アルゴリズム 11－7
（1989. 10. 24）

凸計画問題に対する解析的中心法の計算複雑度

F. Jarre
Universität Würzburg

アブストラクト

この論文では、Sonnenvend (1985) によって提案された解析的中心法を凸計画問題に適用した時のインプリメンテーションについて述べる。Jarre (1988) では、直線探索を用いた時のこの方法の理論的計算複雑度が議論されている。本論文では、高次の外挿手法を用いた方法を二つ導入し、適当な高次の外挿法を用いて加速した時に、この解析的中心法の効率がもっとも良くなることを示す。計算機実験によって、実際に必要な反復回数が、最悪の場合の評価よりかなり少ないことを示す。

COMPUTATIONAL COMPLEXITY OF THE METHOD OF ANALYTIC CENTERS FOR CONVEX PROGRAMMING

F. Jarre
Institut für Angewandte Mathematik und Statistik
Universität Würzburg, Am Hubland
8700 Würzburg, West-Germany
Februar 1989

ABSTRACT

This work examines implementational aspects of the method of analytic centers due to Sonnevend (1985) when applied to solve a class of convex programs – where the objective and also the constraints are given by convex functions whose Hessian matrices fulfill a relative Lipschitz condition with a Lipschitz constant $M \geq 0$. In Jarre (1988) the theoretical properties of a zero order variant of the method are examined: the existence of a two-sided ellipsoidal approximation for the set of feasible points around its center and a proof, that the zero order variant of the algorithm starting from an initial center of the feasible set generates a sequence of stricly feasible points whose objective function values converge to the optimal value. However. this zero order algorithm is of theoretical interest only.
Here we introduce two extrapolation schemes to predict the next center and show. that full efficiency of the method can only be obtained when accelerating it by some (higher order) extrapolation scheme. We further present some numerical experiments which indicate. that by using extrapolation the speed of convergence of the method is considerably faster than the worst case bound of $O(\sqrt{m}|\ln \epsilon|)$ iterations.

In this talk we first present the basic ideas of the method of analytic centers due to Sonnevend (1985). For a zero order variant of this method already the early works of Renegar (1986) and Gonzaga (1987) presented theoretical proofs of a low computational complexity when solving linear programs. In Jarre, Sonnevend and Stoer (1988) it is shown in some preliminary numerical experiments, that higher order variants of the method of analytic centers yield considerably faster algorithms, which are not only of low theoretical complexity but also interesting for practical applications. Here, we will emphasize, that the results obtained for solving linear programs also hold if the method is applied to solve a class of convex programs. We consider the class where the objective and also the constraints are given by twice continously differentiable convex functions whose Hessian matrices fulfill a relative Lipschitz condition with a Lipschitz constant $M \geq 0$. More precisely the problem under study is to find

$$\lambda^* := \min\{f_0(x) \mid x \in P\}.$$

$$\text{Here} \quad P := \{x \in \mathbb{R}^n \mid f_i(x) \leq 0 \quad \text{for} \quad 1 \leq i \leq m\}$$

is supposed to be bounded and to have a nonempty interior P^o, and the $f_i \in C^2(P)$ are smooth convex functions for $0 \leq i \leq m$ whose Hessian matrices fulfill

$$\exists M \geq 0 \quad \forall y, y+h \in P^o \quad \text{and} \quad \forall z \in \mathbb{R}^n :$$

$$|z^T(D^2 f_i(y+h) - D^2 f_i(y))z| \leq M\|h\|_y z^T D^2 f_i(y) z$$

$$\text{where } \|h\|_y^2 := h^T\left(\sum_i \frac{D^2 f_i(y)}{-f_i(y)} + \frac{Df_i^T(y) Df_i(y)}{f_i^2(y)}\right)h.$$

This condition is affine invariant and invariant under scaling of the functions f_i. It can be restricted to values of h with $\|h\|_y \leq 0.5$ and in this modified form the condition allows certain singularities of the constraint functions at the boundary of the feasible domain. (The function $f_1(y) = -\sqrt{y}$ for example fulfills the condition with $M = 8$ for $y \in (0, \infty)$.) A similar condition has also been suggested by Nesterov and Nemirowsky (1988). We mention, that the two-sided ellipsoidal approximation of the feasible set which is known for the linear and the quadratic case (see Jarre (1987), Sonnevend and Stoer (1988)) can be generalized, i.e. that there exist two similar ellipsoids E_{in} and E_{out} both centered at the same point such that

$$E_{in} \subset \{x \in \mathbb{R}^n \mid f_i(x) \leq 0 \text{ for } 1 \leq i \leq m\} \subset E_{out},$$

and E_{in} and E_{out} have a similarity ratio of $O((1+M^{4/3})m)$. A fixed portion of the interior ellipsoid turns out – as in the linear case – to lie in the domain of superlinear convergence of Newton's method for computing the center. Using this result one can give a proof showing, that the zero order variant of the method of analytic centers starting from an initial center of the feasible set generates a sequence of stricly feasible points whose objective function values converge to the optimal value. Concerning the speed of convergence for a problem with m constraints one can show (see Jarre (1988)), that an upper bound for the

gap in between the objective function value and the optimal value is reduced by a factor of ϵ within $O((1+M^2)\sqrt{m}|\ln\epsilon|)$ iterations. Here, each iteration involves the computation of one Newton step. If all constraint functions and the objective function are linear or quadratic, then the Lipschitz constant for the Hessians $D^2 f_i$ is $M = 0$ and the bound of $O(\sqrt{m}|\ln\epsilon|)$ Newton iterations to guarantee an error reduction by a factor of ϵ in the objective function is as good as the ones currently given for *linear* programs. However, this zero order algorithm is of theoretical interest only.
In this talk, we will introduce two extrapolation schemes to predict the next center and present some preliminary numerical experiments for problems with linear and quadratic constraints. The experiments show, that full efficiency of the method can only be obtained when accelerating it by some (higher order) extrapolation scheme. The numerical experiments further indicate, that the speed of convergence is considerably faster than the worst case bound of $O(\sqrt{m}|\ln\epsilon|)$ iterations, and that one can expect almost the same computational effort for solving an arbitrary convex program satisfying above conditions as needed to solve a (dense) linear program of the same size.

Outline of the method:
For $\lambda > \lambda^*$, let $P(\lambda)$ denote the feasible set P intersected with the level set $\{x|\ f_0(x) \le \lambda\}$ where the objective function is "better" than λ,

$$P(\lambda) := P \cap \{x|\ f_0(x) \le \lambda\}.$$

Assume without loss of generality that the objective function $f_0(x) = c^T x$ is a linear function. (If not one can introduce an artificial variable x_{n+1} and an additional constraint $f_0(x) - x_{n+1} \le 0$ and minimize x_{n+1}.) The method follows a homotopy path $\lambda : \infty \to \lambda^*$ of some interior point in $P(\lambda)$ which is "easily" computable and depends smoothly on all constraints: A very convenient point (Sonnevend (1985)) is the *analytic center* $x(\lambda)$ of $P(\lambda)$. In Jarre (1988) it is shown, that for each parameter $\lambda > \lambda^*$ the analytic center $x(\lambda)$ of $P(\lambda)$ exists and is the unique point x in $P(\lambda)^o$ minimizing the strictly convex logarithmic barrier function

$$\varphi(x,\lambda) := -\ln(\lambda - f_0(x)) - \sum_{i=1}^{m} \ln(-f_i(x)).$$

It is *invariant* under affine transformations of P and *invariant* under scaling of the functions f_i. It is further shown, that if a point $y \in P(\lambda)^o$ is "sufficiently close" to $x(\lambda)$, then Newton's method starting at y to minimize φ will converge superlinearly to $x(\lambda)$.

Here we will examine possibilities to find a first estimate (extrapolation) $\hat{y}_{k+1}$ for $x(\lambda_{k+1})$ using a given approximation $y_k \approx x(\lambda_k)$ for some values $\lambda_k > \lambda_{k+1} > \lambda^*$. $\hat{y}_{k+1}$ is to be close enough to $x(\lambda_{k+1})$ such that Newton's method starting from $\hat{y}_{k+1}$ will converge in a few steps to a point y_{k+1} which is up to mashine accuracy equal to $x(\lambda_{k+1})$.

The analytic center $x = x(\lambda)$ of $P(\lambda)$ is characterized by the equation

$$D_x\varphi(x,\lambda) =: g(x,\lambda) = \frac{c^T}{\lambda - c^T x} + \sum_{i=1}^{m} \frac{Df_i(x)}{-f_i(x)} = 0 \qquad (*)$$

One can show (Jarre 1988), that $H(x,\lambda) := D_x g(x,\lambda)$ is positive definite if P^o is nonempty and bounded. Differentiating $(*)$ with respect to λ yields

$$D_x g(x,\lambda)\frac{dx}{d\lambda} + \frac{\partial g(x,\lambda)}{\partial \lambda} = H(x,\lambda)\frac{dx}{d\lambda} - \frac{c}{(\lambda - c^T x)^2} = 0.$$

Defining

$$G(x,\lambda) := H(x,\lambda)\ (\lambda - c^T x)^2$$

we get

$$x'(\lambda) = \frac{dx}{d\lambda} = G^{-1}(x,\lambda)c.$$

The vector $x'(\lambda)$ is thus easily computable and can be used for extrapolation, as done for example by Adler, Resende and Veiga (1986), Fiacco and McCormick (1968), Gonzaga (1987), or Karmarkar (1984). Yet in the case of higher order extrapolation it might be advantageous to use an extrapolation scheme based on the knowledge of the previous centers rather than a scheme based on the Taylor expansion of the curve $x(\lambda)$, see Sonnevend (1986).

Model algorithm:

Let $y_0 = x(\lambda_0)$ be given for some $\lambda_0 > \lambda^*$. Set $k = 0$. Define $\Delta_0 := (\lambda_0 - c^T y_0)/(100\sqrt{m})$.

step 1: Let $\lambda_{k+1} := \lambda_k - \Delta_k$.

step 2: Define $\hat{y}_{k+1} := y_k - \Delta_k G(y_k,\lambda_k)^{-1}c$ as first approximation to $x(\lambda_{k+1})$. (*linear extrapolation*).

step 3: Test whether $\hat{y}_{k+1}$ is in $P(\lambda_{k+1})^o$ and whether it is a "sufficiently good" approximation to $x(\lambda_{k+1})$. If not let $\Delta_k := \gamma_k \Delta_k$ with $0 < \gamma_k < 1$ (*reduce step length*) and goto 1. Else goto 4.

step 4: Test whether $\hat{y}_{k+1}$ is 'too close' to $x(\lambda_{k+1})$. If so (and if the step length had not just been reduced), let $\Delta_k := \kappa_k \Delta_k$ with $\kappa_k > 1$ (*enlarge step length*) and goto 1. Else goto 5.

step 5: Use Newton's method with starting vector $\hat{y}_{k+1}$ to minimize $\varphi(x,\lambda_{k+1})$ i.e. to approximate $x(\lambda_{k+1})$ more closely. Let the result be y_{k+1}.

step 6: Test for convergence, e.g. check whether $\frac{5}{4}|\lambda_{k+1} - c^T y_{k+1}| < \epsilon$. If so, stop, else let $\Delta_{k+1} := \Delta_k$ (try to keep the old step length) and $k := k+1$ and goto 1.

Comments

In step 3 the vector $\hat{y}_{k+1}$ is "sufficiently good" if it can be expected that Newton's method for finding the minimum $x(\lambda_{k+1})$ of $\varphi(x,\lambda_{k+1})$ gives a sufficiently precise approximation

to $x(\lambda_{k+1})$ within few iterations if $\hat{y}_{k+1}$ is used as starting point. This can be expected if $\hat{y}_{k+1} \in P^o(\lambda_{k+1})$ and if the gradient $\hat{g}_{k+1} := D_x \varphi(x, \lambda_{k+1})\big|_{x=\hat{y}_{k+1}}$ is small, say

$$d_{k+1}\|\hat{g}_{k+1}\| \leq \gamma \max_{0 \leq i \leq m} \|Df_i(y_k)\|. \qquad (**)$$

with $d_{k+1} := \min\{(\lambda_{k+1} - f_0(\hat{y}_{k+1})), f_i(\hat{y}_{k+1}) \quad 1 \leq i \leq m\}$ and $\gamma = 0.1$. (See also Jarre, Sonnevend and Stoer (1988).) Similarly the test in step 4 will check whether the gradient $\hat{g}_{k+1}$ is too small, say whether (**) holds with $\gamma = 0.01$. As a consequence of Jarre (1987), the quantity $\frac{5}{4}|\lambda_{k+1} - c^T y_{k+1}|$ used in the stopping test in step 6 is an upper bound for the error $c^T y_{k+1} - \lambda^*$.

In some numerical experiments we also investigated higher order extrapolation schemes which are based on the knowledge of the previous centers as suggested by Sonnevend (1986). The linear extrapolation in step 2 of the model algorithm is then modified in the following way. Let $\lambda_{k-j}, \lambda_{k-j+1}, ..., \lambda_k$ and $y_{k-j}, ..., y_k$ be given as best approximations to $x(\lambda_{k-j}), ..., x(\lambda_k)$ computed so far. Fix λ_{k+1} as in step 1 of the model algorithm and construct an interpolating function $\hat{x}$ with

$$\hat{x}(\lambda_{k-l}) = y_{k-l}, \quad l = 0, 1, ..., j$$

and define $$\hat{y}_{k+1} := \hat{x}(\lambda_{k+1})$$

The choice of $j \leq k$ determines the order of extrapolation.
Other extrapolation schemes may be based on a parameter different from λ and the value of λ_{k+1} might not be given explicitly. This however does not cause any difficulties, since (in case of a linear ojective function f_0) for any point $\hat{x}_{k+1}$ in P one can easily determine a parameter λ such that the norm of the gradient $\hat{g}_{k+1}$ is minimal.

NUMERICAL EXPERIMENTS

To gain insight into the practical behaviour of the method and the possibilities of acceleration offered by different types and orders of extrapolation, it has been tested on small size problems (with up to 300 unknowns and dense matrices). In particular this implementation is intended to illustrate the *global* convergence behaviour and the dependence of the number of iterations on the dimension of the problem.

Example 1
Within these examples we show the influence of the order j of interpolation on the performance of the method. $j + 1$ is the number of "analytic centers" y_l, $k - j \leq l \leq k$, used for extrapolation. The results refer to a randomly chosen problem of the form

$$\min_{x \in P} -\sum_{j=1}^{n} x_j$$

$$P = \{x \geq 0 : Ax \leq 10^4 e,\}$$

where A is an $n \times n$ matrix whose entries are random numbers uniformly distributed between 1 and 1000 and $e = (1,\ldots,1)T \in \mathbb{R}^n$, taken from Iri and Imai (1986). Here, $n = 10$. The objective function was counted 10 times, mainly to improve the performance of the zero order method. The results show, that it is advantageous to use higher than first order interpolation, thereby exploiting the smoothness of the curve $x(\lambda)$ of analytic centers. Using the order $j = 0$ for extrapolation roughly corresponds to a variant of a method used by Renegar (1986) and $j = 1$ or 2 seems to be comparable to the ones of Adler Resende and Veiga (1986). This table however cannot serve as a basis for a comparison of the methods, as the underlying algorithm is laid out for large step extrapolation rather than efficient updates of the Hessian matrix which are easier to devise in the case of short step lengths.

TABLE 1 **The influence of the order j of extrapolation.**

j	#it	#Hess	$f_0(x_{end}) - \lambda^*$
0	556	308	$3.7 \cdot 10^{-6}$
1	34	34	$3.0 \cdot 10^{-6}$
2	28	29	$2.7 \cdot 10^{-6}$
3	27	27	$4.7 \cdot 10^{-6}$
4	26	26	$2.1 \cdot 10^{-6}$
5	26	27	$1.5 \cdot 10^{-6}$
5^*	22	23	$5.2 \cdot 10^{-6}$

All examples were based on extrapolation by polynomials $\hat{x}(\lambda)$ in the parameter λ, with the exception of the last example in row 5^*, where polynomials $\hat{x}(t)$ in the parameter $t = r(\lambda)/(r(\lambda) + \rho)$ were used, which of course corresponds certain rational functions in r. Here the proper choice of ρ is essential for fast convergence. From other experiments we obtained an acceptable value of $\rho \approx c^T x(\lambda_0) - \lambda^*$, which can be obtained from an ellipsoidal approximation of P using the first center. The stopping test always was $c^T x - \lambda^* \leq 10^{-5}$. The results reported here for $n = 10$ are typical: similar results were obtained for other dimensions n as well.

Example 2
This example is a problem with a convex quadratic objective function and linear constraints. The objective function is given by $f_0(x) := x^T Q x + c^T x$ where Q is a diagonal dominant tri-diagonal-matrix with random entries, and c a uniformly distributed vector. The constraints are as in example 1. Each problem was run three times with different random numbers, the average is listed in the table. The number of constraints was chosen $m = 2n$.

TABLE 2 **Quadratic objective function (Phase 2).**

n	#it	#Hess	#grad	ω
10	14	15	153	0.584
20	17	20	229	0.586
30	16	21	249	0.536

Example 3
Another set of test problems with nonlinear constraints was given by:

$$f_i(x) := x^T Q_i x + b_i^T x + \gamma_i$$

where the Q_i are diagonal dominant tri-diagonal-matrices with random entries. the b_i are uniformly distributed vectors and the γ_i are negative real numbers. Each problem was run three times with different random numbers. the average is listed in the table. The number of constraints was chosen $m = n/3$.

TABLE 3 Quadratic constaints (Phase 2).

n	#it	#Hess	#grad	ω
10	10	12	161	0.202
20	12	12	159	0.329
30	12	12	151	0.305
60	13	15	157	0.346
100	13	14	153	0.328

FINAL REMARK

It is shown. that the algorithm besides its nice theoretical properties is also interesting for practical applications. in particular when accelerating the method by some higher order extrapolation. In the description of the implementation we tried to preserve the conceptual simplicity of the method and therefore suppressed many minor details (how to choose the constants. step size correction. or when to update the Hessian matrix) which nevertheless turned out to be useful for improving the running times of the algorithm. Finding good error bounds for higher order extrapolating functions is a difficult open problem. Since an error bound is crucial for the proof of the full efficiency of the above method. we cannot give a tight complexity bound for the accelerated method yet. But in all numerical examples higher order extrapolation gave a remarkable speedup such that the total number of iterations seems to be almost independent from the dimension. This observation supports the basic idea of Sonnevend's method, to use a suitable high order extrapolation scheme for following a single path of analytic centers.

REFERENCES

I. Adler, M.G.C. Resende and G. Veiga, "An Implementation of Karmarkar's Algorithm for Linear Programming," ORC Report, Dept. of Ind. Engin. and OR, University of California. (Berkeley, CA. 1986).

A.V. Fiacco and G.P. Mc Cormick, "Sequential Unconstrained Minimization Techniques." Nonlinear Programming, (Whiley, New York, 1968)

C.C. Gonzaga, "An Algorithm for Solving Linear Programming Problems in $O(n^3L)$ Operations," Memo No. UCB/ERL M87/10, College of Engineering, University of California, (Berkeley, CA, 1987).

M. Iri and H. Imai, "A Multiplicative Barrier Function Method for Linear Programming," Algorithmica 1 (1986) 455 – 482.

F. Jarre, "On the convergence of the method of analytic centers when applied to convex quadratic programs," report No. 35, Schwerpunktprogramm Anwndungsbezogene Optimierung und Steuerung, Inst. für Ang. Math. und Statistik, Universität Würzburg (1987).

F. Jarre, "On the method of analytic centers for solving smooth convex programs," report No. 67, Schwerpunktprogramm Anwndungsbezogene Optimierung und Steuerung, Inst. für Ang. Math. und Statistik, Universität Würzburg, submitted to Lecture Notes in Mathematics in revised form (1988).

F. Jarre, G. Sonnevend, J. Stoer, "An Implementation of the Method of Analytic Centers," in: A. Benoussan, J.L. Lions eds., Lecture Notes in Control and Information Sciences 111 (New York: Springer 1988) pp.297 – 307.

N. Karmarkar, "A New Polynomial-Time Algorithm for Linear Programming," Combinatorica 4 (1984) 373 – 395.

V.Klee and G.J.Minty, "How good is the simplex algorithm?," O. Sisha, ed. Inequalities III (Academic Press, New York, 1972).

S. Mehrotra and J. Sun, "An Interior Point Algorithm for Solving Smooth Convex Programs Based on Newton's Method," technical report 88-08, Dept. of IE/MS, Northwestern University (Evaston IL, 1988).

J.E. Nesterov and A.S. Nemirovsky, "A General Approach to Polynomial-Time Algorithms Design for Convex Programming," report, Central Economical and Mathematical Institute, USSR Acad. Sci. (Moscow, USSR, 1988).

J. Renegar, "A Polynomial Time Algorithm Based on Newton's Method for Linear Programming," report MSRI 07 118-86, Mathematical Sciences Research Institute (Berkeley, CA. 1986).

G. Sonnevend, "An Analytical Centre for Polyhedrons and New Classes of Global Algorithms for Linear (Smooth, Convex) Programming," in: Lecture Notes in Control and Information Sciences 84 (1985) pp. 866 – 876.

G. Sonnevend, "A New Method for Solving a Set of Linear (Convex) Inequalities and its Applications," 5^{th} IFAC-IFORS Conf. Budapest, Dept. of Num. Anal., Inst. of Math., L. Eötvös University, (Budapest, Hungary 1986).