Geometric Notes on Optimization with Equality Constraints

W. C. RHEINBOLDT

Department of Mathematics and Statistics
University of Pittsburgh, Pittsburgh, PA 15260, U.S.A.

Abstract—For minimization problems with nonlinear equality constraints, various numerical tools are shown to become available when the constraint set has a manifold structure. In appropriate local coordinate systems these tools permit the computation, e.g., of the gradient and Hessian of the transformed (unconstrained) objective function. This opens up a new view on the computational solution of the minimization problem which—while leading to algorithms similar in concept and performance to the well-known "reduced Hessian" methods—provides a different theoretical basis for these methods.

Keywords—Optimization, Equality constraints, Manifolds, Local coordinates, Reduced Hessian methods.

1. INTRODUCTION

Consider the constrained optimization problem

$$\min\{g(x) : x \in \mathbb{R}^n, \ F(x) = 0\}, \tag{1.1}$$

where the mappings $g : \mathbb{R}^n \mapsto \mathbb{R}$ and $F : \mathbb{R}^n \mapsto \mathbb{R}^m$, with $n = m + d$, $d > 0$, are of class C^r for $r \geq 2$ on some open subset E of $\mathbb{R}^n$. We restrict ourselves to a set E where

$$\operatorname{rank} DF(x) = m, \qquad \forall x \in E. \tag{1.2}$$

Here $DF(x)$ is the Jacobian of F at x, and we will denote the gradient of g by $\nabla g(x) = Dg(x)^T$ and its derivative—the Hessian matrix—by $\nabla^2 g(x)$.

A well-known necessary condition for $x^* \in E$ to be a local minimizer of (1.1) requires the existence of a Lagrange multiplier $\lambda^* \in \mathbb{R}^m$ such that (x^*, λ^*) is a critical point of the functional

$$h : \mathbb{R}^n \times \mathbb{R}^m \mapsto \mathbb{R}, \qquad h(x, \lambda) = g(x) + \lambda^T F(x). \tag{1.3}$$

A sufficient condition for $x^* \in E$ to be a minimizer is the positive definiteness of the "reduced" Hessian matrix $U^T \nabla^2 h(x^*, \lambda^*) U$ for some, and hence any, $n \times m$ matrix U whose columns span $\ker DF(x^*)$.

These facts have been used in various ways in the development of numerical methods for computing local minimizers of (1.1). One class of these methods works, in essence, with the augmented functional (1.3) and hence with its gradient and Hessian (or some approximation);

Typeset by $\mathcal{A}\mathcal{M}\mathcal{S}$-TEX

see, e.g., the surveys [1,2]. Another class of methods uses projections of $\nabla h(x, \lambda)$ onto the nullspace of $DF(x)$ and involves the mentioned reduced Hessian matrix or some approximation of it; see, e.g., [3], and the survey [2].

The submersion assumption (1.2) guarantees that

$$M = \{x \in E : F(x) = 0\} \tag{1.4}$$

is a d-dimensional C^r-submanifold of $\mathbb{R}^n$. In connection with (1.1) this fact appears to have been used first by Tanabe [4]. Yet it has found little, if any, application in the development of the above indicated classes of methods. The aim of this note is to show that this manifold structure provides the basis for some geometrical observations which, in turn, suggest new tools for the computational solution of (1.1). While the resulting approaches are similar in concept and performance to the "reduced Hessian" methods (e.g., that in [3]), the geometric insight provides a different theoretical basis for these methods.

2. LOCAL COORDINATE SYSTEMS

Central to any analysis on a differentiable manifold is the concept of a chart or its inverse, a local coordinate system. Specifically, for submanifolds of $\mathbb{R}^n$, such as (1.4), a pair (Ω, Φ) is a local coordinate system on M if

(i) Ω is an open neighborhood of $0 \in \mathbb{R}^n$,
(ii) the C^r-mapping $\Phi : \Omega \mapsto \mathbb{R}^n$ is a homeomorphism of Ω onto the (relatively) open subset $\Phi(\Omega)$ of M, and
(iii) Φ is an immersion on Ω.

We call (Ω, Φ) a local coordinate system of any point $x \in M$ such that $x \in \Phi(\Omega)$.

In any local coordinate system (Ω, Φ) on M, the functional g restricted to $\Phi(\Omega) \subset M$ is transformed into the functional

$$\gamma : \Omega \mapsto \mathbb{R}, \qquad \gamma(y) = g(\Phi(y)), \quad \forall y \in \Omega, \tag{2.1}$$

which is of the class C^r on the open set $\Omega \subset \mathbb{R}^d$. Thus, in local coordinates the constraints are always automatically satisfied.

We review briefly some earlier results about certain local coordinate systems for which effective numerical methods have been developed (see, e.g., [5-7]). For this we assume that $\mathbb{R}^n$ is equipped with its canonical inner product which then induces a Riemannian metric on M. For any $x \in M$ the tangent and normal space are denoted by $T_x M$ and $N_x M$, respectively.

Let $x^c \in M$ be a point on M where a local coordinate system is to be constructed. For this we use the splitting $\mathbb{R}^n = T_{x^c} M \oplus N_{x^c} M$ and apply the implicit function theorem to $F(x^c + u + w) = 0$, $u \in T_{x^c} M$, $w \in N_{x^c} M$. This ensures the existence of open neighborhoods $V \subset T_{x^c} M$ of $0 \in T_{x^c} M$ and $W \subset \mathbb{R}^n$ of x^c, and of a C^r mapping $\psi : V \mapsto N_x M$ such that $\psi(0) = 0$ and $M \cap W = \Phi(V)$ where

$$\Phi : V \mapsto \mathbb{R}^n, \qquad \Phi(u) = x^c + u + \psi(u) \in \mathbb{R}^n, \quad u \in V. \tag{2.2}$$

Any choice of a basis of $T_{x^c} M$ provides us with an isomorphism from $T_{x^c} M$ onto $\mathbb{R}^d$ under which the image of V is some open neighborhood Ω of $0 \in \mathbb{R}^d$. Then it follows easily that (Ω, Φ) is a local coordinate system of x^c on M to be called here a tangent coordinate system.

For the mapping ψ in (2.2) we note that $D\Phi(0)u \in T_{x^c} M$ for $u \in T_{x^c} M$ implies that $D\psi(0) = 0$. Moreover, in [8] it was shown that

$$D^2\psi(0)\left(u^1, u^2\right) = V_{x^c}\left(u^1, u^2\right), \qquad \forall u^1, u^2 \in T_{x^c} M \tag{2.3}$$

where $V_x : T_x M \times T_x M \mapsto N_x M$ denotes the second fundamental tensor of M at x.

Several methods exist for the computation of an $n \times d$ matrix $U(x)$ whose columns form a basis of the tangent space $T_x M = \ker DF(x)$ of a given $x \in M$. For instance, a QR-factorization of $DF(x)^T$ obviously provides such matrix with orthonormal columns. It is readily seen (see, e.g., [9]) that there exists a $\delta > 0$ such that for any $y \in \mathbb{R}^d$, $\|y\| \le \delta$ the nonlinear system

$$\begin{pmatrix} F(x) \\ U(x^c)^T(x - x^c) \end{pmatrix} = \begin{pmatrix} 0 \\ y \end{pmatrix} \tag{2.4}$$

has the unique solution $x = \Phi(y)$ which can be easily computed, say, by means of a chord Newton method. If a QR-factorization was used to obtain $U(x^c)$ this can be simplified considerably so that each step involves only one back-solve in $\mathbb{R}^d$.

From (2.4) it follows (see, e.g., [7]) that the derivative $D\Phi(y)$ of Φ at y satisfies

$$\begin{pmatrix} DF(\Phi(y)) \\ U(x^c)^T \end{pmatrix} D\Phi(y) = \begin{pmatrix} 0 \\ I_d \end{pmatrix}. \tag{2.5}$$

For the case when $U(x)$ has orthonormal columns, the explicit solution

$$D\Phi(y) = U(\Phi(y))[U(x^c)^T U(\Phi(y))]^{-1} \tag{2.6}$$

was given in [10] and it was also shown that (2.6) does not depend on the particular choice of $U(\Phi(y))$. This offers another approach for computing $D\Phi$.

Finally, in [11] it was proved that, when Q denotes the orthogonal projection onto $N_x M$ for given $x \in M$, then for any $u^1, u^2 = \ker DF(x)$ we have $V_x(u^1, u^2) = -Qz$ with any solution z of $DF(x)z = D^2 F(x)(u^1, u^2)$. Thus $V_x(u^1, u^2) \in N_x M$ is the unique solution of the linear system

$$\begin{pmatrix} DF(x) \\ U(x)^T \end{pmatrix} V_x(u^1, u^2) = \begin{pmatrix} -D^2 F(x)(u^1, u^2) \\ 0 \end{pmatrix}. \tag{2.7}$$

As before, the computation is considerably simplified when QR-factorizations are used.

3. APPLICATION TO PROBLEM (1.1)

Once a tangent coordinate system (Ω, Φ) has been chosen, the methods sketched in the previous section allow us to compute for any $y \in \mathbb{R}^d$, $\|y\| \le \delta$ the point $x = \Phi(y) \in M$ and with it the value $\gamma(y) = g(\Phi(x))$ of the transformed objective function (2.2), the corresponding gradient

$$\nabla \gamma(y) = D\Phi(y)^T \nabla g(\Phi(y)), \tag{3.1}$$

and, for each $i, j = 1, \ldots, d$, the component of the Hessian

$$e^{i^T} \nabla^2 \gamma(y) e^j = (u^i)^T \nabla^2 g(x) u^j + V_x(u^i, u^j)^T \nabla g(x), \qquad u^i = U(x)e^i, \quad u^j = U(x)e^j. \tag{3.2}$$

The equation (3.2) is of interest in itself. For strictly convex g, $\nabla^2 g(x)$ is positive definite, but the second term of (3.2) may well be negative. This term reflects curvature properties of M at x. Note that, by definition, we have $V_x(u^i, u^j) \in N_x M$, and if x is a local minimizer of g, then the necessary condition involving the functional (1.3) requires that $\nabla g(x) \in \text{rge } DF(x)^T = N_x M$. Thus, the second term in (3.2) certainly need not be small near a local minimizer unless the curvature of M at that point is zero, or the point constitutes a local minimizer of g in $\mathbb{R}^n$.

The computational tools discussed so far can now be applied to the numerical solution of (1.1). Let $x^k \in M$ be a known iterate where a tangent coordinate system (Ω, Φ) has been constructed. Then, a step of our procedure will consist in the computation of a point $x^{k+1} = \Phi(y^{k+1})$ such that $y^{k+1} \in \Omega$ and that some sufficient-decrease condition in g holds. In other words, the corresponding step $g(x^k) - g(x^{k+1}) > 0$ in the g-values should be sufficiently far away from zero.

For the computation we introduce a suitable radius $\delta_k > 0$ such that the closed ball $\mathcal{B}_k = \{y \in \mathbb{R}^d : \|y\| \leq \delta_k\}$ is contained in Ω. Estimates of such a $\delta > 0$ can be obtained easily from the iterative procedure used in the computation of $x = \Phi(y)$ for given y. Such techniques have been indicated in [7,12], and are similar to those discussed in [13]. Of course, if the exact Hessians (3.2) are used then information about the local curvature behavior of M can be derived directly from the second fundamental tensor.

With δ_k the computation of the next point x^{k+1} now requires minimizing γ on $\mathcal{B}_k$. This suggests the application of a trust region method to some quadratic approximation of γ. For example, with the Taylor approximation

$$\hat{\gamma}(y) = \gamma(0) + \nabla\gamma(0)y + \frac{1}{2}y^T\nabla^2\gamma(0)y, \qquad y \in \mathbb{R}^d, \tag{3.3}$$

we wish to solve the problem

$$\min\left\{\hat{\gamma}(y); \; y \in \mathbb{R}^d, \; \|y\| \leq \delta.\right\}. \tag{3.4}$$

Trust region methods have been investigated by several authors (see, e.g., [14,15]). Any such method is designed to produce a point $y^{k+1} \in \mathcal{B}_k$ which satisfies a sufficient decrease condition for $\hat{\gamma}$. From y^{k+1} the next iterate x^{k+1} is obtained by an application of Φ. Of course, in practice, it is desirable to avoid the computation of the exact Hessians in (3.3) by replacing them with certain matrices A_k. Here, a widely used approach is to update these A_k from step to step by means of variable metric methods (see, e.g., [1,3]).

Suppose that for the tangent coordinate system at x^k the set $\Phi(\Omega)$ does not contain a local minimizer of g. Then, we expect $x^{k+1} = \Phi(y^{k+1})$ to be a point for which y^{k+1} is on the boundary of $\mathcal{B}_k$, that is, where $\|y^{k+1}\|/\delta = 1$. Hence, for any reasonable choice of δ_k, we expect x^{k+1} to be almost out of the domain of validity of the coordinate system. Accordingly we evaluate a new tangent coordinate system at x^{k+1} and proceed as before.

On the other hand, suppose that $x^k \in M$ is a point where the set $\Phi(\Omega) \subset M$ contains a local minimizer $x^* = \Phi(y^*)$ of g. Then, we expect that γ has compact level sets in Ω containing y^*, as is the case, for example, when $\nabla^2\gamma(y^*)$ is positive definite. Hence, now the quotient $\|y^{k+1}\|/\delta$ should be below one. If this quotient is less than some tolerance $1 - \eta > 0$ then we retain the old coordinate system constructed at x^k. In other words, our process now becomes a standard unconstrained descent method for γ in Ω which continues in the same tangent coordinate system until either there is convergence to a local minimizer of γ in Ω or the computed points are detected to leave that set.

An essential consequence of the present setting is that any chosen minimization procedure for γ that converges in Ω to y^* will retain exactly the asymptotic convergence properties known for it in the general unconstrained minimization setting. No further proofs are required here. In particular, the variable metric methods involving, e.g., the DFP or BFGS updates will have the superlinear asymptotic convergence behavior proved for such methods in [16] (see also [17]).

The overall procedure is similar in concept to the mentioned reduced Hessian methods, as, e.g., the method in [3]. But our approach provides a strong geometric basis for the algorithm. This shows itself not only in the simple local convergence results but also, for instance, in our "updating algorithm" for the local coordinate system—which in [3] is reflected by the recomputation of the basis of the nullspace of DF. As expected, numerical experiments with the new approach correspond largely to those reported for the known methods. However, we should mention that when the Taylor approximation (3.3) is used, then the new method does not exhibit the well known slow-down in solving the Maratos-problem

$$\min\left\{-x_1 + 10 * \left(x_1^2 + x_2^2 - 1\right) : x \in \mathbb{R}^2, \; x_1^2 + x_2^2 - 1 = 0\right\},$$

as can be seen in Table 1.

Table 1.

Step	x_1	x_2	$\|D\gamma(y)\|$
1	0.8	0.6	0.6
2	0.94735	0.32030	0.33705
3	0.99996	0.83802(−2)	0.88212(−2)
4	1.00000	−0.78728(−5)	0.78731(−5)
5	1.00000	0.16391(−7)	0.16391(−7)

REFERENCES

1. M.J.D. Powell, Variable metric methods for constrained optimization, In *Mathematical Programming: The State of the Art*, (Edited by A. Bachem, M. Groetschel and B. Korte), pp. 288–311, Springer-Verlag, Berlin, (1983).
2. P.E. Gill, W. Murray and M.H. Wright, *Practical Optimization*, Academic Press, New York, (1981).
3. T.F. Coleman and A.R. Conn, On the local convergence of a quasi-Newton method for the nonlinear programming problem, *SIAM J. Num. Anal.* **21**, 755–769 (1984).
4. K. Tanabe, A geometric method in nonlinear programming, *J. Optim. Theory Appl.* **30**, 181–210 (1980).
5. W.C. Rheinboldt, *Numerical Analysis of Parametrized Nonlinear Equations*, J. Wiley and Sons, New York, (1986).
6. W.C. Rheinboldt, On the computation of multi-dimensional solution manifolds of parametrized equations, *Numer. Math.* **53**, 165–181 (1988).
7. F.A. Potra and W.C. Rheinboldt, Differential-geometric techniques for solving differential algebraic equations, In *Real-Time Integration Methods for Mechanical Systems Simulation*, (Edited by R. Deyo and E. Haug), pp. 155–192, Springer-Verlag, New York, (1991).
8. P.J. Rabier and W.C. Rheinboldt, On a computational method for the second fundamental tensor and its application to bifurcation problems, *Num. Math.* **57**, 681–694 (1990).
9. J.M. Ortega and W.C. Rheinboldt, *Iterative Solutions of Nonlinear Equations in Several Variables*, Academic Press, New York, (1970).
10. P.J. Rabier and W.C. Rheinboldt, On the computation of impasse points of quasilinear differential algebraic equations, *Math. Comp.* **62**, 133–154 (1994).
11. P.J. Rabier and W.C. Rheinboldt, On the numerical solution of the Euler Lagrange equations, *SIAM J. Numer. Anal.* **32**, 318–329 (1995).
12. F.A. Potra and J. Yen, Implicit numerical integration for Euler Lagrange equations via tangent space parametrization, *J. Mech. Struct. Mach.* **19**, 76–98 (1991).
13. C. den Heijer and W.C. Rheinboldt, On steplength algorithms for a class of continuation methods, *SIAM J. Num. Anal.* **18**, 925–948 (1981).
14. D.M. Gay, Computing optimal locally constrained steps, *SIAM J. Sci. Stat. Comp.* **2**, 186–197 (1981).
15. J.J. Moré and D.C. Sorensen, Computing a trust region step, *SIAM J. Sci. Stat. Comp.* **4**, 553–572 (1983).
16. J.E. Dennis and J.J. Moré, A characterization of superlinear convergence and its application to quasi-Newton methods, *Math. Comp.* **28**, 549–560 (1974).
17. J.E. Dennis and R.B. Schnabel, *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall, Englewood Cliffs, NJ, (1983).