabla\hat{g}_i - \sum \frac{\rho^+}{g_i}\nabla g_i && (17)\\
&= \mathcal{O}(\rho^{a+1}) + \sum \frac{\rho^+(g_i - \hat{g}_i)}{\hat{g}_i g_i}\nabla\hat{g}_i + \sum \frac{\rho^+}{g_i}(\nabla\hat{g}_i - \nabla g_i) && (18)\\
&= \mathcal{O}(\rho^{a+1}) + \sum \frac{\mathcal{O}(\rho^{a+1})\rho^+}{\hat{g}_i g_i}\nabla\hat{g}_i + \sum \lambda_i \mathcal{O}(\rho^{a+1}). && (19)
\end{aligned}$$

We have $g_i \sim \Theta(\rho^+)$ and $\hat{g}_i = g_i + \mathcal{O}(\rho^{a+1})$, so that $\hat{g}_i \sim \Theta(\rho^+)$ since $\frac{\rho^{a+1}}{\rho^+}$ is bounded. □

This result allows to claim that by using $(a > 1)$–order extrapolations, we get a $\frac{(a+1)}{2}$ order of convergence without even recourse to Newton corrections. Using a first order extrapolation is not enough, and requires a further Newton correction. Indeed, to reach the required approximation criterion, $\nabla\phi(\hat{x}, \rho^+)$ has to be lower than $\|r^+\| = \rho^+$, which implies that $\rho^{a+1} < {\rho^+}^2$.

3.1.1 Computing extrapolations

We rewrite equation Eq.(8) in a simplified notation:

$$\Phi(x, \rho) = c - \rho A^t G^{-1} e. \tag{20}$$

For linear programs, c and A are constant while otherwise, $c = \nabla f(x)$ and $A = \nabla g(x)$. $G = \mathrm{diag}(g_i(x))$, and for linear programs, $g(x) = Ax - b$. At the current point,

$$\Phi(x, \rho) = \tau\bar{\mathbf{r}} \tag{21}$$

for some residual vector $r = \tau\bar{\mathbf{r}}$ with $\bar{\mathbf{r}} = \frac{r}{\|r\|}$. Equation Eq.(21) induces a bi-parameter equation $x(\rho, \tau)$ and the solution searched for is $x^* = x(0, 0)$. We note for the sequel

$$\begin{aligned}
\nabla_x\Phi(x, \rho) &= \rho A^t G^{-2} A + \nabla^2_{xx} l(x, \rho G(x)^{-1} e) && (22)\\
\nabla_\rho\Phi(x, \rho) &= -A^t G^{-1} e && (23)
\end{aligned}$$

and remark that the Lagrangian term vanishes for linear programs.

The implicit function theorem yields

$$\begin{aligned} \nabla_x\Phi(x,\rho)\dot{x}_\rho + \nabla_\rho\Phi(x,\rho) &= 0 \\ \nabla_x\Phi(x,\rho)\dot{x}_\tau - \bar{\mathbf{r}} &= 0. \end{aligned} \tag{24}$$

Thus, the combined extrapolation step reduces to

$$\nabla_x\Phi(x,\rho)\left(\dot{x}_\rho(\rho^+-\rho)+\dot{x}_\tau(-\tau)\right)+\nabla_\rho\Phi(x,\rho)(\rho^+-\rho)+\tau\bar{\mathbf{r}}=0, \tag{25}$$

which, for this first order candidate, simplifies to $\nabla_x\Phi(x,\rho)\hat{x}^1+\Phi(x,\rho^+)=0$.

In order to go further to the expressions of higher order extrapolates, we first note the following:

$$\nabla^2_{x\rho}\Phi(x,\rho)=\nabla^2_{\rho x}\Phi(x,\rho) = A^tG^{-2}A+\nabla^2_{xx}l(x,G(x)^{-1}e) \tag{26}$$

$$\nabla^2_{\rho\rho}\Phi \equiv 0 \tag{27}$$

$$\nabla^2_{\tau\cdot}\Phi=\nabla^2_{\cdot\tau}\Phi \equiv 0 \tag{28}$$

In a nutshell, any derivative of Φ with respect to τ vanishes since Φ does not involve τ, and any high order derivative of Φ with respect to ρ more than once also vanishes since Φ is linear in ρ.

Now, still using the implicit function theorem, this time to equations Eq.(24), we get

$$\begin{aligned} \nabla^2_{xx}\Phi(x,\rho)\dot{x}_\rho\dot{x}_\rho+\left(\nabla_{x\rho}\Phi(x,\rho)+\nabla_{\rho x}\Phi(x,\rho)\right)\dot{x}_\rho+\nabla_x\Phi(x,\rho)\ddot{x}_{\rho\rho} &= 0 \\ \nabla^2_{xx}\Phi(x,\rho)\dot{x}_\tau\dot{x}_\rho+\nabla_{\rho x}\Phi(x,\rho)\dot{x}_\tau+\nabla_x\Phi(x,\rho)\ddot{x}_{\tau\rho} &= 0 \\ \nabla^2_{xx}\Phi(x,\rho)\dot{x}_\rho\dot{x}_\tau+\nabla_{x\rho}\Phi(x,\rho)\dot{x}_\tau+\nabla_x\Phi(x,\rho)\ddot{x}_{\rho\tau} &= 0 \\ \nabla^2_{xx}\Phi(x,\rho)\dot{x}_\tau\dot{x}_\tau+\nabla_x\Phi(x,\rho)\ddot{x}_{\tau\tau} &= 0 \end{aligned} \tag{29}$$

Observe that the four relations above all imply a linear system defined by the same matrix $\nabla_x\Phi(x,\rho)$ and the following four right hand sides, conveniently expressed using $\bar{x}_\tau$ which denotes a constant vector of value $\dot{x}_\tau$, and similarly $\bar{x}_\rho$ is a constant vector of value $\dot{x}_\rho$:

$$\nabla_\rho\left(\nabla_x\Phi(x,\rho)\bar{x}_\rho+\nabla_\rho\Phi(x,\rho)\right) = \nabla^2_{xx}\Phi(x,\rho)\dot{x}_\rho\bar{x}_\rho+\nabla^2_{x\rho}\Phi(x,\rho)\bar{x}_\rho+\nabla^2_{\rho x}\Phi(x,\rho)\dot{x}_\rho \tag{30}$$

$$\nabla_\tau\left(\nabla_x\Phi(x,\rho)\bar{x}_\rho+\nabla_\rho\Phi(x,\rho)\right) = \nabla^2_{xx}\Phi(x,\rho)\dot{x}_\tau\bar{x}_\rho+\nabla^2_{\rho x}\Phi(x,\rho)\dot{x}_\tau \tag{31}$$

$$\nabla_\rho\left(\nabla_x\Phi(x,\rho)\bar{x}_\tau-\bar{\mathbf{r}}\right) = \nabla^2_{xx}\Phi(x,\rho)\dot{x}_\rho\bar{x}_\tau+\nabla^2_{x\rho}\Phi(x,\rho)\bar{x}_\tau \tag{32}$$

$$\nabla_\tau\left(\nabla_x\Phi(x,\rho)\bar{x}_\tau-\bar{\mathbf{r}}\right) = \nabla^2_{xx}\Phi(x,\rho)\dot{x}_\tau\bar{x}_\tau \tag{33}$$

Hereafter, we use the "bar" $\bar{\rho}$ and $\bar{\tau}$ to represent actual extrapolation steps values, as opposed to variables within the equations. Now, $\hat{x}^2=\ddot{x}_{\rho\rho}(\bar{\rho}^+-\bar{\rho})^2+2\ddot{x}_{\tau\rho}(\bar{\rho}^+-\bar{\rho})(-\bar{\tau})+\ddot{x}_{\tau\tau}(\bar{\tau})^2$ so that the right hand sides involving the second derivatives may be combined into $(\bar{\rho}^+-\bar{\rho})((\bar{\rho}^+-\bar{\rho})Eq.(30)-\bar{\tau}Eq.(32))$ and $-\bar{\tau}((\bar{\rho}^+-\bar{\rho})Eq.(31)-\bar{\tau}Eq.(33))$ and, also using the notation that $\hat{\bar{x}}^1$ is a constant vector of value $\hat{x}^1$, is expressed:

$$\begin{aligned} &\nabla_\rho\left(\nabla_x\Phi(x,\rho)\hat{\bar{x}}^1+\nabla_\rho\Phi(x,\rho)(\bar{\rho}^+-\bar{\rho})-\bar{\mathbf{r}}\bar{\tau}\right)(\bar{\rho}^+-\bar{\rho}) \\ &+\nabla_\tau\left(\nabla_x\Phi(x,\rho)\hat{\bar{x}}^1+\nabla_\rho\Phi(x,\rho)(\bar{\rho}^+-\bar{\rho})-\mathbf{r}\bar{\tau}\right)(-\tau) \end{aligned} \tag{34}$$

So establish a recurrence relation to compute the $\hat{x}^p$, it is convenient to define a family of functions

$$\Phi^0(x,\rho,\tau)=\Phi(x,\rho)-\tau\bar{\mathbf{r}} \tag{35}$$

$$\Phi^p(x,\rho,\tau)=(\bar{\rho}^+-\bar{\rho})\nabla_\rho\Phi^{p-1}(x,\rho,\tau)-\bar{\tau}\nabla_\tau\Phi^{p-1}(x,\rho,\tau) \tag{36}$$

Theorem 3.2

$$\hat{x}^p=\sum_{j=0}^{p}\binom{p}{j}\frac{\partial^p x}{\partial\rho^{p-j}\partial\tau^j}(\bar{\rho}^+-\bar{\rho})^{p-j}(-\bar{\tau})^j$$

satisfies $\Phi^p(x,\rho,\tau)=0$.

Proof. The inductive proof has its base verified by the relation Eq.(34). The induction step will use the relation:

$$\hat{x}^{p+1} = \frac{\partial \hat{x}^p}{\partial \rho}(\bar{\rho}^+ - \bar{\rho}) + \frac{\partial \hat{x}^p}{\partial \tau}(-\bar{\tau})$$

The equations Φ^p includes a term $\nabla_x \Phi(x,\rho)\hat{x}^p$ defining the linear system, the remaining of Φ^p corresponding to the right hand side of the linear equation. □

The recurrence Φ^p may be explicitly written as

$$\nabla_x \Phi(x,\rho)\hat{x}^p + \hat{\Phi}^p$$

where $\hat{\Phi}^p$ involves terms of the form $\nabla^j_{x^{j_x}\rho^{j_\rho}}\Phi(x,\rho)v_1^{i_1}v_2^{i_2}\dots v_l^{i_l}$ with $\sum_{k=1}^{l} i_k = j_x$, $j_x + j_\rho = j$ and $1 < j \leq p$. Moreover, each v_k is composed of partial derivatives of x with respect to ρ and/or τ up to order $j_x - 1$. As it happens, the recurrence may be written using only the $\hat{x}^p$ without explicit reference to the (mixed) partials derivatives of x wrt ρ or τ:

$$\Phi^0(x,\rho,\tau) = \Phi(x,\rho) + \tau\bar{\mathbf{r}} \tag{37}$$

$$\Phi^1(x,\rho,\tau) = \nabla_x \Phi(x,\rho)\hat{x}^1 + (\bar{\rho}^+ - \bar{\rho})\nabla_\rho \Phi(x,\rho) - \bar{\tau}\bar{\mathbf{r}} \tag{38}$$

$$\Phi^2(x,\rho,\tau) = \nabla_x \Phi(x,\rho)\hat{x}^2 + 2(\bar{\rho}^+ - \bar{\rho})\nabla^2_{x\rho}\Phi(x,\rho)\hat{x}^1 + \nabla^2_{xx}\Phi(x,\rho)\hat{x}^1\hat{x}^1 \tag{39}$$

$$\begin{aligned}\Phi^3(x,\rho,\tau) =& \nabla_x \Phi(x,\rho)\hat{x}^3 + 3(\bar{\rho}^+ - \bar{\rho})\nabla^2_{x\rho}\Phi(x,\rho)\hat{x}^2 \\ &+ 3\nabla^2_{xx}\Phi(x,\rho)\hat{x}^1\hat{x}^2 + 3(\bar{\rho}^+ - \bar{\rho})\nabla^3_{xx\rho}\Phi(x,\rho)\hat{x}^1\hat{x}^1 + \nabla^3_{xxx}\Phi(x,\rho)\hat{x}^1\hat{x}^1\hat{x}^1\end{aligned} \tag{40}$$

$$\begin{aligned}\Phi^4(x,\rho,\tau) =& \nabla_x \Phi(x,\rho)\hat{x}^4 + 4(\bar{\rho}^+ - \bar{\rho})\nabla^2_{x\rho}\Phi(x,\rho)\hat{x}^3 \\ &+ 3\nabla^2_{xx}\Phi(x,\rho)\hat{x}^2\hat{x}^2 + 4\nabla^2_{xx}\Phi(x,\rho)\hat{x}^1\hat{x}^3 + 12(\bar{\rho}^+ - \bar{\rho})\nabla^3_{xx\rho}\Phi(x,\rho)\hat{x}^1\hat{x}^2 \\ &+ 6\nabla^3_{xxx}\Phi(x,\rho)\hat{x}^2\hat{x}^1\hat{x}^1 + 4(\bar{\rho}^+ - \bar{\rho})\nabla^4_{xxx\rho}\Phi(x,\rho)\hat{x}^1\hat{x}^1\hat{x}^1 \\ &+ \nabla^4_{xxxx}\Phi(x,\rho)(\hat{x}^1)^4\end{aligned} \tag{41}$$

3.1.2 Implementation for linear programming

By introducing the notation $v^1 = A\hat{x}^1$, and $V = \mathrm{diag}(v)$, we may express the high order terms using the following lemma.

Lemma 3.3

$$\nabla_x(v^t V_p G^{-p} u) = -pA^t V V_p G^{-(p+1)} u \tag{42}$$

This allows to write equation Eq.(39) as

$$\rho A^t G^{-2} A\hat{x}^2 - 2\rho A^t V^1 G^{-3} v^1 + 2(\bar{\rho}^+ - \bar{\rho})A^t G^{-2} v^1 = 0 \tag{43}$$

Similarly, equation Eq.(40) leads to the following expression:

$$-6\rho A^t V^1 G^{-3} v^2 - 6(\bar{\rho}^+ - \bar{\rho})A^t V^1 G^{-3} v^1 + 3(\bar{\rho}^+ - \bar{\rho})A^t G^{-2} v^2 + 6\rho A^t (V^1)^2 G^{-4} v^1 \tag{44}$$

As we may observe, each term involves a single matrix–vector computation in addition to several $\mathcal{O}(n)$ diagonal matrices and vector operations, overall yielding cheap right hand sides to compute higher order derivatives. This was to be expected.

3.1.3 General implementation using automatic differentiation (AD)

Using AD tools, we may evaluate higher order derivative cheaply too. Assuming full dense Hessian's—constraint jacobians, the linear system requires $\mathcal{O}(n^3)$ arithmetic operations to factorizes, and further on, back–front substitutions together with right hand side computations reduce to $\mathcal{O}(n^2)$ arithmetic operations. As it happens, we may get the high order right hand sides required for the Taylor coefficients in $\mathcal{O}(n^2)$ complexity, leaving the main burden to obtain and factorize the Jacobian Matrix.

3.2 High order primal-Dual

High order algorithms were introduced soon after the predictor-corrector approach for LP. In this context, one may use higher order Taylor approximation to the central path to improve the predictor step or use several corrector steps to improve the predictor (Shamanskii like method). This latter option, at the heart of the excellent code HOPDM, seems to give the best results.

Mehrotra obtained the following result for the Shamanskii-like high order terms algorithm for LP:

Theorem 3.4 *[7] Let $\{w^l\}$ be generated by the generalized predictor-corrector algorithm using r high order terms. Then,*

1. $\rho_l \to 0$ *with Q-order* $r+2$*;*
2. $Z_i^l s_i^l \to 0$ *with Q-order* $r+2$*.*

In the formulation Eq.(2), for linear complementarity problems (LCPs), encompassing LPs, the only non linearity comes from the terms ZSe and explicit Taylor coefficients are readily available. By using such high order Taylor predictors, Stoer, Wechs and Mizuno [8] obtained similar results, using $r-order$ Taylor extrapolations and a centering corrector step, they obtain $(r+1-\epsilon)$ Q-order of convergence. They further obtain generalization without the strict complementarity assumption.

4 Conclusion and future work

In this paper, we summarized known results about the asymptotic behavior of interior point algorithms in non-linear optimization. We considered both primal and primal-dual variants. Overall, primal-dual variants enjoy better asymptotic properties, but as is well known, global convergence of those variants is trickier to achieve, although clearly possible. On the other hand, primal variants are naturally globally convergent, and one may devise hybrid algorithms which rely on some barrier primal sub-problem to enforce global convergence and on primal-dual equations to achieve almost quadratic convergence order. The following table summarizes the orders discussed above.

variant	primal	primal-dual
corrector predictor-corrector	linear 2-step 4/3 [4]	quadratic [6] cubic
order-$k \geq 2$ pred order-k pred-corr	$\frac{k+1}{2}$ $\frac{2(k+1)}{3}$(NLP)	— $k+2$(LP [7]); $k+1$ (LCP [8])
inexact pred inexact pred-corr	— $\frac{2}{3}(1+\xi)$	$1+\xi$ [2] —

Table 1: Limiting convergence orders for variants discussed in the paper

As can be seen from the table, primal variants suffer from poorer asymptotic convergence order. High order predictors bring them somewhat competitive. It should me mentioned that order-k predictor incur a computational cost of $\mathcal{O}(n^3)$ arithmetic operation to factorize the jacobian matrix plus k times $\mathcal{O}(n^2)$ to obtain the high order terms while the corrector steps involve the solution of a linear system, again $\mathcal{O}(n^3)$. Primal-dual variants involve larger matrices, but less prone to fill in in their factorization when sparse.

4.1 Next

Shamanskii-like variants are much lighter to implement than high order Taylor based extrapolations (see section 3.1.1). This aspect makes them attractive to generalize the results from LP and LCP since for non-linear problems, everything is non-linear, not just the complementarity terms, which complicates the expressions of higher order Taylor coefficients. Also, since primal variants look less disadvantaged by using at least second order extrapolations, some investigation is in order to re-assess their potential in the non-linear context.

References

[1] Paul Armand and Joël Benoist. A local covergence property of primal-dual methods for nonlinear programming. *Mathematical Programming*, 2007.

[2] Paul Armand, Joël Benoist, and Jean-Pierre Dussault. Local path-following property of inexact interior methods, 2010. Optimization online, presented at CIRO 2010.

[3] Stefania Bellavia. Inexact interior-point method. *J. Optim. Theory Appl.*, 96(1):109–121, 1998.

[4] Jean-Pierre Dussault. Numerical stability and efficiency of penalty algorithms. *S.I.A.M. Journal on Numerical Analysis*, 32(1):296–317, February 1995.

[5] Jean-Pierre Dussault and Abdelatif Elafia. On the convergence rate of the logarithmic barrier algorithm. *Computational Optimization and Applications*, 19(1):31–54, apr 2001.

[6] Nicholas I. M. Gould, Dominique Orban, Annick Sartenaer, and Philippe L. Toint. Superlinear convergence of primal-dual interior point algorithms for nonlinear programming. *SIAM J. on Optimization*, 11(4):974–1002, 2000.

[7] Sanjay Mehrotra. Asymptotic convergence in a generalized predictor-corrector method. *Math. Program.*, 74(1):11–28, 1996.

[8] Josef Stoer, Martin Wechs, and Shinji Mizuno. High order infeasible-interior-point methods for solving sufficient linear complementarity problems. *Math. Oper. Res.*, 23(4):832–862, 1998.