TECHNICAL RESEARCH REPORT

A Primal-Dual Interior-Point Method for Nonconvex Optimization with Multiple Logarithmic Barrier Parameters and with Strong Convergence Properties

by T. Urban, A.L. Tits, C.T. Lawrence

T.R. 98-27

A primal-dual interior-point method for nonconvex optimization with multiple logarithmic barrier parameters and with strong convergence properties

Thomas Urban

André L. Tits

Craig T. Lawrence

July 6, 1998

ABSTRACT

It is observed that an algorithm proposed in the 1980s for the solution of nonconvex constrained optimization problems is in fact a primal-dual logarithmic barrier interior-point method closely related to methods under current investigation in the research community. Its main distinguishing features are judicious selection and update of the multiple barrier parameters (one per constraint), use of the objective function as merit function, and careful bending of the search direction. As a pay-off, global convergence and fast local convergence ensue. The purpose of the present note is to describe the algorithm in the interior-point framework and language and to provide a preliminary numerical evaluation. The latter shows that the method compares well with algorithms recently proposed by other research groups.

1. Introduction

Consider the problem

$$\min_{x \in \mathcal{R}^n} \; f(x) \text{ subject to } d(x) \geq 0 \tag{P1}$$

where $f : \mathcal{R}^n \to \mathcal{R}$ and $d : \mathcal{R}^n \to \mathcal{R}^p$ are smooth and the relation $d(x) \geq 0$ is understood componentwise.

Over a decade ago, a feasible-iterate algorithm for solving (P1) was proposed based on the following idea. First, given strictly feasible estimates $\hat{x}$ of a solution and $\hat{z}$ of the corresponding Karush-Kuhn-Tucker (KKT) multiplier vector, compute the Newton (or a quasi-Newton) direction $(\Delta x, \Delta z)$ for the solution of the equalities in the KKT first order necessary conditions of optimality. Note that, if the Hessian (or Hessian estimate) is positive definite, the primal direction Δx is a direction of descent for f but that it may not allow a reasonably long step to be taken inside the feasible set. Second, motivated by this observation, solve again the same system of equations, but with the right-hand side perturbed so as to tilt the primal direction away from the constraint boundaries into the feasible set. The perturbation should be small enough that the tilted primal direction remains a descent direction for f and its size should decrease as a solution is approached, i.e., as $\|\Delta x\|$ decreases, so that a solution point located on the constraint boundaries can be reached. Third, bend the primal direction by means of a second order correction, and perform a line search, with f as a merit function, on the resulting arc. Bending is necessary if a Maratos-like effect is to be avoided, i.e., if a full step of one is to be allowed by the line search criterion close to the solution. These ideas were put forth in [1]. It was shown there that, under standard assumptions, global convergence as well as local superlinear convergence (in the case of approximate Hessian) can be achieved if the amounts of tilting and bending are appropriately chosen. It turns out that the "correct" amounts depend on Δx, Δz as well as z. Note that the central idea in the algorithm of [1] originated in earlier work by Herskovits and others [2, 3, 4]; see [5] for a detailed historical account. Ideas were also borrowed from [6] and [7].

The main purpose of the present note is to present the algorithm of [1] in the framework and language of recent interior-point methods. Indeed, it turns out that the tilted primal direction and the dual direction described above are also the primal and dual directions produced by a logarithmic barrier method very similar to methods recently discussed in the primal-dual interior-point literature (e.g., [8, 9, 10, 11, 12, 13, 14]). In particular, the algorithm of [1] bears striking similarities with that discussed by Gay, Overton and Wright [14]. The main differences are as follows:

- multiple barrier parameters μ_j are used, one per constraint;
- the barrier parameters are updated based on the "untilted" primal direction (obtained with $\mu_j = 0$ for all j) and are allowed to occasionally increase;
- f is used as the merit function and the search is along an arc obtained by bending the primal direction to allow, close to the solution, satisfaction of the line search criterion with a full step of one (Maratos effect avoidance);
- different steplengths are used for the primal and each component of the dual.

The paper is organized as follows. In Section 2, the algorithm is presented in full detail, in the language used in the recent interior-point literature; specifically, we borrow the notation used in [14]. In Section 3, assumptions are listed and convergence results are stated, including a local convergence result not found in [1]. In Section 4, preliminary numerical results are presented, Finally, Section 5 contains some concluding remarks. Throughout, $\|\cdot\|$ denotes the Euclidean norm.

2. Algorithm

In this section we describe the algorithm of [1] from an interior-point perspective. In connection with problem (P1), consider the logarithmic barrier function

$$\beta(x,\mu) = f(x) - \sum_{j=1}^{p} \mu_j \log d_j(x)$$

where $\mu = (\mu_1, \ldots, \mu_p)^T \in \mathcal{R}^p$ and the μ_js are positive. Note that the classical logarithmic barrier function is recovered when all μ_js are equal. While the algorithm we are about to describe was first introduced from a different perspective, in the interior-point context, the selection of different barrier parameters for different constraints can be thought of as allowing higher barrier values near the boundary of constraints closer to being active. The barrier gradient is given by

$$\nabla\beta(x, \mu) = g(x) - B(x)^T D(x)^{-1}\mu, \tag{2.1}$$

where, like in [14], g denotes the gradient of f, B the Jacobian of d and $D(x)$ the diagonal matrix $\mathrm{diag}(d_j(x))$.

Problem (P1) can be tackled via a sequence of unconstrained minimizations of $\beta(x, \mu)$ with $\mu \to 0$. Thus, in view of (2.1), $z = D(x)^{-1}\mu$ can be viewed as an approximation to the KKT multiplier vector associated with a solution of (P1) and the right-hand side of (2.1) as the value at (x, z) of the gradient (w.r.t. x) of the Lagrangian

$$\mathcal{L}(x, z) = f(x) - z^T d(x).$$

Accordingly, and in the spirit of primal-dual interior point methods, we consider using a (quasi-)Newton iteration for the solution of the system of equations in (x, z)

$$g(x) - B(x)^T z = 0, \tag{2.2}$$

$$D(x)z = \mu, \tag{2.3}$$

i.e.,

$$\begin{bmatrix} -W & B(x)^T \\ ZB(x) & D(x) \end{bmatrix} \begin{bmatrix} \Delta x \\ \Delta z \end{bmatrix} = \begin{pmatrix} g(x) - B(x)^T z \\ \mu - D(x)z \end{pmatrix} \tag{2.4}$$

where $Z = \mathrm{diag}(z_j)$ and where W is equal to, or approximates, the Hessian (w.r.t. x) of the Lagrangian $\mathcal{L}(x, z)$. Next, as in [1], we assume that W is symmetric and positive definite. When $\mu = 0$, a primal-dual feasible solution to (2.2)-(2.3) is a KKT point for (P1). Moreover, it turns out that, given (x, z) primal-dual feasible, the primal direction Δx^0 obtained by setting $\mu = 0$ is a descent direction for f at x. In [1], such a property is sought for the search direction. On the other hand, the components of μ should be positive enough to prevent the primal steplength from collapsing, but small enough that the fast local convergence properties associated with the (quasi-)Newton iteration for (2.2)-(2.3) with $\mu = 0$ be preserved. This is achieved in [1] by selecting

$$\mu_j = \rho z_j \|\Delta x^0\|^\nu, \tag{2.5}$$

with $\rho \in (0, 1]$ as large as possible subject to the constraint

$$\nabla f(x)^T \Delta x \leq \theta \; \nabla f(x)^T \Delta x^0,$$

where $\nu > 2$ and $\theta \in (0,1)$ are prespecified.[1]

In [1] the line search criterion includes a decrease of f and strict feasibility. It involves a second order correction $\Delta\tilde{x}$ to allow a full (quasi-)Newton step to be taken near the solution. With index sets I and J defined by

$$I = \{j : d_j(x) \le z_j + \Delta z_j\},$$

$$J = \{j : z_j + \Delta z_j \le -d_j(x)\},$$

$\Delta\tilde{x}$ is the solution of the linear least squares problem

$$\min \frac{1}{2}\Delta\tilde{x}^T W \Delta\tilde{x} \text{ s.t. } d_j\left(x+\Delta x\right) + \nabla d_j\left(x\right)^T \Delta\tilde{x} = \psi, \quad \forall j \in I \tag{2.6}$$

where

$$\psi = \max\left(\|\Delta x\|^\tau, \max_{j\in I}\left|\frac{\Delta z_j}{z_j + \Delta z_j}\right|^\kappa \|\Delta x\|^2\right), \tag{2.7}$$

where $\tau \in (2,3)$ and $\kappa \in (0,1)$ are prespecified. If $J \neq \emptyset$ or $\|\Delta\tilde{x}\| > \|\Delta x\|$, $\Delta\tilde{x}$ is set to 0. Note that I estimates the active index set and that J should be empty near the solution when strict complementarity holds. An (Armijo-type) arc search is then performed as follows: given $\beta \in (0,1)$, compute the first number α in the sequence $\{1, \beta, \beta^2, \dots\}$ such that

$$f\left(x + \alpha\Delta x + \alpha^2\Delta\tilde{x}\right) \le f\left(x\right) + \xi\alpha\nabla f(x)^T \Delta x \tag{2.8}$$

$$d_j\left(x + \alpha\Delta x + \alpha^2\Delta\tilde{x}\right) > 0, \quad \forall j \tag{2.9}$$

$$d_j\left(x + \alpha\Delta x + \alpha^2\Delta\tilde{x}\right) \ge d_j(x), \quad j \in J \tag{2.10}$$

where $\xi \in (0, 1/2)$ is prespecified and where the third inequality is introduced to prevent convergence to points with negative multipliers. The next primal iterate is then set to

$$x^+ = x + \alpha\Delta x + \alpha^2\Delta\tilde{x}.$$

Finally, the dual variable z is reinitialized whenever $J \neq \emptyset$; if $J = \emptyset$ the new value z_j^+ of z_j is set to

$$z_j^+ = \min\{z_{\max}, \max\{z_j + \Delta z_j, \|\Delta x\|\},$$

where $z_{\max}$ is a prespecified (large) number. Thus z_j^+ is allowed to be close to 0 only if $\|\Delta x\|$ is small, indicating proximity to a solution.

It is observed in [1, Section 5] that the total work per iteration (in addition to function evaluations) is essentially one Cholesky decomposition of size p and one Cholesky decomposition of size $\hat{p}$, the number of active constraints at the solution.[2]

---

[1] Note that Δx depends on ρ affinely.

[2] There are two misprints in [1, Section 5]: in equation (5.3) (statement of Proposition 5.1) as well as in the last displayed equation in the proof of Proposition 5.1 (expression for λ_k^0), $M_k B_k^{-1}$ should be $B_k^{-1} M_k$.

3. Convergence

Let $X = \{x : d(x) \geq 0\}$; for $x \in X$, let $I(x) = \{j : d_j(x) = 0\}$; and let x_0 be a strictly feasible initial guess, i.e., $d(x_0) > 0$. In addition to existence of a strictly feasible x_0, we assume the following.

1. $X \cap \{x : f(x) \leq f(x_0)\}$ is bounded.
2. For all $x \in X$, the vectors $\nabla d_j(x)$, $j \in I(x)$, are linearly independent.
3. The sequence $\{W_k\}$ of Hessian values/approximations is bounded and uniformly positive definite.
4. The set of points $x \in X$ such that, for some $z_1, \ldots, z_p \in \mathcal{R}$,

$$\nabla_x \mathcal{L}(x, z) = 0,$$

$$z_j d_j(x) = 0, \quad j = 1, \ldots, p$$

(with no restriction on the signs of the z_js) is finite.[3]

Under these assumptions, all accumulation points of the primal sequence $\{x_k\}$ generated by the algorithm are KKT points.

Suppose now that there is an accumulation point x^* where the second order sufficiency conditions with strict complementarity hold. Suppose further that

$$\frac{\left\| N_k^T \left(W_k - \nabla_{xx}^2 \mathcal{L}\left(x^*, \lambda^*\right)\right) N_k (\Delta x)_k \right\|}{\|(\Delta x)_k\|} \to 0 \quad \text{as } k \to \infty$$

where

$$N_k = I - \hat{B}_k^T \left(\hat{B}_k \hat{B}_k^T \right)^{-1} \hat{B}_k$$

with

$$\hat{B}_k = [\nabla d_j(x_k) \text{ s.t. } j \in I(x^*)]^T \in \mathcal{R}^{|I(x^*)| \times n}.$$

Also suppose $z_{\max}$ is larger than all KKT multipliers at x^*. Then the sequence of primal vectors converges two-step superlinearly to x^*. In particular, the full (quasi-)Newton step is eventually accepted by the line search criterion. Moreover, the sequence of barrier parameter vectors converges to 0, the sequence of dual vectors converges to the associated KKT multipliers, J is eventually empty, and I is eventually the index set of active constraints at x^*.

[3] Such points are referred to in [1] as *stationary points*.

Not shown in [1] is the fact that, if the sequence $\{W_k\}$ satisfies the stronger condition that

$$\frac{\left\|N_k^T\left(W_k-\nabla_{xx}^2\mathcal{L}\left(x^*,\lambda^*\right)\right)N_k(\Delta x)_k\right\|}{\|(\Delta x)_k\|^2}$$

is bounded as $k\to\infty$, then convergence is (at least) two-step quadratic.

Finally, stronger convergence results hold for a variation of the present algorithm, under weaker assumptions, in the LP and convex QP cases (see [5]). In particular, global convergence to the solution set X^* takes place whenever X^* is nonempty and bounded, X has a nonempty interior, and for every $x\in X$ the gradients of the active constraints at x are linearly independent.

4. Numerical Examples

In [1], an initial guess for the dual variable provided by the user is employed both at the start and whenever z is reinitialized because $J\neq\emptyset$ (see end of Section 2 above). The barrier parameter vector μ is then automatically initialized according to (2.5). Instead, when running our numerical tests, we used an adaptation of a scheme used in [14] to the present case of a vector barrier parameter. Given the current primal variable x, first obtain a least squares solution to (2.2)

$$z^{\mathrm{ls}}=\mathrm{argmin}\ \|g(x)-B(x)^Tz\|$$

and, based on (2.3), let

$$\mu^{\mathrm{ls}}=D(x)z^{\mathrm{ls}}.$$

Next "clip" the barrier parameters,

$$\mu_j=\min\{100,\max\{1,|\mu_j^{\mathrm{ls}}|\}\}$$

and select z to again satisfy (2.3), i.e.,

$$z=D(x)^{-1}\mu.$$

Note however that (2.5) will not hold in general if both z and μ are prescribed. This hurdle is cleared by introducing a factor $c_j>0$ in the right-hand side of (2.5) and to reset c_j to an appropriate value every time z and μ have been initialized or reinitialized. As a side point, following a recommendation made in [1], in (2.5) we replace $\left\|\Delta x^0\right\|^\nu$ with $\min\left(.01\left\|\Delta x^0\right\|,\left\|\Delta x^0\right\|^\nu\right)$ to improve the behavior of the algorithm in the early iterations. Thus in the present implementation, (2.5) is replaced with

$$\mu_j=\rho c_j z_j\min\left(.01\left\|\Delta x^0\right\|,\left\|\Delta x^0\right\|^\nu\right)\qquad(4.1)$$

where c_j is kept constant between initializations/reinitializations of z and μ and is reset to satisfy (4.1) with $\rho=1$ whenever z and μ are initialized/reinitialized.

Finally, we again followed [14] in our choice of a stopping criterion, i.e., the algorithm stops when

$$\left\| \begin{bmatrix} g(x) - B(x)^T z \\ \mu - D(x) z \end{bmatrix} \right\| \leq \text{tol} \tag{4.2}$$

and

$$\max_i \mu_i \leq \text{tol} \tag{4.3}$$

where tol > 0 is preselected.

The algorithm thus specified was coded in MATLAB 5.0 and applied to a subset of the Hock and Schittkowski test set [15] selected based on the identification of problems with positive definite Hessians and inequality constraints. The values for the parameters used in the algorithm are listed in Table 1. Exact Hessian information was used throughout. The initial guess for the primal variable was taken from [15], except for problem 65 in which the initial vector given in [15] is not strictly feasible. Following [14], $x_0 = [0, 0, 0]$ was selected instead. The tests were run on a Pentium-Pro 200MHz PC. Results are given in Table 2 where the first columns give the problem number from [15], the second column the number of iterations, and the third column the final value of the objective function. Our results compare well with those given in [14], [9], and [10].

Parameter	Value
ξ	.3
β	.8
θ	.8
ν	3
τ	2.5
κ	.5
$z_{\max}$	1e8
tol	1e-8

TABLE 1. Parameter Values

5. Concluding Remarks

An algorithm first proposed over a decade ago was described in the language of primal-dual interior point methods. This algorithm can be viewed as involving a vector of barrier parameters. It enjoys global convergence and a superlinear or quadratic local convergence rate. Promising preliminary numerical results were given.

Problem Name	# of Iters	Function Value
HS-12	6	-30.000
HS-35	7	0.1111
HS-43	9	-44.000
HS-65	10	0.9535
HS-76	9	-4.6818
HS-100	11	680.6301
HS-113	17	24.3062

TABLE 2. Results on Hock-Schittkowski Test Problems

REFERENCES

1 E.R. Panier, A.L. Tits, and J.N. Herskovits. A QP-free, globally convergent, locally superlinearly convergent algorithm for inequality constrained optimization. *SIAM J. Contr. and Optim.*, 26(4):788–811, July 1988.

2 S. Segenreich, N. Zouain, and J.N. Herskovits. An optimality criteria method based on slack variables concept for large structural optimization. In *Proceedings of the Symposium on Applications of Computer Methods in Engineering*, pages 563–572, Los Angeles, California, 1977.

3 J.N. Herskovits. *Développement d'une Méthode Numérique pour l'Optimization Non-Linéaire*. PhD thesis, Université Paris IX - Dauphine, Paris, France, January 1982.

4 J.N. Herskovits. A two-stage feasible directions algorithm for nonlinear constrained optimization. *Math. Programming*, 36(1):19–38, 1986.

5 A.L. Tits and J.L. Zhou. A simple, quadratically convergent algorithm for linear and convex quadratic programming. In W.W. Hager, D.W. Hearn, and P.M. Pardalos, editors, *Large Scale Optimization: State of the Art*, pages 411–427. Kluwer Academic Publishers, 1993.

6 A.V. Fiacco and G.P. McCormick. *Nonlinear Programming: Sequential Unconstrained Minimization Techniques.* Wiley, New-York, 1968.

7 E. Polak and A.L. Tits. On globally stabilized quasi-newton methods for inequality constrained optimization problems. In R.F. Drenick and E.F. Kozin, editors, *Proceedings of the 10th IFIP Conference on System Modeling and Optimization — New York, NY, August-September 1981*, volume 38 of *Lecture Notes in Control and Information Sciences*, pages 539–547. Springer-Verlag, 1982.

8 M. Argaez and R.A. Tapia. On the global convergence of a modified augmented Lagrangian linesearch interior point Newton method for nonlinear programming. Technical Report 95-38, Department of Computational and Applied Mathematics, Rice University, Houston, Texas, February 1997.

9 R.H. Byrd, J.C. Gilbert, and J. Nocedal. A trust region method based on interior point techniques for nonlinear programming. Technical Report OTC 96/02, Argonne National Laboratory, Argonne, Illinois, 1996.

10 R.H. Byrd, M.E. Hribar, and J. Nocedal. An interior point algorithm for large-scale nonlinear programming. Technical report, Northwestern University, Evanston, Illinois, 1996.

11 A.R. Conn, N.I.M. Gould, and Ph.L. Toint. A primal-dual algorithm for minimizing a nonconvex function subject to bounds and nonlinear constraints. Technical Report RC 20639, IBM T.J. Watson Center, Yorkstown Heights, New York, 1996.

12 A.S. El-Bakry, R.A. Tapia, T. Tsuchiya, and Y. Zhang. On the formulation and theory of the newton interior-point method for nonlinear programming. *J. Opt. Theory Appl.*, 89:507–541, 1996.

13 A. Forsgren and P.E. Gill. Primal-dual interior methods for nonconvex nonlinear programming. Technical Report NA 96-3, Department of Mathematics, University of California, San Diego, California, 1996.

14 D.M. Gay, M.L. Overton, and M.H. Wright. A primal-dual interior method for nonconvex nonlinear programming. Technical Report 97-4-08, Computing Sciences Research Center, Bell Laboratories, Murray Hill, NJ, July 1997.

15 W. Hock and K. Schittkowski. *Test Examples for Nonlinear Programming Codes*, volume 187 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, 1981.