ON THE SOLUTION OF NONLINEAR OPTIMIZATION PROBLEMS OF HIGH DIMENSION

Otu Vaarmann

Tallinn Technical University
Department of Mathematics
Ehitajate tee 5, 19086 Tallinn
vaarmann@edu.ttu.ee

Many optimization problems in economics are of the multiobjective type and highdimensional. Possibilities for solving large-scale optimization problems on a computer network or multiprocessor using a multi-level approach are studied. For the solution of involving decomposition-coordination problems a few rapidly convergent methods are developed, their convergence properties and computational aspects are examined. Problems of their global implementation and polyalgorithmic approach are discussed.

Introduction

A number of problems in economics, engineering and scientific computational (production planning, process control, image restoration, parameter identification, neural network, inverse problems etc.) lead frequently to a large mathematical programming problem

$$\min\{f(x) : x \in Q\}, \tag{1}$$

where Q is a closed subset of R^n. It also contains the problem of finding a fixed point of a nonlinear mapping F, i.e.

$$F(x) = 0 \tag{2}$$

with

$$f(x) = \|F(x)\|^2 \text{ and } Q = R^n, \tag{3}$$

where F is acting between spaces R^n and R^m. On the other hand, the problem of finding an extremum for constrained optimization problem is frequently reduced by means of Lagrange multipliers or penalty functions to seeking stationary points of certain unconstrained functions (functionals).

Arguably, many socio-ecological and industrial optimization problems are of the multiobjective type. Nowadays the large and complex systems are composed of many parts interacting in a more or less complicated way. One of the most complicated problems is the technical, economical and social criteria-based estimation of the construction of new buildings or the renovation of old ones. It is worth noting that even for bicriteria problems, both generating methods and interactive approaches are computationally much more complicated and costly than methods for a single criteria optimization.

In order to cope with large scale problems and to develop many optimum plans a multi-level approach may be useful. The idea of hiearchical decision making is to reduce the overall complex problem into smaller and simpler approximate problems (subproblems) which can be distributed over a larger number of processors. One way to break a problem into smaller subproblems is the use of decomposition-coordination schemes, i.e. by designating the computers (processors) as the master and slaves. Computation of proper values for coordination parameters in a convex programming leads frequently to solving an auxiliary optimization problem or a system of nonlinear equations

$$H(y, \beta) = 0, \tag{4}$$

where $H = (H_1, ..., H_m)^T$ and $\beta = (\beta_1, ..., \beta_m)^T$ while the components of the vector $y = (y_1, ..., y_n)^T$ are to be determined as a solution of nonlinear problems

$$F_i(y_i, \beta) \to \min, \ y_i \in \Gamma_i(\beta), \tag{5}$$

depending on the parameter vector β and where F_i is the performance index of the i-th subproblem and $\Gamma_i(\beta) \subset R^{n_i}$ is its feasible region [1]. Under the assumption that the problems (5) have the solutions $y_i = y_i(\beta)^*$, we shall study the problem for determining the vector $\beta^* \in R^m$ from the equation

$$H(y(\beta), \beta) = 0. \tag{6}$$

Further on, we shall reformulate the problem (6) into the form (2) more suitable for mathematics where $F(x) \equiv H(\cdot, x)$ and x will stand for the desired parameter vector.

Decomposition-coordination problems have some specific features:

· the user has at his disposal only functional values;

· the evaluation of functional values includes, basically, the solution of certain subproblems and therefore it can cause a great computational effort;

· the functions involved are not necessarily differentiable, they may belong to a set of almost differentiable functions [2].

Besides the problem (6) may be ill-conditioned or even ill-posed.

Similar problems arise on the occasion of methods which require with the help of a generating method to assess a vector of weights showing the relative importance of the different criteria. Attribution of numerical values to the weights always implies some degree of arbitrariness and uncertainity.

Therefore, for problems based on a multi-level approach sophisticated algorithms are needed which try to find a trade-off between robustness, stability and efficiency. Methods with the high order of convergence making full use of local information (e.g. functional values, gradient and Hessian) permit sometimes to win in speed and accuracy.

Computational effort is often one of the basic problems in the solution of real-life problems. The total cost of an iterative method is determined by the number of iterations needed to achieve the required accuracy and the cost of each iteration. Implementation of methods with the high order of convergence require for computing a solution with the prescribed accuracy, as a rule, less iterations than methods with a lower convergence order and therefore likely less a total arithmetic.

As to stability which is another important aspect of computation the use of methods with high order of convergence may relieve the stability problem as well. Whereas even very rough approximation to the operator of second derivatives in the methods with the convergence order $p \geq 3$ may provide their numerical stability [3,4] then it as reasonable to develop methods based on a quadratic model

$$F(x_c + d) \approx F(x_c) + B(x_c)d + G(x_c)dd,$$

where x_c denotes the current iterative point, d the increment of the argument, $B(x) \approx F'(x)$ and $G(x) \approx F''(x)$.

METHODS

In this section we assume that F is acting from a Banach space X into another Y. One of the most popular methods of order three for (2) is the method of tangent hyperbolas (or Chebyshev-Halley method)

$$x_{k+1} = x_k - T_k^{-1}\Gamma_k F(x_k), \quad k = 0, 1, ..., \tag{7}$$

where $\Gamma_k = [F'(x_k)]^{-1}$ and $T_k = I - \frac{1}{2}\Gamma_k F''(x_k)\Gamma_k F_(x_k)$.

It can be rewritten as

$$x_{k+1} = x_k - [F'(x_k) - \frac{1}{2}F''(x_k)\Gamma_k F(x_k)]^{-1}F(x_k). \tag{8}$$

Since in problems, arising in mathematical modelling and simulation procedural and rounding errors are unavoidable then the study of methods with approximate operators may give more realistic impression of methods under discussion. In further discussion, in principal, it does not matter what is the origin of approximation but for concreteness, the approach adopted by this report is the use mainly of iterative methods to obtain approximations to T_k^{-1} and/or Γ_k or approximate solutions to the corresponding linear equations.

If A_k and $L(x, x-y)$, $x, y \in X$, approximate the operators Γ_k and the term $F''(x_k)(x-y)$ respectively, then due to $x_k - y_k = A_k F(x_k)$ and $L(x_k, A_k) = L(x_k, A_k F(x_k)) \approx F''(x_k) A_k F(x_k)$ it follows from (3) that

$$x_{k+1} = x_k - U_k^{-1} A_k F(x_k), \tag{9}$$

where $U_k = A_k F'(x_k) - \frac{1}{2} A_k L(x_k, A_k)$.

If in turn instead of U_k^{-1} to use its approximation V_k we get the method

$$x_{k+1} = x_k - V_k A_k F(x_k). \tag{10}$$

Further on we shall suppose the existence and boundedness of the operators $[F'(x_k)]^{-1}$ and U_k^{-1}. Likewise we assume the existence of such constants $\alpha, \beta, \lambda, \Lambda, \mu, M, C, C_1, K, G, G_1 < \infty$ and sequences γ_{1k} and γ_{2k} such that the following inequalities are valid

$$\begin{gathered} \| F'(x) \| \leq M, \ \| F''(x) \| \leq K, \ \| A_k \| \leq \mu_k \leq \mu, \ \| V_k A_k \| \leq \lambda_k \| F(x_k) \| \leq \lambda \| F(x_k) \|, \\ \| V_k \| \leq \Lambda_k \leq \Lambda, \ \| A_k^{-1} \| \leq \beta_k \leq \beta, \ \| I - U_k V_k \| \leq \gamma_{1k}, \\ \max\{\| I - A_k F'(x_k) \|, \ \| I - F'(x_k) A_k \|\} \leq \gamma_{2k}. \end{gathered} \tag{11}$$

Theorem 1. Let $x_0 \in S$, $S = \{x \in X : \| x - x_0 \| \leq \wp\}$ and the following conditions are valid on S:
1^o *operator F is twice Frechet-differentiable;*
2^o *operator of the second derivatives satisfies a Lipschitz-condition* $\| F''(x) - F''(y) \| \leq L_2 \| x - y \|$;
3^o $\| F''(x)(x-y) - L(x, x-y) \| \leq G \| x - y \|^2$, $\| L(x, x-y) \| \leq G_1 \| x - y \|$;
4^o *there exist* $\Gamma(x)$ *and* $U^{-1}(x)$ *with* $\| \Gamma(x) \| \leq C$ *and* $\| U^{-1}(x) \| \leq C_1$;
5^o $\delta = \delta_0^{(i)} < 1$, $i = 1, 2, 3$ *(the quantity δ is defined in different ways in the cases 1) - 3)).*

Then 1) If $\gamma_{ik} \leq \gamma_{i0} < 1, i = 1, 2,$ and $r_1 = \lambda \| F(x_0) \| / (1 - \delta) \leq \wp$, then the equation (2) has a solution x^* in S $\| x^* - x_0 \| \leq r_1$, to which the sequence (10) converges with

$$\| x_k - x^* \| \leq r_1 \delta^k, \ \delta = \delta_0^{(1)};$$

if $\gamma_{i0} \geq \gamma_{i1} \geq \gamma_{i2} \geq ... \geq \gamma_{in} \geq ... \geq 0$, and $\gamma_{ik} \to 0$, as $k \to \infty$, then $\delta_k^{(1)} \to 0$, and the sequence (10) converges superlinearly with

$$\| x_k - x^* \| \leq r_1 \prod_{m=0}^{k-1} \delta_m^{(1)},$$

where $\delta_k^{(1)} = \beta_k \mu_k \gamma_{1k} + \frac{1}{2} \gamma_{1k} \mu_k \lambda_k K \| F(x_k) \| + \frac{1}{2} \gamma_{2k} \lambda_k^2 K \| F(x_k) \| + \frac{1}{4} \mu_k^2 \lambda_k^2 K G_1 \| F(x_k) \|^2 + + \frac{1}{6} \lambda_k^3 L_2 \| F(x_k) \|^2$.

2) If $\gamma_{1k} = C_2 \| F(x_k) \|$, $\gamma_{2k} \leq \gamma_{20}$, $(\gamma_{20}, \ C_2 < \infty)$, $\delta = \delta_0^{(2)} = d_0^{(2)} \| F(x_0) \| < 1$,
$d = \lim_{k \to \infty} d_k^{(2)} > 0$, $d_k^{(2)} = \beta_k \mu_k + \dfrac{1}{2} \gamma_{20} \lambda_k^2 K + \dfrac{1}{2} (\mu_k \lambda_k K C_2 + \dfrac{1}{2} \mu_k^2 \lambda_k^2 K G_1 + \dfrac{1}{3} \lambda_k^3 L_2) \| F(x_k) \|$, *then the equation (2) has a solution x^* in S, $\| x^* - x_0 \| \leq r_2$, to which the sequence (10) converges quadratically*

$$\| x_k - x^* \| \leq \lambda H_k^{(2)}(\delta)/d, \ H_k^{(2)}(\delta) = \sum_{i=k}^{\infty} \delta^{2^i}.$$

3) If $\gamma_{1k} = C_3 \parallel F(x_k) \parallel^2,\ \gamma_{2k} = C_4 \parallel F(x_k) \parallel^2,\ C_3, C_4 < \infty$ *and* $r_3 = H_0^{(3)}(\delta)/d \leq \wp$, *where* $H_k^{(3)}(\delta) = \sum_{i=k}^{\infty} \delta^{3^i},\ \delta = \delta_0^3 = \sqrt{d_0} \parallel F(x_0) \parallel < 1,\ d = d_0^{(3)} = \beta_0\lambda_0 C_3 + \frac{1}{2}(C_3\mu_0\lambda_0 + C_4\lambda_0^2)K + \frac{1}{2}(\lambda_0^3 G + \frac{1}{2}\mu_0\lambda_0 G_1 K + \frac{1}{3}\lambda_0^3 L_2)$,
then the sequence (10) converges cubically

$$\parallel x_k - x^* \parallel \leq (\lambda/\sqrt{d}) H_k^{(3)}(\delta).$$

The proof of this Theorem 1 rests on a more general theorem from [5].

Particularly, approximating the term $F''(x_k)A_kF(x_k$ by the expression $L(x_kA_k) = 2[F'(x_k) - F'(x_k - \frac{1}{2}A_kF(x_k))]$, the method (8) becomes

$$x_{k+1} = x_k - [F'(x_k - \frac{1}{2}A_kF(x_k))]^{-1}F(x_k), \tag{12}$$

because of $F'(x_k) - \frac{1}{2}F''(x_k)\Gamma_kF(x_k) \approx F'(x_k) - \frac{1}{2}L(x_k, A_k) = F'(x_k - \frac{1}{2}A_kF(x_k))$.

Let W_k be an operator which approximates $[F'(x_k - \frac{1}{2}A_kF(x_k))]^{-1}$, then W_k can be written as $W_k = V_kA_k$ with $V_k \approx U_k^{-1}$ because of $[F'(x_k - \frac{1}{2}A_kF(x_k))]^{-1} = A_k^{-1}[A_kF(x_k) - \frac{1}{2}A_kL(x_k, A_k)]^{-1} = = U_k^{-1}A_k$.

If $\parallel I - U_kV_k \parallel \leq \gamma_{1k} = O(\parallel F(x_k) \parallel^2)$ and $\parallel I - F'(x_k)A_k \parallel \leq \gamma_{2k} = O(\parallel F(x_k) \parallel)$ then by Theorem 1 the method (12) is cubically convergent. In particulan case $A_k = \Gamma_k$ the method (12) coincides with the midpoint method

$$x_{k+1} = x_k - [F'(x_k - \frac{1}{2}\Gamma_kF(x_k))]^{-1}F(x_k). \tag{13}$$

To get derivative free methods of the type (13) one can modify it as follows

$$x_{k+1} = x_k - [F(2u_k - x_k; x_k)]^{-1}F(x_k), \tag{14}$$

where $F(v; w)$ denotes the first order divided difference with the basic elements ν and w, $u_k = = x_k - \frac{1}{2}B_kF(x_k)$ and $B_k = [F(2x_k - x_{k-1}; x_k)]^{-1}$ or $B_k = [F(2x_k - u_{k-i}; x_k)]^{-1}$.

Besides the midpoint method can be used for equations with nondifferentiable operators [6].

Another possibility to avoid the evaluation of F'' and thereby to reduce computational costs is to replace it by a fixed bilinear operator

$$x_{k+1} = x_k - [I - \frac{1}{2}A_k\Phi A_kF(x_k)]^{-1}A_kF(x_k), \tag{15}$$

where $\Phi : X \times X \to Y$ is a general bounded bilinear operator.

The execution of one iteration step by the formula (15) is equivalent to solving two perturbed linear equations

$$[F'(x_k) + V_k](y_k - x_k) = -F(x_k),$$

$$[F'(x_k) + V_k](x_{k+1} - y_k) = -\frac{1}{2}\Phi(y_k - x_k)^2,$$

where $V_k = A_k^{-1} - F'(x_k)$ and therefore the method (15) has similar computational costs as Newton method. It is shown in [7] that the method (15) with $A_k = \Gamma_k$ remains faster the Newton method.

Using in the approximate method of tangent parabolas (or Euler-Chebyshev method)

$$x_{k+1} = x_k - A_kF(x_k) - \frac{1}{2}A_kF''(x_k)(A_kF(x_k))^2 \tag{16}$$

for approximating the term $F''(x_k)(A_kF(x_k))^2$ the expression

$$2[F(x_k - A_kF(x_k)) - F(x_k) - A_k^{-1}(x_k - A_kF(x_k)) - x_k]$$

we get

$$x_{k+1} = x_k - A_k F(x_k) - A_k F(x_k - A_k F(x_k)) \tag{17}$$

for which the rate of approximation $O(\| F(x_k) \|)$ for Γ_k is sufficient to obtain convergence order $p = 3$ [5].

The derivative free variant of (17)

$$y_k = x_k - [F(2y_{k-1} - x_{k-1}; x_{k-1})]^{-1} F(x_k) \tag{18}$$

$$x_{k+1} = y_k - [F(2y_k - x_k; x_k)]^{-1} F(y_k) \tag{19}$$

has the asymptotic convergence order equal to 3 provided the second derivative F'' is Lipschitz continuous and corresponding divided differences are Lipschitz continuous [5].

The procedure (18), (19) requires little information per an iteration: two values of F and one values of the divided difference (except for the first iteration), i.e. computational effort is comparable with that of Newton method.

GLOBAL IMPLEMENTATION. POLYALGORITHMIC STRATEGY

For today, there are lot of methods having the high order of convergence $p > 2$, but in practice they are relatively little exploited. This is partially due to the fact that computational schemes of execution of one iteration are laborous, they require frequently the evaluation of derivatives of order greater than one and a good initial guess since their advantages become evident in the close vicinity

Practically to obtain a method that is robust, stable and computationally convenient and efficient, at the some time, is not a trivial task. Since none of existing methods has all the above-listed characteristics then we propose to use polyalgorithmic strategy.

The property of global convergence is a criterion for robustness. One of the most effective ways to guarantee the global convergence or at least greatly expand the domain of convergence is the "continuation strategy". According to this, firstly $F(x) = 0$ must be replaced by a one-parameter family of problems $G(x, \lambda) = 0,\ \lambda \in [o, 1]$, such that $F(x) = G(x, 1)$ and the solution of $G(x, 0) = 0$ is known. Secondly, a series of problems must be solved, where the parameter λ is slowly varied. But all the homotopy methods suffer from the disadvantage that Jacobian at some points may become singular. Therefore the implementation of methods with the convergence order $p \geq 3$ in conjunction with the continuation strategy may be justified. Recall that continuous methods converge globally but slowly, whereas the iterative methods with high order of convergence convergence converge locally. These features can be combined in such a way that a continuous method is used, if necessary, to help get into the domain of convergence of the rapidly convergent method, which, then, will be turned on to improve the accuracy. But the functions involving in decomposition-coordination schemes may be nonsmooth and therefore can cause serious computations for numerical methods. One possibility to handle equations with nonsmooth functions is to approximate the locally Lipschitzian function with a smooth one and to use the derivative of the smooth function in algorithm (e.g. in an extension of the Levenberg-Marquardt method as suggested in [8]) whenever a derivative is needed. In the paper [9] a trust region method for solving nonsmooth equations subject to linear constraints are proposed. With the choice of polyhedral norm to define the trust region and merit function, this method solves a sequence of linear programs.

Another reasonable polyalgorithmic strategy for decomposition-coordination problems is the use a derivative free method with the convergence order $p > 2$ if works, otherwise, to switch over to a modification of the Newton method with finite-difference approximation of the Jacobian. If it does not work either we try to implement the damped Newton-like method or the continuation strategy. If no the method put to the test does not work, we have to take advantage of a slower but a more global method. We choose a method based on the steepest descent direction provided the problem is smooth at the current iterative point, otherwise, we use a method based on subgradients (e.g. methods from [2]) or the approximational gradient. As shown in [10], methods based on the approximational gradient allow to handle many nonsmooth and discontinuous problems successfully. Analogs of the steepest descent and the conjugate gradient methods based on the approximational gradient are studied in [10]. After accomplishing a fixed member of iterations by the global method we attempt to start with a high order method once again.

ACKNOWLEDGEMENT

Finanical support from the Estonian Science Foundation (grant Nr. 5006) is greatly appreciated.

References

1. S. Ulm. *Metody Dekompozitsii dlya Resheniya Zadach Optimizatsii.* Valgus, Tallinn, 1979 (in Russian).
2. N. Shor. *Metody Minimizatsii Nedifferentsiryemykh Funktsii i ikh Prilozheniya.* Naukova Dumka, Kiev, 1979 (in Russian).
3. G.-P. Ehle and H. Schwetlick. *Discretized Euler-Chebyshev multistep method.* SIAM J. Numer. Anal., 1976, 13, Nr. 3, 433–443.
4. B. Döring. *Iterative Lösung Gewisser Randwertprobleme und Integralgleichungen.* Applicae Matematiky, 1978, B. 23, 1–31.
5. O. Vaarmann. *On high order iterative methods for nonlinear problems.* In: Nonlinear Mathematical Analysis and Applications (ed. Th. M. Rassias). Palm Harbor, FL; Hadronic Press, 1998, 275–299.
6. I. Argyros. *An error analysis for the midpoint method.* Tamkang Journal of Mathematics, 1999, 30, Nr. 2, 71–83.
7. J. A. Ezquerro and M. A. Hernandez. *Newton-type iterations as approximations of Chebyshev's method.* Southwest Journal of Pure and Applied Mathematics (Electronical Journal), Issue 1, July 1998, 44–55.
8. Li Qun Qi. *Trust region algorithms for solving nonsmooth equations.* SIAM J. Optim., 1995, 5, Nr. 1, 219–230.
9. A. Gabriel and Jong-Shi Pang. *A trust region method for constrained nonsmooth equations.* In: Large Scale Optimization: State of the Art. (eds. W. W. Hager et al). Kluwer Academic Publishers. Dordrecht, 1994, 155–181.
10. T. B. Bigil'deyeva and V. I. Rol'shikov. *Numerical methods for optimizing discontinuous functions.* J. Comput. Syst. Sci. Int., 1995, 33, Nr. 5, 49–56.