DIFFERENTIABLE LOCAL BARRIER-PENALTY PATHS

C. Grossmann [1]

[1]*TU Dresden, Institut für Numerische Mathematik, D-01062 Dresden, Germany, grossm@math.tu-dresden.de**

Abstract   Perturbations of Karush-Kuhn-Tucker conditions play an important role for primal-dual interior point methods. Beside the usual logarithmic barrier various further techniques of sequential unconstrained minimization are well known. However other than logarithmic embeddings are rarely studied in connection with Newton path-following methods. A key property that allows to extend the class of methods is the existence of a locally Lipschitz continuous path leading to a primal-dual solution of the KKT-system. In this paper a rather general class of barrier/penalty functions is studied. In particular, under LICQ regularity and strict complementarity assumptions the differentiability of the path generated by any choice of barrier/penalty functions from this class is shown. This way equality as well as inequality constraints can be treated directly without additional transformations. Further, it will be sketched how local convergence of the related Newton path-following methods can be proved without direct applications of self-concordance properties.

keywords: Perturbed KKT-systems, general barrier-penalty embedding, differentiable path, path-following methods, interior point methods

1.   Barrier/penalty functions and primal-dual paths

Barrier/penalty methods and its path-following variants form an important class of numerical methods for constrained optimization problems via a family of unconstrained ones (cf. [2], [8], [9]). While large classes of classical barrier/penalty methods are well studied already in [3] path-following methods mainly restrict to log-barrier terms. The aim of the present paper is to provide a convergence concept for a wide range of path-following Newton methods under strong regularity assumptions. The concerning results are derived in detail in [5]. In addition, following [6] for the log-barrier method we discuss the relaxation of the LICQ regularity assumption by MFCQ. In this case we show

that the study of the behavior of the log-barrier method applied to a locally linearized problem provides full information upon the convergence properties of the approximated duals of the original nonlinear problem.

Considered are nonlinear programming problems

$$f(x) \rightarrow \min !$$
$$\text{s.t.} \quad x \in G := \left\{ x \in R^n : g_i(x) = 0, \ i \in I^g, \quad g_i(x) \leq 0, \ i \in I^u \right\}, \tag{1}$$

where

$$I^g := \{1, \ldots, q\}, \quad I^u := \{q+1, \ldots, m\}, \quad I := I^g \cup I^u$$

and $f, g_i : R^n \rightarrow R, \ i \in I$ denote twice Lipschitz continuously differentiable functions. Let abbreviate $g : R^n \rightarrow R^m$ with $g = (g_1, \ldots, g_m)^T$ and

$$G_g := \{x \in R^n : g_i(x) = 0, \ i \in I^g\}, \quad G_u^0 := \{x \in R^n : g_i(x) < 0, \ i \in I^u\}.$$

Problem (1) is supposed to possess some local solution $x^* \in G$ that satisfies the linear independence constraint qualification (LICQ). In particular, this implies that a uniquely defined multiplier vector $y^* \in R^m$ exists such that the KKT-conditions

$$\nabla_x L(x^*, y^*) = 0, \qquad g_i(x^*) = 0, \ i \in I^g,$$
$$g_i(x^*) \leq 0, \quad y_i^* \geq 0, \ i \in I^u, \quad y^{*T} g(x^*) = 0. \tag{2}$$

hold.   In addition to LICQ we assume strict complementarity, i.e. $y_i^* \neq 0 \ \Leftrightarrow \ i \in I_0$ and that x^* satisfies the well-known second order sufficiency condition. Here denotes $I_0 := I_0(x^*) := \{ i \in I : g_i(x^*) = 0 \}$. Further, let $\nabla_x L, \ \nabla_{xx}^2 L$ be the partial gradient and Hessian, respectively, of the Lagrangian L. Taking into account strict complementarity second order sufficiency condition simplifies to (2) and

$$v^T \nabla_{xx}^2 L(x^*, y^*)v > 0 \quad \forall v \in R^n, \quad \nabla g_i(x^*)^T v = 0, \ i \in I_0, \ v \neq 0. \tag{3}$$

Further, we notice that LICQ also implies $G_g \cap G_u^0 \neq \emptyset$ which allows to apply classical barrier methods locally to all the inequality constraints of the optimization problem (1).

In barrier/penalty-methods the constraints of the original nonlinear programming problem (1) are incorporated into the objective in such a way that violations of the constraints are asymptotically avoided by extra costs. Instead of problem (1) we consider the related unconstrained auxiliary problems

$$F(x, s) := f(x) + \sum_{i \in I} \phi_i(g_i(x), s) \rightarrow \min !$$
$$\text{s.t.} \qquad x \in B_s := \{ x \in R^n \mid \phi_i(g_i(x), s) < +\infty, \ i \in I\}. \tag{4}$$

Here $\phi_i(\cdot,\cdot) : R \times R_{++} \to \overline{R}$, $i \in I$, denote barrier/penalty-functions which depend upon the barrier/penalty-parameter $s > 0$. Let

$$R_+ := \{\, t \in R : t \geq 0\,\}, \qquad R_{++} := \{\, t \in R : t > 0\,\}, \qquad \overline{R} := R \cup \{+\infty\}.$$

To ensure differentiability of the local path throughout this paper we assume that for any $s > 0$ the barrier/penalty-functions $\phi_i(\cdot,s) : R \to \overline{R}$ are differentiable in $\mathrm{dom}\,\phi_i(\cdot,s)$ and satisfy

$$\frac{\partial}{\partial t}\,\phi_i(t,s) \;=\; \psi_i\left(\frac{t}{s}\right) \qquad \forall\, t \in \mathrm{dom}\,\phi_i(\cdot,s), \quad s > 0, \tag{5}$$

with some $\psi_i : R \to R$, $i \in I^g$, $\psi_i : R \to \overline{R}$, $i \in I^u$, $\psi_i \not\equiv 0$, $i \in I$. The functions ψ_i we call the generating functions for the barrier/penalty-method. The relation (5) between barrier/penalty-functions and their generating functions was proposed in [4] for path-following algorithms applied to inequality constrained problems.

REMARK 1 *The same structural assumption (5) was considered by Auslender et. al. [1] for saddle point problems. The following supposed properties, however, differ from those made in [1] due to our goal to establish convergence of path-following Newton methods.*

Assumed properties for ψ_i, $i \in I^u$:
U1: $\mathrm{dom}\,\psi_i = (-\infty, d_i)$ with some $d_i \in \overline{R}$ and $\lim\limits_{r \to d_i-} \psi_i(r) = +\infty$.

U2: $\psi_i : R \to \overline{R}$ convex, differentiable in $\mathrm{dom}\,\psi_i$ with ψ_i' locally Lipschitz

$$|\psi_i'(\rho_1) - \psi_i'(\rho_2)| \leq L_1(r)|\rho_1 - \rho_2| \quad \forall\, \rho_1,\, \rho_2 \leq r < d_i$$

and

$$|\psi_i'(\rho_1) - \psi_i'(\rho_2)| \leq L_2(r)\left|\frac{1}{\rho_1} - \frac{1}{\rho_2}\right| \quad \forall\, \rho_1,\, \rho_2 \leq r < \min\{0, d_i\}.$$

with some nondecreasing $L_1(\cdot),\, L_2(\cdot) : R_{++} \to R_{++}$.
U3: $\psi_i'(r) \geq 0 \quad \forall\, r \in \mathrm{dom}\,\psi_i$, $\lim\limits_{r \to -\infty} \psi_i(r) = 0$, $\lim\limits_{r \to -\infty} r^2\,\psi_i'(r)$ exists and is finite.
    Assumed properties for ψ_i, $i \in I^g$:
G1: $\mathrm{dom}\,\psi_i = R$, $R = \psi_i(\mathrm{dom}\,\psi_i)$.
G2: ψ_i differentiable with ψ_i' locally Lipschitz continuous

$$|\psi_i'(\rho_1) - \psi_i'(\rho_2)| \leq L_3(r)\,|\rho_1 - \rho_2| \qquad \forall\, |\rho_1|,\, |\rho_2| \leq r$$

with some nondecreasing $L_3(\cdot) : R_{++} \to R_{++}$.
G3: $\psi_i'(r) \geq 0, \quad \forall\, r \in R \quad$ and $\quad \psi_i(r) \neq 0 \implies \psi_i'(r) > 0$.

Examples for generating functions are:

- $\psi_i(r) := \begin{cases} |d_i - r|^{-p} & \text{, if } r < d_i, \\ +\infty & \text{, if } r \geq d_i, \end{cases}$
  in case $d_i = 0$: $p = 1$ log-barrier, $p = 2$ Fiacco/McCormick's SUMT; otherwise shifted version.

- $\psi_i(r) := \max^{p-1}\{0, r\}$, $r \in R$,
  $p \geq 2$ corresponds to p-th order penalty function.

- $\psi_i(r) = \exp(r)$, $r \in R$
  exponential penalty.

- $\psi_i(r) = \text{sign}(r)|r|^{p-1}$, $r \in R$
  with fixed $p > 1$, (penalty for equality constraints).

We notice that strict complementarity, LICQ and the second order sufficiency conditions guarantee that the wanted minimizers of $F(\cdot, s)$ can be characterized by the necessary and sufficient local optimality condition

$$x(s) \in B_s : \qquad \nabla f(x(s)) + \sum_{i \in I} y_i(s) \nabla g_i(x(s)) = 0$$

with the so-called barrier/penalty multipliers $y_i(s) := \psi_i(\frac{g_i(x)}{s})$. The main result concerning stability behavior of the specific perturbation of the KKT-system is

THEOREM 2 *Under the made assumptions, there exist some $\bar{s} > 0$, $\delta > 0$ such that for any $s \in (0, \bar{s}]$ the parametric system*

$$\begin{aligned} \nabla f(x(s)) + \sum_{i \in I} y_i(s) \, \nabla g_i(x(s)) &= 0 \\ y_i(s) - \psi_i(g_i(x(s))/s) &= 0, \quad i \in I \end{aligned} \tag{6}$$

possesses a unique solution $(x(s), y(s))$ with $x(s) \in B_s \cap U_\delta(x^*)$, and we have

$$\lim_{s \to 0+} (x(s), y(s)) = (x^*, y^*).$$

With

$$x(0) := x^*, \quad y(0) := y^*, \tag{7}$$

the functions $x(\cdot)$, $y(\cdot)$ are continuously differentiable in $(0, \bar{s}]$, possess right sided derivatives at $s = 0$ and these derivatives are bounded for $s \to 0+$.

The proof of this theorem essentially rests on the implicit function theorem applied to the following perturbed KKT-system

$$\begin{aligned} \nabla f(x(s, r)) + \sum_{i \in I_0} y_i(s, r) \, \nabla g_i(x(s, r)) &= r, \\ s \, \psi_i^{-1}(y_i(s, r)) &= g_i(x(s, r)), \quad i \in I_0. \end{aligned}$$

For this the main property (see [5]) is the regularity of the matrix

$$
H(s) := \begin{pmatrix}
\nabla_{xx} L(x(s), y(s)) & \nabla g_1(x(s)) & \cdot & \cdot & \nabla g_m(x(s)) \\
\psi'_1\left(\frac{g_1(x(s))}{s}\right) \nabla g_1(x(s))^T & -s & 0 & \cdot & 0 \\
\psi'_2\left(\frac{g_2(x(s))}{s}\right) \nabla g_2(x(s))^T & 0 & -s & \cdot & 0 \\
& \cdot\cdot & \cdot & \cdot\cdot & \cdot \\
\psi'_m\left(\frac{g_m(x(s))}{s}\right) \nabla g_m(x(s))^T & 0 & 0 & \cdot & -s
\end{pmatrix}
$$

for sufficiently small $s > 0$.

As a direct consequence of Theorem 2 holds

COROLLARY 3 *Under the given assumptions there exist some constants* $s_0 \in (0, \bar{s}]$ *and* $c_L > 0$ *such that*

$$
\left.\begin{array}{l}
\|x(s) - x(t)\| \leq c_L \, |s - t| \\
\|y(s) - y(t)\| \leq c_L \, |s - t|
\end{array}\right\} \qquad \forall\, s,\, t \in [0, s_0]. \tag{8}
$$

2.   Log-Barriers Under Weaker Assumptions

In this section we follow widely [6] and restrict us to inequality constrained optimization problems, i.e. to

$$
f(x) \to \min ! \quad \text{s.t.} \quad x \in G = \{\, x \in R^n \, : \, g_i(x) \leq 0,\ i \in I^u \,\}. \tag{9}
$$

To this problem we apply log-barrier embedding and obtain the auxiliary problem

$$
F(x, s) := f(x) - s \sum_{i \in I^u} \ln(-g_i(x)) \to \min ! \quad \text{s. t.} \quad x \in G^0. \tag{10}
$$

In contrast to the first part of the paper, now the regularity assumptions are relaxed as follows:

(A1) x^* is some local minimizer of (9).

(A2) MFCQ is satisfied at x^*, i.e.,

$$
U^0 := \{u \in R^n : \nabla g_i(x^*)^T u < 0 \quad \forall i \in I_0\} \neq \emptyset.
$$

(A3) the *strict complementarity condition w.r. to* Y^* holds, i.e.,

$$
\exists\, y^* \in Y^* \quad \text{with} \quad y_i^* > 0 \ \forall i \in I_0. \tag{11}
$$

For a general study of regularity conditions and stability in nonlinear programming we refer to [7].

Next we introduce a locally linearized problem and show that log-barrier methods applied to it behave asymptotically like log-barrier methods applied to the original problem. Let denote

$$A := (\ldots \nabla g_i(x^*) \ldots)_{i \in I_0} \qquad \text{(column-wise)}$$

Related to x^* with $d = x - x^*$ we study the locally linearized problem

$$\nabla f(x^*)^T d \to \text{min !} \qquad \text{s.t.} \qquad d \in \mathcal{R}(A), \quad A^T d \le 0. \tag{12}$$

Here $\mathcal{R}(A)$ stands for the range of A. Notice that the point $d^* = 0$ forms the unique solution of problem (12). Log-barriers applied to (12) yield the auxiliary problems

$$\varphi_s(d) = \nabla f(x^*)^T d - s \sum_{i \in I_0} \ln(-\nabla g_i(x^*)^T d) \to \text{min !} \tag{13}$$

$$\text{s. t.} \qquad d \in D^0 = \{ d \in \mathcal{R}(A) : A^T d < 0 \}.$$

For these we have

LEMMA 4 *For any $s > 0$ problem (13) possesses a unique solution $\tilde{d}(s)$. Further, there is a unique solution d^* of the problem*

$$\prod_{i \in I_0} \frac{-\nabla g_i(x^*)^T d}{\nabla f(x^*)^T d} \to \text{max !} \qquad \text{s.t.} \quad d \in D^0, \quad \|d\| = 1, \tag{14}$$

and it holds $\tilde{d}(s) = t_s d^*$ with some $t_s > 0$ for all $s > 0$ as well as $\|\tilde{d}(s)\| = O(s)$.

THEOREM 5 *The log-barrier method (13) yields for the barrier multipliers $\tilde{y}(s)$ related to the solutions $\tilde{x}(s)$ that*

$$\tilde{y}_i(s) := \frac{-s}{\tilde{g}_i(\tilde{x}(s))} \equiv \mu_i, \quad i \in I_0, \tag{15}$$

where

$$\mu_i := -\frac{1}{m_0} \frac{\nabla f(x^*)^T d^*}{\nabla g_i(x^*)^T d^*}, \; i \in I_0 \quad \text{and } d^* \text{ solves (14).} \tag{16}$$

Setting $\mu_i = 0 \; \forall i \in I_1$, μ is a multiplier of the original problem (9).

Next we study the nonlinear problem

$$F(x, s) = f(x) - s \sum_{i=1}^{m} \ln(-g_i(x)) \to \text{min !}$$

$$\text{s. t.} \quad x \in G^\varepsilon := \{ x \in R^n : g_i(x) < 0, \; i = 1, \ldots, m, \; \|x - x^*\| < \varepsilon \},$$

In addition to (A1), (A2), (A3) assume:

(A4) The second-order optimality condition holds:

$$u^T \nabla^2_{xx} L(x^*, y) u > 0 \text{ for all } y \in Y^* \text{ and all } u \in U^*, u \neq 0,$$

where

$$U^* := \{u \; : \; \nabla f(x^*)^T u = 0, \; \nabla g_i(x^*)^T u \leq 0 \; \forall \, i \in I_0\}$$

is the critical cone for x^*.

THEOREM 6 *There are $\bar{s} > 0$ and $\varepsilon > 0$ such that for all $s \in (0, \bar{s})$, the function $F(\cdot, s)$ on G^ε has a global minimizer $x(s)$ which is the unique stationary point of $F(\cdot, s)$ on G^ε. The associated multipliers $y(s)$ converge to μ given in Theorem 5 where*

$$\text{dist}((x(s), y(s)), (x^*, Y^*)) \leq C^* s \quad \text{with some} \quad C^* > 0,$$

the Hessian $\nabla^2 F(x(s), s)$ is uniformly positive definite and $x(\cdot)$ is continuously differentiable on $(0, \bar{s})$.

3. Path-Following Primal-Dual Methods

We consider the convergence of Newton's method applied to the complete primal-dual system (6). Unlike in primal methods its first part stabilizes the approximation of the duals. However, as in the primal approach system, (6) also becomes increasingly ill-conditioned as $s \to 0+$.

Let denote

$$z := \begin{pmatrix} x \\ y \end{pmatrix} \in Z := R^n \times R^m$$

the vector of all primal and dual components. Further, let $T : Z \to Z$ denote the mapping

$$T(z, s) := \begin{pmatrix} T_1(z) \\ T_2(z, s) \end{pmatrix}$$

with

$$T_1(z) := \nabla_x L(x, y), \qquad T_2(z, s) := \psi\left(\frac{g(x)}{s}\right) - y.$$

With these notations (6) can be written as the following parametric system of nonlinear equations

$$T(z, s) = 0. \tag{17}$$

For a fixed barrier/penalty-parameter $s > 0$ a single Newton-step maps an old guess $z \in Z$ to a new approximate $\hat{z} \in Z$ of the solution $z(s)$ of (17) by

$$T'(z, s)(\hat{z} - z) + T(z, s) = 0. \tag{18}$$

Taking into account the structure of (6), the Jacobian has the form

$$T'(z, s) := \begin{pmatrix} Q & B \\ DB^T & -I \end{pmatrix}$$

with

$$Q \quad := \quad Q(x, y) := \nabla^2_{xx} L(x, y),$$

$$B \quad := \quad B(x) := (\nabla g_1(x), \dots, \nabla g_m(x)),$$

$$D \quad := \quad D(x) := \text{diag} \left\{ \tfrac{1}{s} \psi'_i \left(\tfrac{g_i(x)}{s} \right) \right\}_{i \in I}.$$

Since system (6) is increasingly ill-conditioned for $s \to 0+$ an adapted analysis for the Newton system is required to obtain sharp error bounds. In connection with log-barrier interior point methods, self-concordance (cf. [8], [9]) forms a common tool. We apply a different approach (cf. [4]) that analyzes such ill-posed systems directly in the Euclidean norm.

Let remark that in case of log-barriers we have $\Psi(r) = 1/r$ which allows to rewrite the second part as
$$g_i(x) \, y_i \, - \, s \, = \, 0, \quad i \in I.$$

A similar transformation is possible recommended if ψ_i is strictly monotone in domψ_i. This transformation stabilizes the numerical process, but does not remove the generic asymptotic singular behavior of the system (17) for $s \to 0+$.

In path-following Newton methods for a fixed barrier/penalty-parameter $s_k > 0$, and known $z^k \in Z$, we define the new iterate $z^{k+1} \in Z$, by only one Newton-step, i.e.

$$T'(z^k, s_k)(z^{k+1} - z^k) + T(z^k, s_k) = 0 \qquad (19)$$

and update the parameter by $s_{k+1} = \gamma \, s_k$ with some $\gamma \in (0, 1)$. This yields the long-step version of a

Path-Following Algorithm

Step 1: Select parameters ε, c, $s_0 > 0$, and $\nu \in (0, 1)$.
Find $x^0 \in B_{s_0}$ such that

$$\| z^0 - z(s_0) \| \leq c \, s_0. \qquad (20)$$

Set $k := 0$.

Step 2: Determine $z^{k+1} \in Z$ via the linear system

$$\begin{aligned} T'(z^k, s_k) \, d^k &= -T(z^k, s_k)) \\ z^{k+1} &:= z^k + d^k \end{aligned} \qquad (21)$$

Step 3: If $s_k \leq \varepsilon$ then stop. Otherwise set $s_{k+1} := \nu s_k$ and go to Step 2
with $k := k + 1$.

THEOREM 7 *For sufficiently small $s_0 > 0$ and $c > 0$ there exists some parameter $\nu \in (0, 1)$ such that then the given path-following algorithm is well defined and generates iterates $z_k \in Z$ that satisfy*

$$\|z^k - z(s_k)\| \leq c s_k, \qquad k = 0, 1, \dots . \tag{22}$$

Furthermore, the algorithm terminates after at most $k^* := \lceil \ln(\varepsilon/s_0)/\ln(\nu) \rceil$ steps and the estimate

$$\|z^{k^*} - z^*\| \leq (c_L + c) \varepsilon \tag{23}$$

holds, where c_L denotes the Lipschitz constant from Corollary 3.

For the proof as well as for further details we refer to [5].
    To ensure a larger range of convergence the given path-following algorithm has to be endowed with an additional step size procedure in step 2, i.e. we apply

$$z^{k+1} := z^k + \alpha_k d^k$$

with some $\alpha_k > 0$ appropriately defined, e.g. by Armijo's rule.
    An additional stabilization can be obtained by the use of the available approximations of the Lagrangian multipliers. The basic idea rests on

$$\psi_i\left(\frac{g_i(x(s))}{s}\right) \approx y_i(s) \to y_i^*, \quad i \in I.$$

Taking into account $g_i(x^*) = 0$, $i \in I_0$ the generating function is modified by shifts to satisfy

$$\psi_i(0) = y_i(s), \quad i \in I_0. \tag{24}$$

Standard IP-methods like log-barrier do not allow this, but shifted methods that are also covered by the assumed properties of ψ_i do.
    Consider shifted log-barrier, i.e.

$$\psi_i(r) := \begin{cases} 1/(d_i - r) & \text{, if } r < d_i, \\ +\infty & \text{, if } r \geq d_i, \end{cases}$$

In this case this leads to the update

$$\hat{d}_i = 1/y_i(s), \quad i \in I_0, \tag{25}$$

with I_0 approximately identified via the magnitude of $y_i(s)$.

The case of shifted quadratic loss penalties

$$\psi_i(r) = \max\{0, d_i + r\}$$

leads to

$$\hat{d}_i = y_i(s), \quad i \in I_0. \tag{26}$$

This is directly related to augmented Lagrangian techniques.

References

[1] A. Auslender, R. Cominetti, M. Haddou. Asymptotic analysis for penalty and barrier methods in convex and linear programming. *Math. Oper. Res.* 22:43-62, 1997.

[2] A. Forsgren, P.E. Gill, M.H. Wright. Interior methods for nonlinear optimization. *SIAM Review* 44:525-597, 2002.

[3] A.V. Fiacco, G.P. McCormick. *Nonlinear programming: Sequential unconstrained minimization techniques*, Wiley, New York, 1968.

[4] C. Grossmann, Penalty/Barrier path-following in linearly constrained optimization. *Discussiones Mathematicae, Differential Inclusions, Control and Optimization* 20:7-26, 2000.

[5] C. Grossmann, M. Zadlo. General primal-dual barrier-penalty path-following Newton methods for nonlinear programming. TU Dresden, Preprint TUD-NM-10-03, 2003 (to appear in Optimization)

[6] C. Grossmann, D. Klatte, B. Kummer. Convergence of primal-dual solutions for the non-convex log-barrier method without LICQ. *Kybernetika* 40: 571-584, 2004.

[7] D. Klatte, B. Kummer. *Nonsmooth equations in optimization - regularity, calculus, methods and applications.* Kluwer, Dordrecht, 2002.

[8] Y. Nesterov, A. Nemirovskii. *Interior-point polynomial algorithms in convex programming*, SIAM Publications, Philadelphia, 1994.

[9] S.J. Wright. *Primal-Dual Interior-Point Methods.* SIAM Publications, Philadelphia, 1997.