Journal of Engineering Physics and Thermophysics, Vol. 76, No. 2, 2003

EFFICIENT METHOD OF OPTIMIZATION OF PHYSICAL PROCESSES

V. K. Tolstykh

UDC 519.6:536

New extremum algorithms which provide a uniform convergence to an optimum value in the space of functions are suggested. Their distinctive feature lies in analytical determination of the gradient of the objective functional of the problem of optimization and in original control of the direction of descent of a relatively infinite-dimensional gradient.

The problems of optimization of physical processes — their optimum control, identification of their mathematical models, improvement of geometrical characteristics of processes — are of frequent occurrence. Direct extremum algorithms in both infinite-dimensional and finite-dimensional spaces obtained after finite-difference transformations of differential equations or after the expansion of optimized parameters in basis functions are the most general approaches to their solution. In contrast to the traditional principle of maximum, dynamic programming, and other methods of optimum control which are far from applicable to all partial-differential equations [1], in the direct approach one seeks the solution on the basis of direct minimization of the objective $J(u)$.

For direct optimization of distributed and nonstationary processes use is conventionally made of the gradient algorithms [2–4]

$$u^{k+1}(\tau) = u^k(\tau) + b^k p\,(u^k; \tau)\,, \quad b^k > 0\,, \quad k = 0, 1, 2, ..., \tag{1}$$

where $u^k(\tau)$ is the sought parameter function on the kth iteration, $p(u^k; \tau)$ is the antigradient, or the conjugate gradients of the objective functional $J(u)$, and b^k is the step of the method. Implementation of algorithm (1), especially in thermophysical processes, involves fundamental difficulties caused by the substantial nonlinearity of the processes, in particular, the processes of heat transfer. Finite changes in thermophysical parameters can lead to infinitesimal changes in ∇J. Moreover, there are no justifications of the convergence of infinite-dimensional algorithms of the type (1) during a finite number of iterations even for quadratic J. Therefore, (1) requires modernization.

As an example, we consider a one-dimensional process of heat transfer described by the parabolic equation in the region $(t, x) \in [t_a, t_b] \times [x_0, x_1]$

$$C\rho \frac{\partial T(t, x)}{\partial t} - \lambda \frac{\partial^2 T(t, x)}{\partial x^2} = 0\,, \quad \lambda \frac{\partial T}{\partial x}\bigg|_{x_0} = q\,, \quad \lambda \frac{\partial T}{\partial x}\bigg|_{x_1} = u\,, \quad T\big|_{t_a} = T_a\,. \tag{2}$$

At the boundary x_1, we must find the heat flux $u(t)$ which gives a minimum to the functional

$$J(u) = \int_{t_a}^{t_b} (T - T_*)^2\big|_{x_0}\, dt \to \min\,. \tag{3}$$

Applying the direct approach [3], we determine the gradient

$$\nabla J(u; t) = -f(t, x)\,, \quad (t, x) \in (t_a, t_b) \times x_1\,,$$

Donetsk National University, Donetsk, Ukraine; email: tvk@dongu.donetsk.ua. Translated from Inzhenerno-Fizicheskii Zhurnal, Vol. 76, No. 2, pp. 160–162, March–April, 2003. Original article submitted February 5, 2002; revision submitted September 3, 2002.

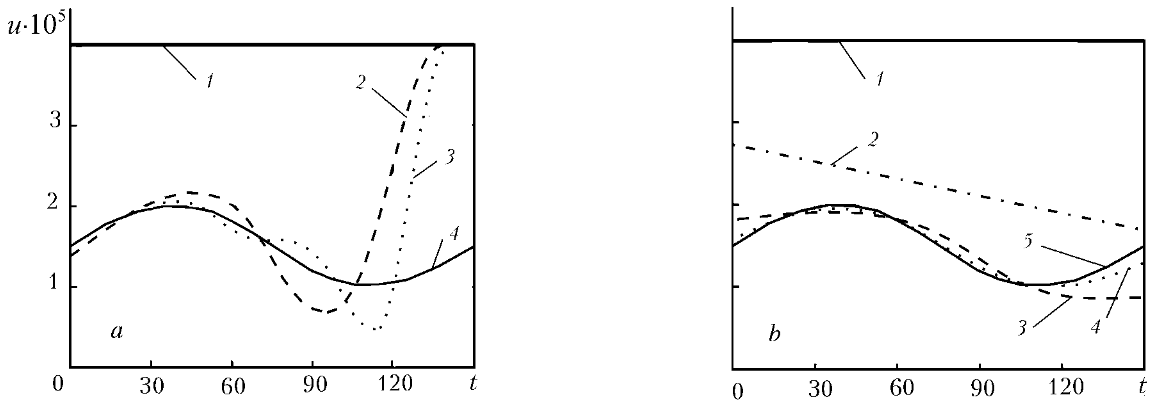

Fig. 1. Optimization of the heat flux: a) method of conjugate gradients and L-BFGS [1) u^0; 2) u^{10}; 3) u^{37}; 4) u_*]; b) method (4) based on conjugate gradients [1) u^0; 2) φ; 3) u^{20}; 4) u^{50}; 5) $u^{75} = u_*$]. u, J/(m^2·sec); t, sec.

where $f(t, x)$ is the solution of the conjugate problem

$$C\rho \frac{\partial f}{\partial t} + \lambda \frac{\partial^2 f}{\partial x^2} = 0 , \quad \lambda \frac{\partial f}{\partial x}\bigg|_{x_0} = 2\,(T - T_*) , \quad \lambda \frac{\partial f}{\partial x}\bigg|_{x_1} = 0 , \quad f\big|_{t_b} = 0 .$$

Figure 1a illustrates an unsuccessful attempt at solving problem (2), (3) by the infinite-dimensional algorithm (1) with the direction of descent p selected according to the method of conjugate Polak–Ribiere gradients. The approximately time-uniform deviation of the heat flux $u^0(t)$ from the optimum one ($u_*(t)$) gives a substantially t-nonuniform (in particular, by seven orders of magnitude) value of the gradient and, as a consequence, a nonuniform convergence to u_*.

The same "poor" curve was obtained in expansion of the function $u(t) = \sum_{i=1}^{n-1} u_i B_i(t)$ in B-splines of zero order (step functions) with carriers equal to the step $\Delta t = (t_b - t_a)/n$ similarly to [5]. Here, the finite-dimensional control $u \in R^{n-1}$ ($n = 100$) was found by the quasi-Newtonian method with an economical memory L-BFGS [6, 7], where the inverse Hesse matrix (Hessian) was represented by five pairs of vectors from R^{n-1}.

We replace algorithm (1) in problem (2), (3) by the following algorithm:

$$u^{k+1}(t) = u^k(t) - b^k \alpha^k(t)\, p\,(u^k; t) , \quad b^k > 0 , \quad k = 0, 1, 2, ..., \tag{4}$$

where the function $\alpha(t)$ controls the convergence $u^k \to u_*$, according to the necessary optimality condition [3], in the form

$$\nabla J\,(u^k; t) \to 0 \tag{5}$$

uniformly in t for $u^k \to u_*$.

Algorithm (4) with the implementation of condition (5) allows one to solve infinite-dimensional problems of optimization during a finite number of iterations without transformation of controls to finite vectors, as is often done [2, 4, 5].

The problems of practical implementation of method (4) lie in selection of the function $\alpha^k(t)$ in order to satisfy condition (5). Here, we consider one way of approximate implementation of (5) on the initial iterations.

We introduce the notion of pattern approximations. Let the first approximation $u^1(t)$ be a certain *a priori* known function $\varphi(t)$ (pattern function) whose gradient $\nabla J(\varphi; t)$ satisfies condition (5) for $k = 1$ (i.e., uniformly decreases after the first iteration). In this case, from (4) we can find

TABLE 1. Convergence of the Algorithms

Method	Iteration k, 0	Functional $J(u^k)$, $4.11 \cdot 10^4$	Residual $\|\|u^k - u_*\|\|$, $3.05 \cdot 10^6$
Conjugate gradients	20	1.82	$1.23 \cdot 10^6$
	37	$5.62 \cdot 10^{-1}$	$1.17 \cdot 10^6$
L-BFGS	10	$5.04 \cdot 10^{-1}$	$1.18 \cdot 10^6$
	24	$1.49 \cdot 10^{-1}$	$1.10 \cdot 10^6$
Method (4) based on the gradient	20	1.33	$2.66 \cdot 10^5$
	50	$1.61 \cdot 10^{-1}$	$1.59 \cdot 10^5$
	75	$5.69 \cdot 10^{-2}$	$8.59 \cdot 10^4$
	9110	$6.47 \cdot 10^{-7}$	$6.05 \cdot 10^3$
The same based on conjugate gradients	20	$9.39 \cdot 10^{-2}$	$7.28 \cdot 10^4$
	50	$4.61 \cdot 10^{-4}$	$4.51 \cdot 10^4$
	75	$2.07 \cdot 10^{-4}$	$2.93 \cdot 10^4$
	324	$1.59 \cdot 10^{-6}$	$4.28 \cdot 10^3$

$$\alpha^0 (t) = \left| \frac{\varphi (t) - u^0 (t)}{\nabla J (u^0; t)} \right| , \quad \nabla J (u^0; t) \neq 0 \ \forall t , \quad t \in (t_a, t_b) .$$

On the following iterations, the parameter $\alpha(t)$ does not change. In the present method, a researcher has to make several trial first iterations in order to select the appropriate pattern function $\varphi(t)$ satisfying (5).

Figure 1b shows the solution of problem (2), (3) with a fundamentally new result of the uniform convergence to $u_*(t)$. To implement (5), we have selected $\varphi = (0.8 + 0.0033t)u^0$.

Table 1 gives the characteristics of convergence of all four algorithms studied. Everywhere, to find the step b^k we used the algorithm of one-dimensional minimization that is based on the Wulf condition with cubic interpolation and is the most efficient at present [6]. The convergence of the studied algorithms was assumed to be completed when the relative change in the functional and the control was less than 10^{-10}. In the table, the last iteration of each method corresponds to this accuracy.

It is seen from the results of the calculations that the new methods based on algorithm (4) minimize the functional J five to six orders of magnitude better than the traditional methods. They also make it possible to approach the optimum solution of u_* three orders of magnitude closer. It is necessary to note that the traditional methods considerably decrease the functional J (by five orders of magnitude in calculations), but, in this case, they do not converge to the exact solution. This means that if the optimization algorithms do not provide a uniform convergence, as algorithm (4) can do, one cannot judge the quality of optimization by a decrease in the objective functional. Unfortunately, the authors of most publications restrict themselves simply to this unsatisfactory evaluation.

Thus, for optimization of physical processes we can suggest algorithm (4), which demonstrates good convergence in the space of controls.

NOTATION

u, optimized parameter which enters into the equations of the process; τ, space-time variable; x, space variable; t, time; J, objective functional of the optimization problem; p, direction of descent; b, parameter of length of a descent step; k, iteration number; ∇J, gradient of the objective functional (linear functional); C, ρ, and λ, heat capacity, density, and thermal conductivity, respectively; T and T^*, temperature and optimum temperature of the process; q, heat flux; f, conjugate variable (linear functional); α, parameter of control of the minimization direction; u^0 and u_*, initial approximation and optimum value of the sought parameter; u^1, u^{10}, u^{20}, u^{37}, u^{50}, u^{75}, and u^k, terms of iterations; φ, pattern approximation to a minimum; Δt, step of a finite-difference grid; n, number of grid meshes; R^{n-1}, finite-dimensional Euclidean space of optimized parameters of dimensionality $n - 1$; B, B-spline (Base spline). Sub- and superscripts: 0 and 1, left and right boundaries of the one-dimensional process; a and b, start and end of the process time.

REFERENCES

1. A. I. Egorov, *Optimum Monitoring of Thermal and Diffusion Processes* [in Russian], Moscow (1978).
2. O. M. Alifanov, E. A. Artyukhin, and S. V. Rumyantsev, *Extremum Methods for Solving Ill-Posed Problems and Their Applications to Inverse Problems of Heat Transfer* [in Russian], Moscow (1988).
3. V. K. Tolstykh, *Direct Extremum Approach to Optimization of Systems with Distributed Parameters* [in Russian], Donetsk (1997).
4. F. L. Chermous'ko and I. V. Banichuk, *Variational Problems of Mechanics and Control* [in Russian], Moscow (1973).
5. R. H. W. Hoppe and S. I. Petrova, "Applications of the Newton Interior-Point Method for Maxwell's Equations," in: M. Deville and R. Owens (eds.), *Proc. 16th IMACS World Congr. on Sci. Computat., Appl. Math. and Simulat.*, Lausanne, Switzerland (2000). CD ROM ISBN 3-9522075-1-9.
6. J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York (1999).
7. C. T. Kelly, *Iterative Methods for Optimization*, SIAM, Philadelphia (1999).