Quadratic Internal Model Principle in Mathematical Programming

Neculai Andrei[1]

1. Introduction. Starting with an initial point x_0 every algorithm for solving the general nonlinear optimization problem

$$\begin{aligned} &\min f(x) \\ &\text{subject to} \\ &h(x) = 0, \end{aligned} \tag{1}$$

where $f: R^n \to R$ and $h: R^n \to R^m$, can be considered as a generator of a sequence of points $\{x_k\}$ which satisfy the constraints of the problem in such a way that $f(x_k) \to f(x^*)$, where x^* is a local solution of the problem. The *line search methods* are characterized by two main actions. At the iteration k a search direction d_k is generated and then a suitable point $x_k + \alpha_k d_k$ is computed by a step length α_k so that a reduction of the minimizing function or of a merit function (a penalty function) is obtained. The main action in any optimization algorithm is the design of the generator of directions d_k. The steplength is computed using the standard procedures of Armijo or of Wolfe in order to reduce the values of the function f or of a merit function. Plenty of nonlinear optimization algorithms are known and there are a lot of papers and books presenting them from the viewpoint of theoretical and computational aspects.

To solve problem (1), or a more general version of it with inequality constraints, each optimization algorithm must "understand" it. There is a large diversity of optimization algorithms. Many of them solve a constrained optimization problem by converting it to a sequence of unconstrained problems via Lagrangian multipliers or via penalty or barrier functions. Another class of methods solves nonlinear programming problems by moving from a feasible point to a new improved one along a feasible direction. However, every optimization algorithm, in one way or another, is based on the Karush-Kuhn-Tucker optimality conditions. Generally, these conditions are expressed as a nonlinear algebraic system. In the framework of the Newton machine this nonlinear system is reduced to a sequence of linear algebraic systems, which is equivalent to a sequence of quadratic programming problems. The quadratic internal model principle in mathematical programming states that "***an optimization algorithm must encapsulate implicitly or explicitly a quadratic internal model of the problem to be solved***". Every optimization algorithm uses its own quadratic internal model which takes into account the main ingredients defining the algorithm. This is the minimal part that must be encapsulated by the algorithm in order to solve the problem.

The philosophical motivation behind the quadratic internal model principle in mathematical programming is as follows. As known, the mathematical model of a physical reality is based on the conservation laws. In physics, a conservation law states that a particular measurable property of an isolated system does not change while the system evolves. Any particular conservation law is a mathematical identity to certain symmetry of a physical

[1] Research Institute for Informatics, Center for Advanced Modeling and Optimization, 8-10, Averescu Avenue, Bucharest 1, Romania. Member of Academy of Romanian Scientists, 54, Splaiul Independentei, Bucharest 5, Romania. E-mail: nandrei@ici.ro

system. For systems which obey the principle of the least action and therefore have a Lagrangian [Lagrange, 1797], [Fourier, 1798] the Noether's theorem [Noether, 1971] expresses the equivalence between conservation laws and the invariance of physical laws with respect to certain transformations called symmetries. The behavior of a physical system can often be expressed in terms of a specific function of the system variables, called Lagrangian. The system follows a path through the phase space such that the integral of the Lagrangian is stationary. For a system with Lagrangian L of the variables q and $\dot{q} = dq/dt$ the equation of motion is

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}}\right) = \frac{\partial L}{\partial q}.$$

From this equation Noether specified that if the quantity on the right hand term is zero (meaning that L is symmetrical over q), then the rate of change of the quantity in parentheses on the left side is also zero, i.e. it is a conserved quantity. Generally, any symmetry of the Lagrangian function corresponds to a conserved quantity, and vice versa. It seems that at the fundamentals of our cognoscible universe lies the concept of symmetry. But, mathematically symmetries are expressed by quadratic forms – a homogeneous polynomial of degree two in a number of variables. It is worth saying that the quadratic forms are central objects in mathematics and they are ubiquitous in physics and chemistry. Quadratic forms occur in number theory, Riemannian geometry, Lie theory and they always express energy of a system, particularly in relation to the L^2 norm, which leads us to the use of the concept of Hilbert spaces. Therefore, it is quite natural to see that at the heart of every mathematical model is a quadratic form. This quadratic form must be replicated in an optimization algorithm in order to get a solution of the corresponding problem.

2. Quadratic Internal Model Principle. In his synthesis, Yuan [2003] has shown that using the Newton machine for every method for solving a constrained optimization problem a linear algebraic system can be associated in a most natural way. All the linear systems corresponding to different methods are similar in form and can be expressed as:

$$\begin{bmatrix} W(x_k,\lambda^k)+T_k & -\nabla h(x_k)^T \\ \nabla h(x_k) & S_k \end{bmatrix}\begin{bmatrix} d \\ \eta \end{bmatrix} = -\begin{bmatrix} \nabla f(x_k)+\varepsilon_1 \\ h(x_k)+\varepsilon_2 \end{bmatrix}, \tag{2}$$

where $W(x_k,\lambda^k) = \nabla^2 f(x_k) - \sum_{i=1}^{m} \lambda_i^k \nabla^2 h_i(x_k)$ is the Hessian matrix of the Lagrange function $L(x,\lambda) = f(x) - \lambda^T h(x)$, $T_k \in R^{n\times n}$ is a symmetric matrix, $S_k \in R^{m\times m}$ is a null or a diagonal matrix whose elements are nonpositive, $\varepsilon_1 \in R^n$ and $\varepsilon_2 \in R^m$ are two vectors. In (2) d is the searching direction and η is an auxiliary vector which for some methods could be the Lagrange multiplier. In the following we shall consider two cases.

1) Let us assume that $S_k = 0$. Therefore, from (2) we get

$$\begin{bmatrix} W(x_k,\lambda^k)+T_k & -\nabla h(x_k)^T \\ \nabla h(x_k) & 0 \end{bmatrix}\begin{bmatrix} d \\ \eta \end{bmatrix} = -\begin{bmatrix} \nabla f(x_k)+\varepsilon_1 \\ h(x_k)+\varepsilon_2 \end{bmatrix}, \tag{3}$$

It is easy to see that the system (3) corresponds to the Newton method for optimization with equality nonlinear constraints or to the sequential quadratic programming method. The augmented system (3) can be considered as the necessary condition for d to be a solution of the following quadratic programming problem

$$\begin{aligned} &\min \frac{1}{2} d^T (W(x_k,\lambda^k)+T_k)d - (\nabla f(x_k)+\varepsilon_1)^T d \\ &\text{subject to} \\ &\nabla h(x_k)d + (h(x_k)+\varepsilon_2) = 0. \end{aligned} \tag{4}$$

It is well known that if $(W(x_k,\lambda^k)+T_k)$ is positive definite on the null space of $\nabla h(x_k)$ and $\nabla h(x_k)$ is a full-rank matrix, then the quadratic problem (4) has a unique global solution d. This solution can be obtained by solving the augmented system (3), where d is the solution of the problem and η is the Lagrange multiplier associated to the equality constraint. The problem (4) is *the quadratic internal model of problem (1) associated to the methods involving the linear system (3) (with $S_k = 0$).*

For example, *the quadratic internal model of the problem (1) corresponding to the Newton method* is:

$$\begin{aligned} &\min \frac{1}{2} d^T W(x_k,\lambda^k) d - (\nabla f(x_k) + \nabla h(x_k)\lambda^k)^T d \\ &\text{subject to} \\ &\nabla h(x_k) d + h(x_k) = 0. \end{aligned} \tag{5}$$

2) Let us suppose that $S_k \neq 0$. Normally, S_k is a diagonal matrix, whose diagonal elements are all negative. In this case it is easy to see that the system (2) corresponds to the following methods: the augmented Lagrange function, the inverse barrier function, the log-barrier function, the interior point algorithms, the path-following methods, the affine scaling interior point methods etc. From (2) we get

$$\eta = -S_k^{-1}\nabla h(x_k) d - S_k^{-1} h(x_k) - S_k^{-1}\varepsilon_2. \tag{6}$$

Therefore, using (6) in (2) it follows that

$$\begin{aligned} &\left[W(x_k,\lambda^k) + T_k + \nabla h(x_k)^T S_k^{-1}\nabla h(x_k)\right] d \\ &\qquad = -\nabla h(x_k)^T S_k^{-1} h(x_k) - \nabla h(x_k)^T S_k^{-1}\varepsilon_2 - \nabla f(x_k) - \varepsilon_1. \end{aligned} \tag{7}$$

But, (7) is equivalent with the following quadratic problem

$$\begin{aligned} &\min \frac{1}{2} d^T \left[W(x_k,\lambda^k) + T_k + \nabla h(x_k)^T S_k^{-1}\nabla h(x_k)\right] d \\ &\qquad -\left[-\nabla h(x_k)^T S_k^{-1} h(x_k) - \nabla h(x_k)^T S_k^{-1}\varepsilon_2 - \nabla f(x_k) - \varepsilon_1\right]^T d \end{aligned} \tag{8}$$

which is called *the quadratic internal model of problem (1) associated to the methods involving the linear system (2) (with $S_k \neq 0$).*

Therefore, an optimization algorithm for solving (1) must encapsulate a procedure for solving (in an iterative way) the quadratic internal model (8), which represents the essence of the problem from the view point of the algorithm involving (2).

Observe that the quadratic internal model of (1), as expressed by (8), is dependent on the algorithm we consider for solving the problem (1). In particular, for example, *the quadratic internal model of problem (1) corresponding to the augmented Lagrange function method,* in which the augmented Lagrange function is

$$L(x,\lambda,\sigma) = f(x) + \lambda^T h(x) + \frac{1}{2}\sigma \left\| h(x) \right\|_2^2, \tag{9}$$

where $\lambda \in R^m$ is the Lagrange multiplier and $\sigma > 0$ is the penalty parameter, is:

$$\begin{aligned} &\min \frac{1}{2} d^T (W(x_k,\lambda^k h(x_k)) + \sigma \nabla h(x_k)^T \nabla h(x_k)) d - \\ &\qquad (-\nabla f(x_k) + \nabla h(x_k)^T (\lambda^k - \sigma h(x_k))) d. \end{aligned} \tag{10}$$

In this case $\eta \in R^m$ is given by $\eta = \sigma \nabla h(x_k) d + \sigma h(x_k)$.

It is worth saying that for the unconstrained problem ($m = 0$) the Newton step can be obtained by solving the following quadratic problem

$$\min \frac{1}{2} d^T \nabla^2 f(x_k) d + \nabla f(x_k)^T d \tag{11}$$

which is *the quadratic internal model of the problem* $\min f(x)$ *corresponding to the Newton method*. Of course, the Newton step $d = -(\nabla^2 f(x_k))^{-1} \nabla f(x_k)$ is obtained by solving the linear system $(\nabla^2 f(x_k))d = -\nabla f(x_k),$ but as we know, it comes from the quadratic problem (11). Similarly, we can say that for the unconstrained problem the quasi-Newton step can be obtained by solving the following quadratic problem

$$\min \frac{1}{2} d^T B_k d + \nabla f(x_k)^T d \tag{12}$$

which is *the quadratic internal model of the problem* $\min f(x)$ *corresponding to the quasi-Newton method,* where B_k is a positive definite matrix satisfying the quasi-Newton equation $B_k(x_{k+1} - x_k) = \nabla f(x_{k+1}) - \nabla f(x_k).$

3. Conclusion. To solve a mathematical programming problem an algorithm must encapsulate in an implicitly or explicitly manner a quadratic internal model of the problem. This is the quadratic internal model principle in mathematical programming. This quadratic internal model reflects the ingredients of the algorithm and represents its essence. The philosophical support of this principle is coming from the Noether Theorem which expresses the equivalence between the conservation laws and symmetries which can be represented by quadratic forms. These quadratic forms are the fundamentals of every line search optimization algorithm.

References

Fourier, J.B.J., (1798) *Mémoire sur la statique, contenant la démonstration du principe des vitesses virtuelles et la théorie des moments*. Journal de l'École Polytechnique, tome 2, chaier 5, 1798.

Lagrange, J,-L., (1797) *Théorie des functions analytiques*. Imp. De la République, Paris, 1797.

Noether, E., (1971) *Invariante Varlationsproblems*. Nachrrichten der Königlichen Gesellschaft der Wissenschaften zu Göttingen, Math-phys. Klasse (1918), pp.235-257. [English translation, M.A. Travel, *Invariant Variation Problems*, Transport Theory and Statistical Physics 1 (3) 1971, pp.1183-207.

Yuan, Y., (2003) *Linear systems associated with numerical methods for constrained optimization.* Journal of Computational Mathematics, 21 (2003), pp.71-84.

April 23, 2009