Riemannian Convexity in Programming (II)

Constantin Udrişte

Abstract

The book [7] emphasizes three relevant aspects: first, the fact that the notion of convexity is strongly metric-dependent either through geodesics or through the Riemannian connection; second that Riemannian convexity of functions is a coordinate-free concept, and consequently it can be easily connected with symbolic computation; third, that the Riemannian structure is involved essentially in formulating and solving programs by means of induced distance, geodesics, Riemannian connection, sectional curvature, etc.

The preceding arguments justify the effort to generalize the optimization theory on Euclidean spaces to the Riemannian manifolds. The generalization is obtained by selecting a suitable Riemannian metric, by passing from vector addition to the exponential map, by changing the search along straight lines with a search along geodesics, and by using covariant differentiation instead of partial differentiation.

§1 shows that some difficulties appearing in the free-minimization problems belong to a wrong understanding of the suitable Riemannian structure of the space. §2 deals with Newton algorithm on Riemannian manifolds for finding zeros of a C^∞ vector field or generally for a C^∞ tensor field. §3 describes the path of centers attached to a convex program on a Riemannian manifold and analysis the monotonicity of the objective function along this curve. §4 gives theorems regarding the Newton method near the path of centers of a convex program (one unit Newton step stays inside the feasible set, quadratic convergence results, upper bound for the difference of two Huard distance function values, etc), using simultaneously the original Riemannian metric and a Hessian Riemannian metric. §5 gives upper bounds for the total number of outer iterations and inner iterations needed by the center algorithm on a Riemannian manifold.

The theorems in §3-5 have their origin in the Euclidean variants exposed in [1], [2] and in the Riemannian point of view about convex programming developed in [7].

1991 Mathematics Subject Classification: 49M45, 49M15
Key words: Objective functions, Newton method, Riemannian convexity, path of centers, complexity analysis

Balkan Journal of Geometry and Its Applications, Vol.1, No.1, 1996, pp. 99-109

1 Commentary on test objective functions

The test functions used in the numerical free minimization have been constructed in such a way as to present various computational difficulties when we look for the finding of a critical point. A list of some of more frequently used test functions on (R^n, δ_{ij}), together with an initial estimate x_1 of the minimizer (critical point) x_*, and a brief description of the computational difficulties presented by the optimization problem is given for example in [11].

For some of such functions we shall point out that some difficulties belong to a wrong understanding of the suitable Riemannian structure of the space, which can create or destroy the convexity. Of course in the minima problems we are interested to create the convexity of the objective function because this assures the convergence of the numerical methods of optimization towards a minimum point.

Let (M, g) be a complete n-dimensional Riemannian manifold, let ∇ be the Riemannian connection and $\gamma(t) = \exp_x(tX), t \in [0, 1]$ be the geodesic determined by the initial conditions $\gamma(0) = x \in M$, $\dot{\gamma}(0) = X \in T_xM$.

If $f : M \to R$ is of class C^∞, and x_* is a minimum point (critical point) of f, then the *general descent algorithm* for finding x_* is:

Let x_1 be an estimate of x_*.

1) Set i = 1.
2) Compute a vector X_i such that $df(X_i) < 0$.
3) Compute a number t_i such that $f\big(\exp_{x_i}(t_iX_i)\big) < f(x_i)$.
4) Compute x_{i+1} from $x_{i+1} = \exp_{x_i}(t_iX_i)$.
5) If x_{i+1} satisfies the given convergence criteria, then stop.
6) Set $i = i + 1$ and go to 2).

The *Newton algorithm* is the variant of the general descent algorithm with $X = -H^{-1}df$, where the Hessian $H = \nabla(\nabla f)$ is supposed to be nondegenerate.

For details, see [7] and the explanations in §2.

1. Rosenbrock banana function (1960)

$$f : R^2 \to R, \quad f(x) = 100(x_2 - x_1^2)^2 + (1 - x_1)^2, \quad x = (x_1, x_2)$$

$$e = (-1 \cdot 2; 1 \cdot 0), \quad x_* = (1, 1), \quad f(x_*) = 0.$$

Difficulties. The graph of this function looks like a steep-sided sharply curving valley, the bottom of which follows the parabola $x_2 = x_1^2$. An Euclidean descent method must in general take a very short steps in order to negociate the sharp bend near (0,0) and must therefore provide new search directions very frecvently.

Let us show that there exists a Riemannian metric on R^2 which produces the convexity of Rosenbrock banana function [7].

Denote by

$$\left(R^2, g_0 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$

the Euclidean plane. The function

$$F : R^2 \to R, F(y) = 100y_2^2 + y_1^2, \quad y = (y_1, y_2)$$

is convex with respect to g_0.

Now we consider the nonlinear coordinate transformation

$$y_1 = 1 - x_1, y_2 = x_2 - x_1^2.$$

The Riemannian manifold (R^2, g_0) is changed into

$$\left(R^2, g(x) = \begin{pmatrix} 4x_1^2+1 & -2x_1 \\ -2x_1 & 1 \end{pmatrix}\right), x = (x_1, x_2)$$

and F is changed into Rosenbrock banana function f which is convex with respect to g.

The geodesics of (R^2, g_0) are straight lines, $y_1 = at + b, y_2 = ct + d, t \in R$. The geodesics of $(R^2, g(x))$ are parabolas, $x_1 = et + f$, $x_2 = e^2t^2 + gt + h, t \in R$.

Consequently, in order to minimize f is suitable to use the descent algorithms with respect to the Riemannian metric $g(x)$, and not those with respect to the Euclidean structure g_0. Then all the above difficulties dissapear.

2. Powell function (1966)

$$f : R^2 \to R, f(x) = x_1^4 + x_1x_2 + (1 + x_2)^2, x = (x_1, x_2)$$

$$e = (0, 0), x_* = (?, ?)$$

Difficulty. For this problem

$$\min_t f(e - tH^{-1}df(e)) = f(e)$$

so that no progress can be made beyond the estimate e of x_* using Euclidean Newton method together with an exact straight line search. Indeed,

$$df = (4x_1^3 + x_2, x_1 + 2(1 + x_2)), \quad df(0, 0) = (0, 2)$$

$$H = \begin{pmatrix} 12x_1^2 & 1 \\ 1 & 2 \end{pmatrix}, \quad H(0, 0) = \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix},$$

$$H(0, 0)^{-1} = \begin{pmatrix} -2 & 1 \\ 1 & 0 \end{pmatrix}, \quad H^{-1}df(0, 0) = \begin{pmatrix} 2 \\ 0 \end{pmatrix},$$

$$\varphi(t) = f(-2t, 0) = 2^4t^4 + 1, \quad t = 0 \quad \text{is critical point,}$$

$$\min \varphi(t) = \varphi(0) = 1 = f(e).$$

The Riemannian structure which exceeds the preceding difficulty is those whose geodesics are parabolas. Indeed, let $x_1 = 2t, x_2 = -bt^2, t \in R$, be a geodesic passing through the point (0,0) with the direction (2,0) at the moment $t = 0$. We obtain

$$\varphi(t) = f(x_1(t), x_2(t)) = 2^4t^4 - 2bt^3 + (1 - bt^2)^2.$$

Imposing $\varphi'(t) = 0$, i.e., $(32 + 2b^2)t^2 - 3bt - 2b = 0$ we obtain two real roots t_1, t_2. Hence the Riemannian Newton method works.

Remark. Some recent papers [3] suggest that sometimes the changing of coordinates (which imply the changing of the Riemannian metric) can be benefic for a concrete optimization problem. But, of course, this idea is included in our general idea [4]-[9] that is enough to change conveniently the Riemannian metric in order to obtain a desired result.

2 Newton method on Riemannian manifolds

Let (M, g) be a complete n-dimensional Riemannian manifold and ∇ be the Riemannian connection determined by the metric g. The completeness implies the fact that any two points of M can be joined by a minimal geodesic.

Let X be a C^∞ vector field on M. Using the mathematical apparatus (M, g, ∇), we want to formulate a numerical algorithm for finding zeros of X. For these we need the Riemannian version of Taylor theorem together the generalizations of numerical techniques on Euclidean space to a Riemannian manifold. These are realized via an intrinsic approach which leads one from the extrinsic idea of vector addition to the exponential map and parallel translation, from the search along straight lines to search along geodesics, and from partial differentiation to covariant differentiation.

Let $\gamma : [0,1] \to M, \gamma(t) = \exp_x(tY)$ be the geodesic fixed by the initial conditions $\gamma(0) = x \in M$, $\dot\gamma(0) = Y \in T_xM$.

2.1. Meanvalue theorem. *Let W_x be a normal neighborhood of the point $x \in M$, let $\bar Y$ be a vector field on W_x adapted to $Y \in T_xM$, and X be a C^∞ vector field on W_x. Denote τ_b the parallel translation with respect to $\exp_x(tY)$ for $t \in [0, b]$, and $x_b = \exp_x(bY)$. Then there exists $\epsilon > 0$ such that for every $b \in [0, \epsilon)$, there is $a \in [0, b]$ satisfying*

$$\tau_b^{-1} X_{x_b} = X_x + b(\nabla_{\bar Y} X)_{x_a} \circ \tau_a.$$

2.2. Taylor theorem. *Same hypothesis as in theorem 2.1. Then there exists $\epsilon > 0$ such that for every $b \in [0, \epsilon)$, there is $a \in [0, b]$ satisfying*

$$\tau_b^{-1} X_{x_b} = X_x + b(\nabla_{\bar Y} X)_x + \ldots + \frac{b^{n-1}}{(n-1)!}\left(\nabla_{\bar Y}^{n-1} X\right)_x + \frac{b^n}{n!}\left(\nabla_{\bar Y}^{n} X\right)_{x_a} \circ \tau_a.$$

Let $x_* = x_b$ be a zero of the C^∞ vector field X and x be an estimate of x_*. If x is sufficiently close to x_*, we have

$$\tau_b^{-1} X_{x_b} = X_x + b(\nabla_{\bar Y} X)_x + \frac{b^2}{2}\left(\nabla_{\bar Y}^2 X\right)_{x_a} \circ \tau_a$$

and the term $\left(\nabla_{\bar Y}^2 X\right)_{x_a}$ can be neglected. Consequently $X(x_*) = 0$ is replaced by

$$b(\nabla_{\bar Y} X)_x \approx -X_x.$$

If $(\nabla X)_x : T_xM \to T_xM$ is nondegenerate, then we obtain

$$b\bar Y_x \approx -(\nabla X)_x^{-1} \circ X_x$$

and hence

$$x_* \approx \exp_x(-(\nabla X)_x^{-1} \circ X_x).$$

Consequently the most suitable direction and sense of moving from the point x towards the point x_* is given by the Newton vector

$$N_x = -(\nabla X)_x^{-1} \circ X_x.$$

Of course zeros of X are global minimum points, and hence critical points, of the energy $f = \frac{1}{2}g(X, X)$. Also, if $X(x) \neq 0$, then N_x gives a descent direction and sense for the energy f at x since $df(N_x) = -2f(x) < 0$.

The preceding considerations suggest that, given an initial estimate x_1 of x_*, we could estimate x_* with arbitrary accuracy by generating the sequence $\{x_i\}$ from

$$x_{i+1} = \exp_{x_i} N_i, \quad N_i = -(\nabla X)^{-1}_{x_i} \circ X_{x_i}.$$

This procedure is the *Newton method* for estimating a zero of X, and is embodied in the following

2.3. Newton algorithm. *Let x_1 be an initial estimate of x_*. Suppose that ∇X is nondegenerate.*

1) Set $i = 1$.

2) Compute the Newton vector $N_i = -(\nabla X)^{-1}_{x_i} \circ X_{x_i}$.

3) Compute x_{i+1} from $x_{i+1} = \exp_{x_i} N_i$.

4) Set $i = i + 1$ and go to 2).

Under certain conditions, the sequence $\{x_i\}$ generated by the Newton algorithm 2.3 converges to the zero x_* of X. In this sense the following theorems holds for general vector fields, though they are formulated for $X = gradf = g^{-1} \circ df$, where g is the Riemannian metric on M and $f : M \to R$ is a C^∞ function. Of course, in the theory appears also the (1,1) tensor field $\mathcal{H}(f) = \nabla(gradf) = g^{-1} \circ \nabla(df)$, which is symmetric with respect to the Riemannian metric, i.e., $g(\mathcal{H}(f)Y, Z) = g(Y, \mathcal{H}(f)Z)$ for any C^∞ vector fields Y, Z on M. The tensor field $\mathcal{H}(f)$ is called *(positive or negative) definite* if the associated (0,2) tensor field $H(f) = g \circ \mathcal{H}(f)$ is (positive or negative) definite. The tensor field $H(f)$ is called the *Hessian* of the function f. If x_* is a critical point of f (i.e.,$gradf(x_*) = 0$), and $H(f)(x_*)$ is nondegenerate, then the critical point x_* is called *nondegenerate.*

2.4. Theorem. *If x_* is a nondegenerate critical point of the function $f \in C^\infty(M)$, then there exists a neighborhood U of x_* such that for any $x_1 \in U$, the iterates of the Newton algorithm 2.3, for $X = gradf$, are well defined and converge quadratically to x_*.*

2.5. Corollary. *If $H(f)(x_*)$ is positive (negative) definite and the sequence generate by the Newton algorithm 2.3 converges to x_*, then the sequence converges quadratically to a local minimum (maximum) of f.*

Remark. Let T be a C^∞ tensor field on M and $f = \frac{1}{2} \parallel T \parallel^2$ be its energy. The zeros of T, i.e., the solutions of the algebraic system $T(x) = 0$, are global minimum points, and hence critical points, of the energy f. Therefore an extended descent method, for example an extended Riemannian-Newton method, can be used to find zeros of any tensor field.

3 Path of centers of a convex program

Let (M, g) be a Riemannian manifold and

$$(P) \qquad \max f_0(x) \text{ subject to } f_\alpha(x) \leq 0, \quad \alpha = 1, \ldots, m; \quad x \in M$$

be a convex program, i.e.,

1) the interior F^0 of the feasible region F is nonempty and bounded,

2) the functions $-f_0, f_\alpha$ are C^2 convex functions on F^0.

The convexity of the functions f_α implies the total convexity of the set F^0.

Let z be a lower bound for the optimal value z_*, and q be a given positive integer. To the program (P) we attach the *Huard distance function*

$$\phi(\alpha, z) = -q \ln(f_0(x) - z) - \sum_{\alpha=1}^{m} \ln(-f_\alpha(x)).$$

Suppose that ϕ has positive definite Hessian. Then ϕ achieves the minimal value in its domain (for fixed z) at a unique point $x(z)$, called *center*. The center $x(z)$ satisfies

$$-q\frac{df_0(x)}{f_0(x) - z} + \sum_{\alpha=1}^{m} \frac{df_\alpha(x)}{-f_\alpha(x)} = 0.$$

Also this point satisfies the Karush-Kuhn-Tucker conditions

$$f_\alpha(x) = 0$$

$$\sum_{\alpha=1}^{m} y_\alpha \, grad\, df_\alpha(x) = \, grad\, df_0(x), \quad y_\alpha \geq 0$$

$$f_\alpha(x) y_\alpha = \frac{f_0(x) - z}{q}, \quad \alpha = 1, \dots, m.$$

The set $\{x(z)\}$ is called the *path of centers*. The method of centers for solving the program (P) works as follows. Given z, we try to reach the vicinity of the center $x(z)$. Then we increase the lower bound z, and we try to reach the vicinity of the new center, etc. To find (an iterate close to) the center, which is in fact equivalent to minimizing ϕ, we use Newton method with approximate geodesic search procedures [1], [5]-[9], [11]. Denoting $-f_\alpha(x) = f_0(x) - z$ for $\alpha = m + 1, \dots, m + q$, the function ϕ is transcribed as

$$\phi(x, z) = -\sum_{\alpha=1}^{m+q} \ln(-f_\alpha(x)),$$

and appear a bounded totally convex set $F_z = \{x \mid f_\alpha(x) \leq 0, \alpha = 1, \dots, m + q\}$ whose interior will be denoted F_z^0. Also, for Newton method we need

$$d\phi(x, z) = \sum_{\alpha=1}^{m+q} \frac{df_\alpha(x)}{-f_\alpha(x)}$$

$$H(x, z) = H(\phi)(x, z) = \sum_{\alpha=1}^{m+q} \left[\frac{H(f_\alpha)(x)}{-f_\alpha(x)} + \frac{df_\alpha(x) \otimes df_\alpha(x)}{f_\alpha^2(x)} \right],$$

where

$$grad\phi = g^{-1} \circ d\phi,$$

and the Hessian $H(\phi)$ are built using the Riemannian metric g. Because $H(\phi)$ is positive definite (by hypothesis), it can be used like a new Riemannian metric.

Differentiating and manipulating the relations defining $x(z)$ and $y_\alpha(z)$ we can prove the following theorem about the monotonicity along the path of centers.

3.1. Theorem. *The primal objective $f_0(x(z))$ is monotonically increasing, the dual objective $f_0(x(z)) - \sum_{\alpha=1}^m y_\alpha(z) f_\alpha(x(z))$ and $f_0(x(z)) - z$ are monotonically decreasing if z increases.*

Also we can obtain an upper bound for the gap z_* - z.

3.2. Lemma. *One has*

$$z_* - z \leq \left(1 + \frac{m}{q}\right)(f_0(x(z)) - z).$$

Proof. The center $x(z)$ minimizes $\phi(x, z)$, the necessary and sufficient conditions being those of Karush - Kuhn - Tucker. It follows that $(x(z), y(z))$ is dual feasible, and we know that dual objective value is always greater than or equal to the optimal value, i.e.,

$$z_* \leq f_0(x(z)) - \sum_{\alpha=1}^{m} y_\alpha(z) f_\alpha(x(z)).$$

Hence

$$z_* - f_0(x(z)) \leq -\sum_{\alpha=1}^{m} y_\alpha(z) f_\alpha(x(z)) = \frac{m}{q}(f_0(x(z)) - z).$$

Consequently

$$(z_* - z) - (f_0(x(z)) - z \leq \frac{m}{q}(f_0(x(z)) - z)$$

or

$$z_* - z \leq \left(1 + \frac{m}{q}\right)(f_0(x(z)) - z).$$

4 Properties near the path of centers

Suppose that ϕ is k-self-concordance with $k \geq 1$. The boundedness of F^0 and the self-concordance of ϕ imply the strict convexity of ϕ, i.e., $H(\phi)(x, z) > 0$. In order to prove some properties of approximately centered points, we shall use simultaneously [1], [2], [8], [9] the Riemannian metric g and the Riemannian metric $H(\phi) = H$.

In the center method, the program (P) is replaced by a sequence of minimizing $\phi(x, z)$ using Newton method with (approximate) geodesic search procedures. Of course, the Newton vector field is $N = -H^{-1}d\phi$, and, if we use H like a Riemannian metric, $N = -grad\phi$. Note that $\| N \|_H = 0$ iff $x = x(z)$.

4.1. Theorem. *Let $x \in F_z^0, X \in T_xM$ and $\gamma(t) = \exp_x(tX), t \in [0, 1]$ be the geodesic determined by $\gamma(0) = x, \dot\gamma(0) = X$. If $\| X \|_{H(x,z)} < \frac{1}{k}$, then $\gamma(1) \in F_z^0 \subset F^0$.*

4.2. Theorem. *Let $\gamma(t) = \exp_x(tN)$ be the geodesic $\gamma : [0,1] \to M$ which verifies the initial conditions $\gamma(0) = x, \dot{\gamma}(0) = N$.*

If $\| N \|_{H(x,z)} < \frac{1}{k}$, then $x_+ = \gamma(1) \in F_z^0$, and

$$\| N(x_+, z) \|_{H(x_+,z)} \leq \frac{k}{(1 - k \| N \|_H)^2} \| N \|_H^2 .$$

For $\| N \|_H < \frac{3-\sqrt{5}}{2k}$, we find

$$\| N(x_+, z) \|_{H(x_+,z)} < \| N \|_H$$

and so the Newton algorithm is convergent. For $\| N \|_H \leq \frac{1}{3k}$, we obtain

$$\| N(x_+, z) \|_{H(x_+,z)} \leq \frac{9}{4} k \| N \|_H^2 .$$

4.3. Theorem. *If $\| N \|_H \leq \frac{1}{3k}$ and x is an approximation of the exact center $x(z)$, then*

$$\phi(x, z) - \phi(x(z), z) \leq \frac{\| N \|_H^2}{1 - (\frac{9}{4} k \| N \|_H)^2}.$$

4.4. Theorem. *If $\| N \|_H \leq \frac{1}{3k}$ and $q \geq \frac{\sqrt{m}}{k}$, then*

$$f_0(x(z)) - z \leq \left(1 + \frac{2\sqrt{m}}{q} \frac{\| N \|_H}{1 - \frac{9}{4} k \| N \|_H}\right)(f_0(x) - z).$$

Proof. We have $d\phi(N) = - \| N \|_H^2$. Since

$$d\phi(x, z) = \frac{q}{f_0(x) - z} df_0(x) - \sum_{\alpha=1}^{m} \frac{df_\alpha(x)}{-f_\alpha(x)},$$

it follows

$$df_0(x)(N) = \frac{f_0(x) - z}{q} \left(\| N \|_H^2 + \sum_{\alpha=1}^{m} \frac{df_\alpha(x)(N)}{-f_\alpha(x)}\right).$$

On the other hand [1], [9]

$$\left| \sum_{\alpha=1}^{m} \frac{df_\alpha(x)(N)}{-f_\alpha(x)} \right| \leq \sqrt{m} \| N \|_H .$$

Consequently

$$df_0(x)(N) \leq \frac{f_0(x) - z}{q} (\| N \|_H^2 + \sqrt{m} \| N \|_H) \leq 2 \| N \|_H \frac{f_0(x) - z}{q} \sqrt{m}.$$

If $x_1 = x$, and $x_2, x_3, \ldots$ is the Newton sequence starting at x_1, then

$$\begin{aligned} \| N(x_i, z) \|_{H(x_i,z)} &\leq \tfrac{9}{4} k \| N(x_{i-1}, z) \|_{H(x_{i-1},z)}^2 \\ &\vdots \\ &\leq \left(\tfrac{9}{4} k\right)^{2^i - 1} \| N(x_1, z) \|_{H(x_1,x)}^{2^i} \end{aligned}$$

and hence

$$\lim_{k\to\infty} \parallel N(x_i,z) \parallel_{H(x_i,z)}= 0.$$

Since $\{x_i\}$ is included in a bounded region, and $\parallel N(x,z) \parallel_{H(x,z)}= 0$ iff $x = x(z)$, all limit points of the sequence are $x(z)$. In other words, the Newton sequence converges to $x(z)$.

Using the inequalities and the convexity of $-f_0(x)$ we obtain

$$f_0(x_{i+1}) - z = f_0(x_i) - z + f_0(x_{i+1}) - f_0(x_i) \leq$$

$$\leq f_0(x_{i+1}) - z + df_0(x_i)(N(x_i,z)) \leq$$

$$\leq \left(1 + 2 \parallel N(x_i,z) \parallel_{H(x_i,z)} \frac{\sqrt{m}}{q}\right)(f_0(x_i) - z)$$

and hence

$$f_0(x_{i+1}) - z \leq (f_0(x_1) - z)\Pi_{j=1}^{i}\left(1 + 2 \parallel N(x_j,z) \parallel_{H(x_j,z)} \frac{\sqrt{m}}{q}\right) \leq$$

$$\leq (f_0(x_1) - z)\Pi_{j=1}^{i}\left(1 + 2\left(\frac{9}{4}k\right)^{2^j-1} \parallel N \parallel_H^{2^j} \frac{\sqrt{m}}{q}\right).$$

If $q \geq \frac{\sqrt{m}}{k}$, we obtain [1]

$$f_0(x(z)) - z \leq (f_0(x_1) - z)\Pi_{j=1}^{i}\left(1 + 2\left(\frac{9}{4}k\right)^{2^j-1} \parallel N \parallel_H^{2^j} \frac{\sqrt{m}}{q}\right) \leq$$

$$\leq \left(1 + \frac{2\sqrt{m}}{q}\frac{\parallel N \parallel_H}{1 - \frac{9}{4}k \parallel N \parallel_H}\right)(f_0(x_1) - z).$$

5 Complexity analysis

Using the results in the preceding paragraph, we can find upper bounds for the total number of outer iterations and inner iterations needed by the Center algorithm stated in [1], [8].

The analysis refers to long-, medium-, and short-step variants with $\tau = \frac{1}{3k}$.

The proofs of the propositions are similar to those in [1], [9].

5.1. Theorem. *The Center algorithm ends up with an ϵ-optimal solution for (P) after at most*

$$\frac{4}{\theta}\left(1 + \frac{m}{q}\right)\ln\frac{4\left(1 + \frac{m}{q}\right)(z_* - z_1)}{\epsilon}$$

outer iterations.

5.2. Lemma. *If* $\gamma(t) = \exp_x(tN), t \in [0,1]$ *and* $\bar{t} = \frac{1}{1+k\parallel N \parallel_H}$, *then*

$$\phi(x,z) - \theta(\gamma(\bar{t}),z) \geq \frac{1}{k^2}(k \parallel N \parallel_H - \ln(1 + k \parallel N \parallel_H)).$$

5.3. Theorem. *The total number of inner iterations during an arbitrary outer iteration is at most*

$$\frac{22}{3}+22qk^2\theta\left(\frac{\theta}{1-\theta}+\frac{3\sqrt{m}}{qk+3\sqrt{m}}\right).$$

Combining theorems 5.1 and 5.3, we obtain the total number of iterations.

5.4. Theorem. *An upper bound for the total number of Newton iterations is given by*

$$88\left(1+\frac{m}{q}\right)\left(\frac{1}{3\theta}+qk^2\left(\frac{\theta}{1-\theta}+\frac{3\sqrt{m}}{qk+3\sqrt{m}}\right)\right)\ln\frac{4\left(1+\frac{m}{q}\right)(z_*-z_1)}{\epsilon}.$$

Consequently, to obtain an ϵ-optimal solution and setting $q=\theta(m)$, the algorithms needs

$-\mathcal{O}\big(k^2 m\ln\frac{z_*-z_1}{\epsilon}\big)$ Newton iterations for the long-step variant $(0<\theta<1)$;

$-\mathcal{O}\big(k^2\sqrt{m}\ln\frac{z_*-z_1}{\epsilon}\big)$ Newton iterations for the medium-step variant for $\theta=\frac{\nu}{\sqrt{m}}>0$;

$-\mathcal{O}\big(k\sqrt{m}\ln\frac{z_*-z_1}{\epsilon}\big)$ Newton iterations for $\theta=\frac{\nu}{k\sqrt{m}}>0$.

For the short-step variant we states how the norm changes when z is increased.

5.5. Lemma. *If* $z_+=z+\theta(f_0(x)-z)$*, then*

$$\parallel N(x,z_+)\parallel_{H(x,z_+)}\leq\parallel N\parallel_H+\frac{\theta}{1-\theta}\sqrt{q}.$$

5.6. Lemma. *Let* $x_+=\gamma(1)$ *and* $z_+=\theta(f_0(x)-z)$*, where* $\theta=\frac{1}{22k\sqrt{q}}$*. If*

$$\parallel N(x,z)\parallel_{H(x,z)}\leq\frac{1}{3k},$$

then

$$\parallel N(x_+,z_+)\parallel_{H(x_+,z_+)}\leq\frac{1}{3k}.$$

Consequently, if θ is enough small, then one unit Newton step is sufficient to reach the vicinity of $x(z_+)$. By theorem 5.1, a short-step algorithm requires $\mathcal{O}\big(k\sqrt{m}\ln\frac{z_1-z_0}{6}\big)$ Newton iterations, using $q=\theta(m)$.

Aknowledgements. A version of this paper was presented at The Second International Workshop on Differential Geometry and Its Applications, Ovidius University of Constantza, Romania, Sept.25-28, 1995.

Many thanks to all the participants at my oral presentation, especially for comments and suggestions.

References

[1] D.den Hertog, *Interior point approach to linear, quadratic and convex programming*, Mathematics and Its Applications, 277, Kluwer Academic Publishers, 1994.

[2] Y.Nesterov, A.Nemirovsky, *Interior-point polynomial algorithms in convex programming*, Siam, 1994.

[3] T.Rapcsak, T.Csendes, *Nonlinear coordinate transformations for unconstrained optimization*, Journal of Global Optimization, 3(1993), 359-375.

[4] C.Udrişte, *Kuhn-Tucker theorem on Riemannian manifolds*, Kolloquia Math.Soc. Janos Bolyai, Topics in Diff. Geometry, Debrecen, Hungary (1984), 1247-1259.

[5] C.Udrişte, *Convergence of minimization methods on Riemannian manifolds*, International Workshop on Differential Geometry and Its Applications, Bucharest, July 25-30, 1993, Scientific Bulletin, Applied Mathematics and Physics, Politehnica University of Bucharest, 55, 3-4 (1993), 247-254.

[6] C.Udrişte, *Minimization algorithms on Riemannian manifolds*, Proceedings of the $23-rd$ Conference on Geometry and Topology, Cluj- Napoca, Romania, 1993, 185-193.

[7] C.Udrişte, *Convex functions and optimization methods on Riemannian manifolds*, Mathematics and Its Applications, 297, Kluwer Academic Publishers, 1994.

[8] C.Udrişte, *Optimization methods on Riemannian manifolds*, Proceedings of IRB International Workshop, Monteroduni, Italy, August 8-12, 1995, (in press).

[9] C.Udrişte, *Riemannian convexity in programming (I)*, Proceedings of The 25-th National Conference on Geometry and Topology, Al.I.Cuza University of Iassy, Romania ,September 18-23, 1995, (in press).

[10] C.Udrişte, *Riemannian convexity*, The Second International Workshop on Differential Geometry and Its Applications, Ovidius University of Constantza, Romania, Sept.25-28, 1995; this Journal pp.111-116.

[11] Ph.Wolfe, *Numerical methods for unconstrained optimization*, Van Nostrand Reinhold Company, 1987.

University Politehnica of Bucharest
Department of Mathematics I
Splaiul Independenţei 313
77206, Bucharest, Romania
e-mail: udriste@mathem.pub.ro
Fax: (401)/312.53.65