On Quadratic Internal Model Principle in Mathematical Programming

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

Abstract. We show that for mathematical programming problems every optimization algorithm must encapsulate implicitly or explicitly a quadratic internal model of the problem to be solved which represents the essence of the problem from the view point of the algorithm. Optimization models are coming from the conservation laws. For systems which obey the principle of least action the Noether's theorem expresses the equivalence between the conservation laws and symmetries. But mathematically, symmetries are expressed by quadratic forms. Therefore, at the heart of every real optimization model is a quadratic form. The quadratic internal model principle says that this quadratic form modified in order to imbed the main ingredients of the optimization algorithm represents the quadratic internal model of the optimization algorithm.

MSC: 49M07, 49M10, 90C06, 65K
Keywords: Nonlinear optimization, line search methods, Newton system, quadratic programming, Noether theorem.

1. Introduction

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

$$\min f(x)$$
subject to
$$h(x) = 0,$$

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 the above problem, 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 [1].

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 system. For systems which obey the principle of the least action and therefore have a Lagrangian (see [8], [6]) the Noether's theorem [9] 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 lie 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 [13]. 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. This is the quadratic internal model principle in mathematical optimization.

The structure of this paper is as follows. In section 2 following the synthesis of Yuan [16] we present some ingredients showing that, in the context of Karush-Kuhn-Tucker theory, every optimization algorithm for solving general equality constrained optimization problems reduce to a nonlinear algebraic system. Using the Newton machine, this nonlinear system reduces further to a sequence of linear algebraic systems, known as Newton systems. We show that these Newton systems have the same structure, however being different for specific optimization methods. In section 3 the quadratic internal model principle for nonlinear optimization is presented, showing that at the heart of every optimization algorithm there is a quadratic model (quadratic programming problem) imbedding the main ingredients of the algorithm [1].

2. Linear systems used in algorithms for solving constrained optimization problems

In this section, following the synthesis of Yuan [16], we consider a number of methods for solving constrained optimization problems and show that the main ingredient into the corresponding algorithms of these methods is the solution of a linear algebraic system of equations with a special structure.

2.1. Newton method for nonlinear equality constraints

Let us consider the equality constrained problem:

$$\min f(x) \tag{1a}$$

$$h(x) = 0, \tag{1b}$$

where $f : R^n \to R$ and $h : R^n \to R^m$ are smooth nonlinear functions. For this problem, a point $x \in R^n$ is a regular point of the constraints if the vectors $\nabla h_i(x)$, $i = 1, \dots, m$ are linear independent. From the Karush-Kuhn-Tucker theory we know that at a local solution x^* of (1) which is a regular point there exists the Lagrange multipliers λ_i, $i = 1, \dots, m$, such that

$$\nabla f(x^*) - \nabla h(x^*)^T \lambda^* = 0, \tag{2a}$$

$$h(x^*) = 0. \tag{2b}$$

To have λ^* unique in (2), x^* must be a regular point of the problem. Therefore, the algorithms that solve the first order optimality conditions (2) require the assumption that all iterates x_k are regular points of the problem. Denote:

$$F(x, \lambda) = \begin{bmatrix} \nabla f(x) - \nabla h(x)^T \lambda \\ h(x) \end{bmatrix}. \tag{3}$$

For solving the system $F(x, \lambda) = 0$, let x_k be the current iterate point and λ^k be the corresponding Lagrange multiplier. The Lagrange-Newton step for the above nonlinear system is

$$\begin{bmatrix} W(x_k, \lambda^k) & -\nabla h(x_k)^T \\ \nabla h(x_k) & \end{bmatrix} \begin{bmatrix} d_k \\ \eta_k \end{bmatrix} = \begin{bmatrix} -\nabla f(x_k) + \nabla h(x_k)^T \lambda^k \\ -h(x_k) \end{bmatrix}, \tag{4}$$

where

$$W(x_k, \lambda^k) = \nabla^2 f(x_k) - \sum_{i=1}^{m} \lambda_i^k \nabla^2 h_i(x_k) \tag{5}$$

is the Hessian matrix of the Lagrange function $L(x, \lambda) = f(x) - \lambda^T h(x).$ Observe that the system (3) is linear in $\lambda.$ Therefore, the Newton method gives λ^* in one iteration. In fact, taking $x_k = x^*$ in (4) we have that $\lambda^k \to \lambda^*$ when $\{x_k\} \to x^*.$ This behavior of the Newton method for solving the system $F(x, \lambda) = 0$ is very important because to get the global convergence in (x, λ) variables we can consider a line search strategy based only on the x variables. Of course, since in (3) we need ∇h in order to evaluate $F,$ it seems we can substitute $W(x_k, \lambda^k)$ in (4) by its quasi-Newton approximation. Observe that if x_k is a regular point and $W(x_k, \lambda^k)$ or its quasi-Newton approximation is positive definite, then it can be proved that the solution of (4) is unique.

2.2. Sequential Quadratic Programming

Another method for solving the equality constrained problem (1) is the sequential quadratic programming method proposed by Wilson [15] and interpreted by Beale [2]. In this method the search directions are computed by minimizing a quadratic approximation of the objective function $f(x)$ subject to linear approximation of the constraints. For (1), in the current point x_k the quadratic programming approximation has the following form

$$\min \frac{1}{2} d_k^T B_k d_k + \nabla f(x_k)^T d_k, \tag{6a}$$

$$h(x_k) + \nabla h(x_k) d_k = 0, \tag{6b}$$

where $d_k \in R^n$ is the unknown and B_k is a quasi-Newton approximation of the Hessian $\nabla^2 f(x_k)$ [10]. The global convergence of this method is given by Han [7]. Powell [11] proved its superliniar convergence. Solving this quadratic approximation of the problem we get d_k with which another iterate $x_{k+1} = x_k + d_k$ can be computed. Again from the Karush-Kuhn-Tucker theory a solution d_k of this problem and its corresponding Lagrange multiplier λ^k consist a saddle point of the Lagrange function

$$L(d,\lambda) = \frac{1}{2} d^T B_k d + \nabla f(x_k)^T d - \lambda^T h(x_k) - \lambda^T \nabla h(x_k) d . \tag{7}$$

Namely, (d_k, λ^k) is a solution of the following linear algebraic system of equations

$$\begin{bmatrix} B_k & -\nabla h(x_k)^T \\ \nabla h(x_k) & \end{bmatrix} \begin{bmatrix} d \\ \lambda \end{bmatrix} = \begin{bmatrix} -\nabla f(x_k) \\ -h(x_k) \end{bmatrix}. \tag{8}$$

Therefore, we can see that ignoring the right-hand-side term, the linear system (4) obtained from the Lagrange-Newton method is the same as that of the sequential quadratic programming method (8) if the exact Hessian matrix $W(x,\lambda)$ of the Lagrange function $L(x,\lambda) = f(x) - \lambda^T h(x)$ of the original problem, or its quasi-Newton approximation, is the same as the quasi-Newton approximation of the Hessian $\nabla^2 f(x_k)$ given by matrix B_k.

2.3. Augmented Lagrange function

The augmented Lagrange function is

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

where $\lambda \in R^m$ is the Lagrange multiplier and $\sigma > 0$ is the penalty parameter. The stationary condition of augmented Lagrange function is

$$\nabla f(x) + \nabla h(x)^T \lambda + \sigma \nabla h(x)^T h(x) = 0. \tag{10}$$

Now, applying the Newton method in the current point (x_k, λ^k) we get the following linear system

$$\left[\nabla^2 f(x_k) + \sum_{i=1}^m \lambda_i^k \nabla^2 h_i(x_k) + \sigma \sum_{i=1}^m \nabla h_i(x_k) \nabla h_i(x_k)^T + \sum_{i=1}^m h_i(x_k) \nabla^2 h_i(x_k) \right] d =$$
$$-\left[\nabla f(x_k) + \nabla h(x_k)^T \lambda^k + \sigma \nabla h(x_k)^T h(x_k) \right]. \tag{11}$$

Let us define [16]

$$\sigma \nabla h(x_k) d + \sigma h(x_k) = \eta,$$

where $\eta \in R^m$. With this we have

$$\sigma \sum_{i=1}^m \nabla h_i(x_k) \nabla h_i(x_k)^T d = \sigma \nabla h(x_k)^T \nabla h(x_k) d = \nabla h(x_k)^T \left(\eta - \sigma h(x_k) \right).$$

i.e. the Newton system (11) becomes

$$W(x_k, \lambda^k + \sigma h(x_k)) d + \nabla h(x_k)^T \eta = -\left[\nabla f(x_k) + \nabla h(x_k)^T \lambda^k \right].$$

Therefore (11) can be rewritten as

$$\begin{bmatrix} W(x_k, \lambda^k + \sigma h(x_k) & \nabla h(x_k)^T \\ \nabla h(x_k) & -\frac{1}{\sigma} I \end{bmatrix} \begin{bmatrix} d \\ \eta \end{bmatrix} = -\begin{bmatrix} \nabla f(x_k) + \nabla h(x_k)^T \lambda^k \\ h(x_k) \end{bmatrix}. \tag{12}$$

2.4. Inverse barrier function

For inequality constrained problems the inverse barrier function is

$$f(x)+\frac{1}{\sigma}\sum_{i=1}^{m}\frac{1}{h_i(x)}. \tag{13}$$

The necessary condition for minimization of the inverse barrier function is

$$\nabla f(x)-\frac{1}{\sigma}\sum_{i=1}^{m}\frac{1}{h_i^2(x)}\nabla h_i(x)=0. \tag{14}$$

Let us define the following diagonal matrix

$$D(x)=diag(h_1(x),\cdots,h_m(x)).$$

Therefore, (14) can be written as

$$\nabla f(x)-\frac{1}{\sigma}\nabla h(x)^T D(x)^{-3}h(x)=0.$$

In the current point x_k the Newton system for (14) is

$$\left[\nabla^2 f(x_k)-\frac{1}{\sigma}\sum_{i=1}^{m}\frac{1}{h_i^2(x_k)}\nabla^2 h_i(x_k)+\frac{2}{\sigma}\nabla h(x_k)^T D(x_k)^{-3}\nabla h(x_k)\right]d=$$

$$-\nabla f(x_k)+\frac{1}{\sigma}\nabla h(x_k)^T D(x_k)^{-3}h(x_k). \tag{15}$$

As in Yuan [16] let as define

$$\frac{1}{\sigma}D(x_k)^{-3}\left(2\nabla h(x_k)d-h(x_k)\right)=\eta. \tag{16}$$

After some simple algebra, the above system (15) can be written as

$$\begin{bmatrix} W(x_k,-\frac{1}{\sigma}D(x_k)^{-2}) & \nabla h(x_k)^T \\ \nabla h(x_k) & -\frac{\sigma}{2}D(x_k)^3 \end{bmatrix}\begin{bmatrix} d \\ \eta \end{bmatrix}=\begin{bmatrix} -\nabla f(x_k) \\ \frac{1}{2}h(x_k) \end{bmatrix}. \tag{17}$$

2.5. Log-barrier function

The log-barrier function for problem (1) is

$$f(x)-\frac{1}{\sigma}\sum_{i=1}^{m}\log(h_i(x)). \tag{18}$$

Similarly as in the case of the inverse barrier function the necessary condition for minimum is

$$\nabla f(x)-\frac{1}{\sigma}\sum_{i=1}^{m}\frac{1}{h_i(x)}\nabla h_i(x)=0, \tag{19}$$

which can be rewritten as

$$\nabla f(x)-\frac{1}{\sigma}\nabla h(x)^T D(x)^{-2}h(x)=0. \tag{20}$$

The Newton system associated to (19) is

$$\left[\nabla^2 f(x_k)-\frac{1}{\sigma}\sum_{i=1}^{m}\frac{1}{h_i(x_k)}\nabla^2 h_i(x_k)+\frac{1}{\sigma}\nabla h(x_k)^T D(x_k)^{-2}\nabla h(x_k)\right]d=$$

$$-\nabla f(x_k)+\frac{1}{\sigma}\nabla h(x_k)^T D(x_k)^{-2}h(x_k). \tag{21}$$

As in Yuan [16] let us define

$$\frac{1}{\sigma}D(x_k)^{-2}\left[\nabla h(x_k)d-h(x_k)\right]=\eta. \tag{22}$$

Since

$$\frac{1}{\sigma}\nabla h(x_k)^T D(x_k)^{-2}\nabla h(x_k)d = \nabla h(x_k)^T \eta + \frac{1}{\sigma}\nabla h(x_k)^T D(x_k)^{-2} h(x_k),$$

it follows that the above system is reduced to

$$\begin{bmatrix} W(x_k, -\frac{1}{\sigma}D(x_k)) & \nabla h(x_k)^T \\ \nabla h(x_k) & -\sigma D(x_k)^2 \end{bmatrix} \begin{bmatrix} d \\ \eta \end{bmatrix} = \begin{bmatrix} -\nabla f(x_k) \\ h(x_k) \end{bmatrix}. \tag{23}$$

2.6. Log-barrier function for inequality constraints

For the problem with inequality constraints

$$\min f(x) \tag{24a}$$

$$c(x) \geq 0, \tag{24b}$$

where $c: R^n \to R^p$, the best known merit function is the logarithmic barrier function

$$f(x) - \mu\sum_{i=1}^{p}\log(c_i(x)),$$

where μ is a positive scalar barrier parameter. We use the technique as in the case of log-barrier function for equality constraints. The first order optimality conditions are

$$\nabla f(x) - \nabla c(x)^T \lambda = 0,$$
$$C(x)\lambda = 0,$$
$$c(x) \geq 0 \text{ and } \lambda \geq 0,$$

where λ are the Lagrange multipliers and $C(x) = diag(c_1(x), \cdots, c_p(x))$. Now consider the perturbed problem

$$\nabla f(x) - \nabla c(x)^T \lambda = 0,$$
$$C(x)\lambda = \mu e,$$
$$c(x) \geq 0 \text{ and } \lambda \geq 0,$$

where $\mu > 0.$ As we know, the primal-dual path following methods aim to track solutions to the system

$$\nabla f(x) - \nabla c(x)^T \lambda = 0, \tag{25a}$$

$$C(x)\lambda - \mu e = 0, \tag{25b}$$

as $\mu \to 0$ while maintaining the constraints $c(x) > 0$ and $\lambda > 0$. For the nonlinear system (25) the Newton correction (d, η) satisfies

$$\begin{bmatrix} W(x_k, \frac{\mu}{c_i(x_k)}) & -\nabla c(x_k)^T \\ \nabla c(x_k)\Lambda^k & C(x_k) \end{bmatrix} \begin{bmatrix} d \\ \eta \end{bmatrix} = -\begin{bmatrix} \nabla f(x_k) - \nabla c(x_k)^T \lambda^k \\ C(x_k)\lambda^k - \mu e \end{bmatrix},$$

where $\Lambda^k = diag(\lambda_1^k, \ldots, \lambda_p^k)$ and $e = [1, \ldots, 1] \in R^p$.

2.7. Interior point algorithm for inequality constraints

Consider the problem (24) with inequality constraints. After adding the nonnegative slack variables $v = (v_1, \ldots, v_p)$ we obtain the equivalent formulation of the problem (24) as

$$\min f(x) \tag{26a}$$

$$c(x) - v = 0, \quad v \geq 0. \tag{26b}$$

The interior-point method introduces the slacks in a barrier term, thus obtaining the following problem

$$\min\; f(x)-\mu\sum_{i=1}^{p}\log v_i$$
$$c(x)-v=0,$$
where μ is the barrier parameter. The solution of this problem satisfies the following primal-dual system [3]

$$\nabla f(x)-\nabla c(x)^T\lambda=0, \tag{27a}$$
$$-\mu e+V\Lambda e=0, \tag{27b}$$
$$c(x)-v=0, \tag{27c}$$

where $\lambda\in R^p$ is the vector of Lagrange multipliers, Λ and V are diagonal matrices with elements λ_i and v_i respectively and $e=[1,\dots,1]\in R^p$. Appling the Newton method to the system (27) we get the following linear system for the Newton directions

$$\begin{bmatrix} W(x_k,\dfrac{\mu}{c_i(x_k)}) & \nabla c(x_k)^T \\ \nabla c(x_k) & V\Lambda^{-1}\end{bmatrix}\begin{bmatrix} d\\ \eta\end{bmatrix}=\begin{bmatrix}\nabla f(x_k)-\nabla c(x_k)^T\lambda^k \\ v_k-c(x_k)+V\Lambda^{-1}(\mu V^{-1}e-\lambda)\end{bmatrix}. \tag{28}$$

2.8. Path-following method in linear programming

For a couple of linear programming problems

$$\begin{array}{ll} \max c^T x & \min b^T y \\ Ax+w=b, & A^T y-z=c, \\ x,w\ge 0, & y,z\ge 0, \end{array} \tag{29}$$

where $x\in R^n$ and $y\in R^m$, the first order optimality conditions can be expressed as

$$\begin{aligned} &Ax+w=b,\\ &A^T y-z=c,\\ &XZe=\mu e,\\ &YWe=\mu e, \end{aligned} \tag{30}$$

where μ is the barrier parameter. Observe that this is a system with $2n+2m$ equations in $2n+2m$ unknowns. The only nonlinear expressions in these equations are simple multiplications like $x_i z_i$ and $y_i w_i$. The presence of these extremely simple nonlinearities makes the subject of linear programming nontrivial. Now, if a primal feasible set has a nonempty interior and is bounded, then for each $\mu>0$ there exists a unique solution $(x_\mu,w_\mu,y_\mu,z_\mu)$ to (30). The path $\left\{(x_\mu,w_\mu,y_\mu,z_\mu):\mu>0\right\}$ is called *the primal-dual central path.* Our aim is to solve (30), i.e. to find $(\Delta x,\Delta w,\Delta y,\Delta z)$ such that the new point $(x+\Delta x,w+\Delta w,y+\Delta y,z+\Delta z)$ lies on the primal-dual central path at the current point $(x_\mu,w_\mu,y_\mu,z_\mu)$. The Newton method, after dropping the nonlinear terms leads to the following system

$$\begin{bmatrix} -XZ^{-1} & & -I & \\ & & A & I \\ -I & A^T & & \\ & I & & YW^{-1}\end{bmatrix}\begin{bmatrix}\Delta z\\ \Delta y\\ \Delta x\\ \Delta w\end{bmatrix}=\begin{bmatrix}-\mu Z^{-1}e+x\\ \rho\\ \sigma\\ \mu W^{-1}e-y\end{bmatrix}, \tag{31}$$

where $\rho=b-Ax-w$ and $\sigma=c-A^T y+z$ are the primal infeasibility and the dual infeasibility , respectively. This is known as the *Karush-Kuhn-Tucker system* (KKT system). Solving this system subject to Δz and Δw as

$$\Delta z = X^{-1}(\mu e - XZe - Z\Delta x)$$
$$\Delta w = Y^{-1}(\mu e - YWe - W\Delta y),$$

then (31) reduces to

$$\begin{bmatrix} -Y^{-1}W & A \\ A^T & X^{-1}Z \end{bmatrix}\begin{bmatrix} \Delta y \\ \Delta x \end{bmatrix} = \begin{bmatrix} b - Ax - \mu Y^{-1}e \\ c - A^T y + \mu X^{-1}e \end{bmatrix} \tag{32}$$

which is a symmetric system known as the *reduced KKT system* [14]

2.9. Affine scaling interior point method for linear inequality constraints

For the problem

$$\min f(x), \qquad Ax \geq b, \tag{33}$$

where $x \in R^n$ and $b \in R^m$ Coleman and Li [5] presented a trust region and affine scaling interior point method where ignoring the primal and the dual feasibility constraints as usual in interior point methods the first order necessary optimality conditions can be expressed as

$$\nabla f(x) - A^T\lambda = 0, \tag{34a}$$
$$diag(Ax - b)\lambda = 0. \tag{34b}$$

Denoting $D(x) = diag(Ax - b)$ and $D_k = D(x_k)$, then the Newton step (d, η) for (34) satisfies

$$\begin{bmatrix} \nabla^2 f(x_k) & -A^T \\ diag(\lambda_k)A & D_k \end{bmatrix}\begin{bmatrix} d \\ \eta \end{bmatrix} = -\begin{bmatrix} \nabla f(x_k) - A^T\lambda_k \\ D_k\lambda_k \end{bmatrix}. \tag{35}$$

Far away from the solution the Newton step d may not be a descent direction for the function $f(x)$. Therefore, a globalization consists of replacing $diag(\lambda_k)$ by $\Lambda_k = diag(|\lambda_k|)$. With this the modified Newton step satisfy

$$\begin{bmatrix} \nabla^2 f(x_k) & -A^T \\ \Lambda_k A & D_k \end{bmatrix}\begin{bmatrix} d \\ \eta \end{bmatrix} = -\begin{bmatrix} \nabla f(x_k) - A^T\lambda_k \\ D_k\lambda_k \end{bmatrix}. \tag{36}$$

But from (36) $\eta = -\lambda_k - D_k^{-1}\Lambda_k Ad$. Therefore, the modified Newton step d is a minimizer of the augmented quadratic

$$\frac{1}{2}d^T(\nabla^2 f(x_k) + A^T D_k^{-1}\Lambda_k A)d + \nabla f(x_k)^T d, \tag{37}$$

which can be considered as a quadratic convex regularization of (33). Thus a trust region subproblem can be defined as

$$\min_{d\in R^n} \frac{1}{2}d^T(\nabla^2 f(x_k) + A^T D_k^{-1}\Lambda_k A)d + \nabla f(x_k)^T d$$
$$\left\| d; D_k^{-1/2}Ad \right\| \leq \Delta_k.$$

Considering $\tau \geq 0$ the Lagrange multiplier of the above subproblem we get the following necessary optimality conditions

$$\nabla f(x_k) + (\nabla^2 f(x_k) + A^T D_k^{-1}\Lambda_k A)d + \tau(I + A^T D_k^{-1}A)d = 0, \tag{38a}$$
$$\tau(\Delta_k - \left\| d; D_k^{1/2}Ad \right\|) = 0. \tag{38b}$$

Now, defining

$$D_k^{-1}(\Lambda_k Ad + \tau Ad) + \lambda_k = -\eta \tag{39}$$

(38a) can be rewritten as

$$\begin{bmatrix} \nabla^2 f(x_k) + \tau I & -A^T \\ \Lambda_k A + \tau A & D_k \end{bmatrix}\begin{bmatrix} d \\ \eta \end{bmatrix} = -\begin{bmatrix} \nabla f(x_k) - A^T\lambda_k \\ D_k\lambda_k \end{bmatrix}.$$

The extension to nonlinear inequality constrains was given by Yuan [16]

2.10. Simple bounded optimization

For the simple bounded problem

$$\min\left\{ f(x) : l \le x \le u \right\} \tag{40}$$

where $l, u \in R^n$, the interior point method introduces the slack variables $v, w \in R^n$, $v \ge 0,\ w \ge 0,$ in a barrier term, thus obtaining the following problem

$$\min\ f(x) - \mu \sum_{i=1}^{n} \log w_i - \mu \sum_{i=1}^{n} \log v_i \tag{41a}$$

$$x - w = l, \tag{41b}$$

$$x + v = u, \tag{41c}$$

where μ is the barrier parameter. Using the Lagrangean

$$L(x, w, v, p, q) = f(x) - \mu \sum_{i=1}^{n} \log w_i - \mu \sum_{i=1}^{n} \log v_i - p^T(x - w - l) + q^T(x + v - u)$$

where $p, q \in R^n$ are the Lagrange multipliers, then the first order optimality conditions are

$$\nabla f(x) - p + q = 0, \tag{42a}$$

$$(X - L)Pe - \mu e = 0, \tag{42b}$$

$$(U - X)Qe - \mu e = 0. \tag{42c}$$

After some algebra, the Newton method leads to the following linear symmetric system

$$\begin{bmatrix} \nabla^2 f(x_k) & -I & I \\ -I & -P_k^{-1}(X_k - L) & \\ I & & -Q_k^{-1}(U - X_k) \end{bmatrix} \begin{bmatrix} d \\ \eta_1 \\ \eta_2 \end{bmatrix} = -\begin{bmatrix} \nabla f(x_k) - p_k - q_k \\ -(X_k - L)e + \mu P_k^{-1} e \\ -(U - X_k)e + \mu Q_k^{-1} e \end{bmatrix}. \tag{43}$$

2.11. Celis-Dennis-Tapia (CDT) method

For equality constrained optimization the subproblem corresponding to the CDT method is

$$\min\ g_k^T d + \frac{1}{2} d^T B_k d$$

$$\text{subject to} \tag{44}$$

$$\left\| h(x_k) + \nabla h(x_k)^T d \right\|^2 \le \xi^2,$$

$$\|d\| \le \Delta_k,$$

where ξ^2 is a parameter between $\min_{\|d\| \le \Delta_k} \left\| h(x_k) + \nabla h(x_k)^T d \right\|$ and $\|h(x_k)\|$, (see [4], [12]). The solution d_k of the CDT subproblem is obtained through the optimality conditions for (44) as:

$$B_k d_k + g_k + \tau d_k + \mu(\nabla h(x_k)^T d_k + h(x_k)) = 0, \tag{45a}$$

$$\tau(\Delta_k - \|d_k\|) = 0, \tag{45b}$$

$$\mu\left(\xi^2 - \left\| h(x_k) + \nabla h(x_k)^T d \right\|^2 \right) = 0, \tag{45c}$$

where $\tau \ge 0$ and $\mu \ge 0$ are the Lagrange multipliers. Considering

$$\eta = -\mu(\nabla h(x_k)^T d + h(x_k)),$$

the following system is obtained

$$\begin{bmatrix} B_k + \tau I & -\nabla h(x_k) \\ -\mu \nabla h(x_k)^T & -I \end{bmatrix} \begin{bmatrix} d \\ \eta \end{bmatrix} = \begin{bmatrix} -g_k \\ \mu h(x_k) \end{bmatrix}. \tag{46}$$

If $\mu > 0,$ than (46) can be rewritten as

$$\begin{bmatrix} B_k + \tau I & -\nabla h(x_k) \\ \nabla h(x_k)^T & \dfrac{1}{\mu} I \end{bmatrix} \begin{bmatrix} d \\ \eta \end{bmatrix} = -\begin{bmatrix} g_k \\ h(x_k) \end{bmatrix}. \tag{47}$$

3. Quadratic Internal Model Principle

Therefore, 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:

$$\begin{bmatrix} W(x_k, \lambda^k) + T_k & -\nabla h(x_k)^T \\ \nabla h(x_k) & S_k \end{bmatrix} \begin{bmatrix} d \\ \eta \end{bmatrix} = -\begin{bmatrix} \nabla f(x_k) + \varepsilon_1 \\ h(x_k) + \varepsilon_2 \end{bmatrix}, \tag{48}$$

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 (48) 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 (48) we get

$$\begin{bmatrix} W(x_k, \lambda^k) + T_k & -\nabla h(x_k)^T \\ \nabla h(x_k) & 0 \end{bmatrix} \begin{bmatrix} d \\ \eta \end{bmatrix} = -\begin{bmatrix} \nabla f(x_k) + \varepsilon_1 \\ h(x_k) + \varepsilon_2 \end{bmatrix}, \tag{49}$$

It is easy to see that the system (49) corresponds to the Newton method for optimization with equality nonlinear constraints or to the sequential quadratic programming method. The augmented system (49) can be considered as the necessary condition for d to be a solution of the following quadratic programming problem

$$\begin{aligned} &\min \frac{1}{2} d^T (W(x_k, \lambda^k) + T_k) d - (\nabla f(x_k) + \varepsilon_1)^T d \\ &\text{subject to} \\ &\nabla h(x_k) d + (h(x_k) + \varepsilon_2) = 0. \end{aligned} \tag{50}$$

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 (50) has a unique global solution d. This solution can be obtained by solving the augmented system (49), where d is the solution of the problem and η is the Lagrange multiplier associated to the equality constraint. The problem (50) is *the quadratic internal model of problem (1) associated to the methods involving the linear system (49) (with $S_k = 0$).*

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

$$\begin{aligned} &\min \frac{1}{2} d^T W(x_k, \lambda^k) d - (\nabla f(x_k) + \nabla h(x_k) \lambda^k)^T d \\ &\text{subject to} \\ &\nabla h(x_k) d + h(x_k) = 0. \end{aligned} \tag{51}$$

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 (48) 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 (48) we get

$$\eta = -S_k^{-1}\nabla h(x_k)d - S_k^{-1}h(x_k) - S_k^{-1}\varepsilon_2. \tag{52}$$

Therefore, using (52) in (48) it follows that

$$\left[W(x_k,\lambda^k) + T_k + \nabla h(x_k)^T S_k^{-1}\nabla h(x_k)\right]d$$
$$= -\nabla h(x_k)^T S_k^{-1}h(x_k) - \nabla h(x_k)^T S_k^{-1}\varepsilon_2 - \nabla f(x_k) - \varepsilon_1. \tag{53}$$

But, (53) is equivalent with the following quadratic problem

$$\min \frac{1}{2} d^T \left[W(x_k,\lambda^k) + T_k + \nabla h(x_k)^T S_k^{-1}\nabla h(x_k)\right]d$$
$$-\left[-\nabla h(x_k)^T S_k^{-1}h(x_k) - \nabla h(x_k)^T S_k^{-1}\varepsilon_2 - \nabla f(x_k) - \varepsilon_1\right]^T d \tag{54}$$

which is called *the quadratic internal model of problem (1) associated to the methods involving the linear system (48) (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 (54), which represents the essence of the problem from the view point of the algorithm involving (48).

Observe that the quadratic internal model of (1), as expressed by (54), 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

$$L(x,\lambda,\sigma) = f(x) + \lambda^T h(x) + \frac{1}{2}\sigma \left\|h(x)\right\|_2^2, \tag{55}$$

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

$$\min \frac{1}{2} d^T (W(x_k, \lambda^k h(x_k)) + \sigma \nabla h(x_k)^T \nabla h(x_k))d -$$
$$(-\nabla f(x_k) + \nabla h(x_k)^T (\lambda^k - \sigma h(x_k)))d. \tag{56}$$

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

$$\min \frac{1}{2} d^T \nabla^2 f(x_k)d + \nabla f(x_k)^T d \tag{57}$$

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 (57). Similarly, we can say that for the unconstrained problem the quasi-Newton step can be obtained by solving the following quadratic problem

$$\min \frac{1}{2} d^T B_k d + \nabla f(x_k)^T d \tag{58}$$

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 $B_k(x_{k+1} - x_k) = \nabla f(x_{k+1}) - \nabla f(x_k)$.

4. 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

[1] Andrei, N., Quadratic Internal Model Principle in Mathematical Programming, Newsletter 16 of EUROPT – The Continuous Optimization Working Group of EURO, September 2009.

[2] Beale, E.M.L., *Numerical Methods,* in Nonlinear Programming, edited by J. Abadie, North-Holland, Amsterdam, 1967.

[3] Benson, H.Y., Shanno, D.F., Vanderbei, R.J., *Interior point methods for nonconvex nonlinear programming: jamming and comparative numerical testing.* Mathematical programming, 99 (2004), pp.35-48.

[4] Celis, M.R., Dennis, Jr., Tapia, R.A., *A trust region algorithm for nonlinear equality constrained optimization*, in P.T. Boogs, R.H. Byrd and R.B. Schnabel (Eds.) *Numerical Optimization*, SIAM, Philadelphia, 1985, pp.71-82.

[5] Coleman, T.F., Li, Y., *A trust region and affine scaling interior point method for nonconvex minimization with linear inequality constraints.* Mathematical Programming, 88 (2000), pp.1-31.

[6] Fourier, J.B.J., *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.

[7] Han, S.P., *A globally convergent method for nonlinear programming.* Journal for Optimization Theory and Applications, 22 (1977), pp.297-309.

[8] Lagrange, J,-L., *Théorie des functions analytiques*. Imp. De la République, Paris, 1797.

[9] Noether, E., *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.

[10] Palomares, G., Mangasarian, O.L., *Superlinearly convertgent quasi-Newton algorithms for nonlinearly constrained optimization problems.* Mathematical Programming, 11 (1976), pp.1-13.

[11] Powell, M.J.D., *The convergence of variable metric methods for nonlinearly constrained optimization calculations,* in Nonlinear Programming 3, edited by O.L. Mangasarian, R.R. Meyer and S.M. Robinson, Academic Press, London, 1978.

[12] Powell, M.J.D., Yuan, Y., *A trust region algorithm for equality constrained optimization.* Math. Prog. 49 (1991) 189-211.

[13] Scharlau, W., *On the history of the algebraic theory of quadratic forms.* In Quadratic Forms and their Applications, Proceedings of the Conference on Quadratic Forms and their Applications, July 509, 1999, University College Dublin, E. Bayer-Fluckiger, D. Lewis and A. Ranicki (Eds.). Published as Contemporary Mathematics 272, A.M.S., 2000, pp. 229-259.

[14] Vanderbei, R.J., *Linear Programming. Foundations and extensions.* Kluwer Academic Publishers, Boston, 1996.

[15] Wilson, R.B., *A simplicial algorithm for concave programming.* Ph.D. Thesis, Harvard University, Graduate School of Business Administration, 1963.

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

May 15, 2009