Sufficient Decrease Principle on Riemannian Manifolds

Constantin Udrişte

Abstract

Tools from Riemannian geometry (suitable Riemannian metric, exponential map, search along geodesics, covariant differentiation, sectional curvature, etc) are now used in Mathematical Programming to obtain deeply theoretical results and practical algorithms [3]-[11].

§1 lists basic propositions appearing in the numerical finding of a critical point of a real function defined on a Riemannian manifold. 2 – 3 develop the steplength analysis in terms of geodesics and Riemannian version of Taylor formula (which contains the parallel translation along geodesics), insisting on sufficient decrease principle. 4 analyses the strong influence of the sectional curvature on descent algorithms. 5 proves that the central path of a convex program is in fact a minus gradient line with respect to a suitable Riemannian metric.

The main theorems refer to the convergence of the sequence

$$x_{i+1} = \exp_{x_i}(\omega_i t_i X_i),$$

produced by a descent method, to a critical point of a function f, the convergence of the sequence

$$\{df(x_i)(e_i) \mid e_i = X_i / \parallel X_i \parallel\}$$

to zero, and the convergence of the sequence of distances $\{d(x_i, x_{i+1})\}$ to zero.

Mathematics Subject Classification: 49M10, 65K05, 53C21, 53C22, 58C27
Key words: zeros, geodesics, sufficient decrease, forcing functions, reverse modulus, curvature and descent algorithms, central path.

1 Numerical methods for finding zeros of a tensor field

Let (M, g) be a complete finite-dimensional Riemannian manifold. The Riemannian metric g produces:

Balkan Journal of Geometry and Its Applications, Vol.1, No.2, 1996, pp. 111-123

1) the energy (halph of square of the norm) of a tensor field; in particular, for a vector field X we have the energy

$$f = \frac{1}{2} \parallel X \parallel^2 = \frac{1}{2} g(X, X);$$

2) the length $L(\gamma) = \int_a^b \parallel \dot{\gamma}(t) \parallel dt$ of a piecewise C^1 curve

$$\gamma : [a, b] \to M;$$

3) the Riemannian connection ∇, the parallel translation, and the geodesics;

4) the distance

$$d(x, y) = \inf_{\gamma \in \Gamma} L(\gamma),$$

where Γ is the set of all piecewise C^1 regular curves $\gamma : [a, b] \to M$ joining the points $x, y \in M$, i.e., $\gamma(a) = x, \quad \gamma(b) = y$;

5) the raising and the lowering of the indices of tensor components;

6) the sectional curvature of the manifold; etc.

The topology induced by the distance d on M coincides with the manifold topology of M. Also (M, d) is a complete metric space. The open ball in (M, d) with center x_0 and radius r is denoted by $B(x_0, r)$.

The completeness of (M, g) implies the fact that any geodesic $\gamma : [0, 1] \to M$ can be extended to a geodesic of type $\gamma : R \to M$ and that any two points of M can be joined by a minimal geodesic.

Let $\gamma : [0, 1] \to M$ be a geodesic joining the points $\gamma(0) = x, \gamma(1) = y$. The parallel translation from x to y along γ will be denoted by τ_{xy}.

Let $\gamma(t) = \exp_x(tX)$ be the geodesic which verifies the initial conditions $\gamma(0) = x$, $\dot{\gamma}(0) = X \in T_x M$. We know that for any $x \in M$ there exists $\epsilon > 0$ such that $X \in T_x M, \parallel X \parallel < \epsilon$ imply $d(x, \exp_x X) = \parallel X \parallel$.

Generally, zeros of a C^∞ vector field or 1–form are global minimum points, and hence critical points, of their energies. The numerical methods used for the finding of such zeros has as base iterative procedures of the type [3]–[11]

$$x_{i+1} = \exp_{x_i}(r_i t_i X_i), \tag{1}$$

where the vector $X_i \in T_{x_i} M$ indicates the direction and sense of moving from the starting point x_i, the number t_i determines the steplength on the geodesic which starts from x_i tangent to X_i, and the number r_i is a *relaxation parameter*. These procedures do not depend on the local coordinate system, but depend on geodesics and the sectional curvature of the manifold.

Let us consider a C^∞ 1–form ω. If x_i is not a zero of the 1–form ω, then we select $X_i \in T_{x_i} M$ by the condition $\omega(x_i)(X_i) < 0$ having in mind at least two reasons:

1) if $\omega = df$, $f : D \subset M \to R$, the preceding inequality shows that X_i determines a direction and sense of decreasing of f, i.e., $df(X_i) < 0$;

2) if $f = \frac{1}{2} g^{-1}(\omega, \omega)$ is the energy of ω and $\nabla\omega(x_i)$ is nondegenerate, then the equalities

$$df(Y_i) = g^{-1}(\nabla_{Y_i}\omega, \omega) = \omega(g^{-1}\nabla_{Y_i}\omega), \quad X_i = g^{-1}\nabla_{Y_i}\omega$$

and the inequality $\omega(x_i)(X_i) < 0$ show that Y_i determines a descent direction and sense of the energy f (this idea corresponds to Riemann–Newton method for finding zeros of a 1–form).

Remark. The ideas in this paper hold true for general tensor fields, though they are formulated here for 1–forms. 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 extendend Riemann–Newton method, can be used to find zeros of any tensor field.

Let x_* be a zero of the 1–form ω. If $(\nabla\omega)(x_*)$ is nondegenerate, then the zero x_* is called *nondegenerate*. The Riemann–Newton method for finding of the point x_* was studied in, [3]–[11].

In this paper we refer especially to the case $\omega = df$, where $f : D \subset M \to R$ is a C^1 function. The solutions of the system $df(x) = 0$ are called *critical points* of f. The study of the convergence of the numerical procedure (1) towards the critical point x_* of f, or to a minimizer f_* of f, is based on the following propositions [6]

(2) $f(x_i + 1) \leq f(x_i), \quad i = 1, 2, \ldots$

(3) $\lim\limits_{i\to\infty} x_i = x_*, df(x_*) = 0$

(4) $\lim\limits_{i\to\infty} df(x_i)(e_i) = 0$, where $e_i = X_i / \parallel X_i \parallel$

(5) $\lim\limits_{i\to\infty} df(x_i) = 0$

(6) $df(x_i)(X_i) \leq -\varepsilon \parallel X_i \parallel, \quad \epsilon > 0, \quad \forall i \geq i_0$

(7) $\lim\limits_{i\to\infty} d(x_i, x_{i+1}) = 0.$

In §2 - §3 we shall analyse especially the basic proposition (4) showing that its validity depends only upon the steplength of the algorithm and on very mild conditions for f itself. In the theorems 2.4, 3.4 appears also the proposition (7).

2 Sufficient decrease principle on Riemannian manifolds

Let (M, g) be a complete, finite–dimensional Riemannian manifold. In this paragraph we shall develop the steplength analysis and we shall prove that proposition (4) is true for certain steplength algorithms and arbitrary $X_i \neq 0$. The decreasing condition (2) is not usually enough to imply (4), even if the inequality is strict. There exist however decreasing conditions which imply (4). These type of decreasing is called a *sufficient decrease* and it is usual described using the forcing functions.

2.1. Definition. A function $\sigma : [0, \infty) \to [0, \infty)$ for which the convergence of the sequence $\{\sigma(t_i)\}$ to zero implies the convergence of $\{t_i\}$ to zero, for any sequence $\{t_i\}$, is called a *forcing function.*

Example. Any function $\sigma : [0, \infty) \to [0, \infty)$ which is increasing and positive definite is a forcing function.

2.2. Theorem (Sufficient decrease principle). *If $f : D \subset M \to R$, and*

1) f is of class C^1,

2) f is bounded below on $D_0 \subset D$,

3) there exists a forcing function σ such that

$$f(x_{i+1}) - f(x_i) \leq -\sigma(\mid df(x_i)(e_i) \mid),\ e_i = X_i/ \parallel X_i \parallel, \tag{8}$$

then the proposition (4) is satisfied.

Proof. By (2)–(3), the sequence $\{f(x_i)\}$ is convergent as a decreasing bounded sequence. Hence $\lim_{i\to\infty}(f(x_i) - f(x_{i+1})) = 0$, and consequently by (3) the proposition (4) holds true.

In the following we shall look for estimations of type (8), for various steplength algorithms. For these it is necessary to be sure that the sequence generated by (1) remains in D_0. Denote $D_1 = \{x \in D \mid f(x) \leq f(x_1)\}$ and D_{10} the path–connected component of the sublevel set D_1 containing x_1.

2.3. Lemma. *If $f : D \to R$, where $D \subset M$ is an open set, and*

1) f is continuous on D,

2) f is of class C^1 on the compact set D_{10} for some $x_1 \in D$,

then for any $x \in D_{10}$ and $X \in T_xM$ with $df(x)(X) < 0$, and $\gamma(t) = \exp_x(tX)$, there exists t_* such that $f(x) = f(\gamma(t_*))$ and $\gamma(t) \in D_{10}$ for any $t \in (0, t_*]$.

If $s > 0$ is a number satisfying $f(\gamma(t)) < f(x),\quad \forall\gamma(t) \in \gamma([0, s]) \cap D_{10}$, then $\gamma([0, s]) \subset D_{10}$.

Proof. Let

$$J = \{t_1 > 0 \mid \gamma([0, t_1]) \subset D,\ f(\gamma(t)) < f(x),\quad \forall t \in (0, t_1]\}$$

and $t_* = \sup J$. The number t_* is well–defined since the set J is nonvoid. By compactness of D_{10}, we have $t_* < \infty$ and $\gamma([0, t_*]) \subset D_{10}$. Suppose $f(\gamma(t_*)) < f(x)$. Since D is open and f is continuous, we can select $\delta > 0$ such that $\gamma(t) \in D$ and $f(\gamma(t)) < f(x),\ \forall t \in [t_*, t_* + \delta]$, in contradiction with the definition of t_*. It rests $f(x) = f(\gamma(t_*))$. The last statement is immediate because $s < t_*$.

Denote by τ_{xy} the parallel translation from x to y along a geodesic joining the points x, y.

2.4. Theorem (Majoration principle). *Let $f : D \to R$, where $D \subset M$ is open and f is of class C^1. Suppose that D_{10} is a compact set and*

$$\parallel df(x) - \tau_{xy}^{-1} df(y) \parallel \leq ad(x, y),\quad \forall x, y \in D_{10}.$$

If the sequence $\{x_i\}$ generated by (1) satisfies the conditions:

- $\{X_i\}$ is a sequence of nonzero vectors, each vector being fixed by the conditions $X_i \in T_{x_i}M, df(x_i)(X_i) \leq 0$;

- the steplength t_i and the relaxation parameter r_i satisfy

$$t_i = -(a \parallel X_i \parallel)^{-1} df(x_i)(e_i),\ \epsilon \leq r_i \leq 2 - \epsilon,\quad i = 1, 2, ...$$

with

$$e_i = X_i/ \parallel X_i \parallel,\ \epsilon \in (0, 1),$$

then the sequence $\{x_i\}$ remains in D_{10}, and the propositions (4), (7) are satisfied.

Proof. We use the complete induction. Suppose $x_i \in D_{10}$. If $df(x_i)(X_i) = 0$, then $x_i = x_{i+1}$. Therefore we impose $df(x_i)(X_i) < 0$.

Let $\gamma_i(t) = \exp_{x_i}(tX_i)$. From

$$f(\gamma_i(t)) = f(x_i) + tdf(x_i)(X_i) + t\int_0^1 (\tau^{-1}df(\gamma_i(st)) - df(x_i))(X_i)ds,$$

$$d(x, y) = st \parallel X_i \parallel,$$

it follows

$$f(\gamma_i(t)) - f(x_i) \le tdf(x_i)(X_i) + \frac{1}{2}at^2 \parallel X_i \parallel^2,$$

whenever the geodesic $\gamma_i(t)$ is included in D. Also, Lemma 2.3, with $s = (2-\epsilon)t_i$, ensures $x_{i+1} \in D_{10}$. We have

$$f(x_i) - f(x_{i+1}) \ge -r_i t_i df(x_i)(X_i) - \frac{a}{2}(r_i t_i \parallel X_i \parallel)^2 =$$

$$= \frac{r_i}{a}(df(x_i)(e_i))^2 - \frac{r_i^2}{2}(df(x_i)(e_i))^2 = \frac{1}{2a}\left(2r_i - r_i^2\right)(df(x_i)(e_i))^2 \ge$$

$$\ge \frac{1}{2a}\epsilon(2-\epsilon)(df(x_i)(e_i))^2$$

since $2r_i - r_i^2 = 1 - (1 - r_i)^2 \ge 1 - (1-\epsilon)^2 = \epsilon(2-\epsilon)$. We remark that $\sigma(t) = \frac{\epsilon}{2a}(2-\epsilon)t^2$ is a forcing function and hence

$$\lim_{i\to\infty} df(x_i)(e_i) = 0.$$

Finally, if the selected geodesic is minimal, then

$$d(x_i, x_{i+1}) = r_i t_i \parallel X_i \parallel = -r_i a^{-1} df(x_i)(e_i)$$

implies

$$\lim_{i\to\infty} d(x_i, x_{i+1}) = 0.$$

3 Reverse modulus of continuity of a 1–form as forcing function

In the sequel, the forcing function ct^2 is replaced by a more complex forcing function. In this sense the following ideas hold true for general 1–forms, though they are formulated for $\omega = df$, where $f : D \subset M \to R$ is a C^1 function.

Let τ_{xy} be the parallel translation along a geodesic joining x to y.

3.1. Definition. Let $f : D \subset M \to R$ be a function of class C^1. Assume that on some $D_0 \subset D$ we have

$$s = \sup\left\{\parallel df(x) - \tau_{xy}^{-1}df(y) \parallel, \quad x, y \in D_0\right\} > 0,$$

and denote

$$\beta(t) = \inf\left\{d(x,y) \mid x,y \in D_0, \| df(x) - \tau_{xy}^{-1} df(y) \| \geq t\right\}, \quad t \in [0,s).$$

The function

$$\delta : [o,\infty) \to [0,\infty), \delta(t) = \begin{cases} \beta(t) & \text{for} \quad t \in [0,s) \\ \lim\limits_{t \nearrow s} \beta(t) & \text{for} \quad t \in [s,\infty) \end{cases}$$

is called the *reverse modulus of continuity of the* 1*–form* df *on* D_0.

The function δ is increasing and $\delta(0) = 0$. The following lemma shows that δ is a forcing function.

3.2. Lemma. *If the 1–form* df *is uniformly continuous on* $D_0 \subset D$ *and* s *is strictly positive, then* $\delta(t) > 0, \forall t > 0$.

Proof. Suppose $\delta(t) = 0$ for some $t > 0$. Then, given $\epsilon > 0, \exists x,y \in D_0$ with $\| df(x) - \tau_{xy}^{-1} df(y) \| \geq t$ and $d(x,y) \leq \epsilon$, contradicting the uniform continuity of df.

Let $\gamma(t) = \exp_{x_i}(tX_i), t \in [0,1]$. The sufficient decrease principle reduces to the hypotheses that $\tau_{x_i\gamma(t)}^{-1} df(\gamma(t))(X_i)$ is sufficiently smaller with respect to $df(x_i)(X_i)$. In other words, for $\mu \in [0,1)$, the suitable steplength t_i can be defined as

$$t_i = \min\left\{t \geq 0 \mid \tau_{x_i\gamma(t)}^{-1} df(\gamma(t))(X_i) = \mu df(x_i)(X_i)\right\}. \tag{9}$$

3.3. Theorem. *Suppose* $f : D \subset M \to R$ *is of class* C^1 *on the open set* D*. If* D_{10} *is compact,* $\mu \in [0,1), \epsilon \in (0,1]$ *and the iterative process (1) works under* $df(x_i)(X_i) \leq 0, X_i \neq 0, \epsilon \leq r_i \leq 1$, *and the condition (9), then the sequence* $\{x_i\}$ *is included in* D_{10} *, it is strongly downward and the proposition (4) is satisfied.*

Proof. We use the complete induction. Suppose $x_i \in D_{10}$. If $df(x_i)(X_i) = 0$, then $t_i = 0$ and hence $x_{i+1} = x_i$. Consequently it is necessary $df(x_i)(X_i) < 0$. Then Lemma 2.3 assures the existence of $s_i > 0$, with the geodesic $\gamma(s) = \exp_{x_i}(sX_i), s \in [0,s_i]$ included in D_{10} and $f(x_i) = f(x_{i+1})$. By the meanvalue theorem, there exists $\hat{s} \in (0,s_i)$ with $\tau_{x_i\gamma(\hat{s})}^{-1} df(\gamma(\hat{s}))(X_i) = 0$. Using the continuity of df, the equation in (9) has a solution in $(0,s_i)$ and $df(x_i)(X_i) < 0$ implies the existence of $t_i > 0$. Since $r_i \leq 1$, the point x_{i+1} is well defined and $x_{i+1} \in D_{10}$.

On the other hand

$$\tau_{x_i\gamma(t)}^{-1} df(\gamma(t))(X_i) = \mu df(x_i)(X_i) < 0, \quad \forall t \in [0, r_i t_i), \tag{10}$$

and hance $f(\gamma(t))$ is decreasing on $[0, r_i t_i]$. Hence

$$f(x_i) \geq f\left(\gamma_{x_i x_{i+1}}(t)\right) \geq f(x_{i+1}), \quad \forall t \in [0,1],$$

i.e., the sequence $\{x_i\}$ is strongly downward in D_{10}.

For the last part suppose $\mu > 0$. Then the meanvalue theorem and (10) give

$$f(x_i) - f(x_{i+1}) = -r_i t_i \tau_{x_i\gamma(t)}^{-1} df(\gamma(t))(X_i) \geq -t_i \epsilon \mu df(x_i)(X_i). \tag{11}$$

To avoid the triviality, we suppose that f is nonconstant on D_{10}. Lemma 3.2 shows that the reverse modulus of continuity of the 1–form df is a forcing function on D_{10}. From $e_i = X_i / \| X_i \|$,

$$(\mu-1)df(x_i)(e_i)=\tau^{-1}_{x_i\gamma(t_i)}df(\gamma(t_i))(e_i)-df(x_i)(e_i)\leq$$

$$\leq\| df(x_i)-\tau^{-1}_{x_i\gamma(t_i)}df(\gamma(t_i))\|$$

and the definition of δ, we find

$$t_i\parallel X_i\parallel\geq\delta[(\mu-1)df(x_i)(e_i)].$$

Then (11) is continued by

$$\text{(12)}\qquad f(x_i)-f(x_{i+1})\geq -t_i\parallel X_i\parallel\epsilon\mu df(x_i)(e_i)\geq\sigma(-df(x_i)(e_i)),$$

where

$$\sigma(t)=\mu\epsilon t\delta((1-\mu)t),\quad t\geq 0.$$

Since σ is a forcing function, the proof is finished for $\mu>0$.

Let $\mu=0$. We replace the steplength t_i with $\bar{t}_i$ given by Theorem 3.3 for $\mu=\frac{1}{2}$. Denoting $\bar{x}_{i+1}=\gamma(\bar{t}_i)$ it follows $f(\bar{x}_{i+1})\geq f(x_{i+1})$. Further the estimation (12) takes place with $\bar{x}_{i+1}$ instead of x_{i+1} for $\sigma(t)=\frac{1}{2}\epsilon t\delta\left(\frac{t}{2}\right)$ and hence

$$f(x_i)-f(x_{i+1})\geq f(x_i)-f(\bar{x}_{i+1})\geq\sigma(-df(x_i)(e_i)).$$

Since this last idea is very important, it will be punctuated like

Comparison principle. Suppose that two different steplength algorithms I, II produce from x_i the points x^I_{i+1} and x^{II}_{i+1}. If

$$f(x_i)-f\left(x^I_{i+1}\right)\geq\sigma(|\,df(x_i)(e_i)\,|),\ e_i=X_i/\parallel X_i\parallel,$$

where σ is a forcing function, then it is enough to prove

$$f\left(x^I_{i+1}\right)\geq f\left(x^{II}_{i+1}\right)$$

in order to obtain

$$f(x_i)-f\left(x^{II}_{i+1}\right)\geq f(x_i)-f\left(x^I_{i+1}\right)\geq\sigma(|\,df(x_i)(e_i)\,|).$$

Remark. The preceding theorem permits only relaxation factors satisfying $r_i\leq 1$. Results for $r_i>1$, require stronger conditions on f.

3.4.Theorem. *Let $D\subset M$ be open and $f:D\to R$ be of class C^2. Suppose that D_{10} is compact and*

$$a\parallel X_x\parallel^2\leq\text{Hess}f(X_x,X_x)\leq b\parallel X_x\parallel^2,\ \forall x\in D_{10},\ \forall X_x\in T_xM,$$

where $b\geq a>0$. If $\mu\in[0,1),\epsilon\in(0,1)$ and the sequence (1) is fixed by $df(x_i)(X_i)\leq 0, X_i\neq 0$, with t_i in Theorem 3.3 and $1\leq r_i\leq\bar{r}==1+\left(\frac{a}{b}\right)^{\frac{1}{2}}\cdot(1-\epsilon)$, then the sequence $\{x_i\}$ is included in D_{10}, and the propositions (4), (7) are satisfied.

Proof. We use the complete induction. Suppose $x_i\in D_{10}$ and $df(x_i)(X_i)<0$. Already we know that t_i is well–defined and $\gamma([0,t_i])\subset D_{10}, f(x_{i+1})=f(\gamma(t_i))<f(x_i)$. By the continuity of f, there exists $t\in(t_i,\bar{r}t_i]$ such that

$\gamma([0,t]) \subset D_{10}$. Denoting $y_{i+1} = \gamma(\beta), \beta \in (0, \epsilon t_i), \tau = \tau_{x_i x_{i+1}}$, the Taylor formula with the rest in the integral form implies

$$f(x_i) - f(\gamma(t)) = (f(x_i) - f(y_{i+1})) + (f(y_{i+1}) - f(x_{i+1}))-$$

$$-(f(\gamma(t)) - f(x_{i+1})) = f(x_i) - f(y_{i+1}) + (\beta - t_i)\tau^{-1} df(x_{i+1})(X_i)+$$

$$+(\beta - t_i)^2 \int_0^1 (1-s)\tau^{-1}\text{Hess} f(\gamma_{x_{i+1}}(s(t_i - \beta)X_i))(X_i, X_i)ds-$$

$$-(t - t_i)\tau^{-1} df(x_{i+1})(X_i)-$$

$$-(t_i - t)^2 \int_0^1 (1-s)\tau^{-1}\text{Hess} f(\gamma_{x_{i+1}}(s(t_i - t)X_i))(X_i, X_i)ds \geq$$

$$\geq f(x_i) - f(y_{i+1}) + (\beta - t)\tau^{-1} df(x_{i+1})(X_i)+$$

$$+\frac{1}{2}(t_i - \beta)^2 a \parallel X_i \parallel^2 - \frac{1}{2}(t_i - t)^2 b \parallel X_i \parallel^2 \geq$$

$$\geq f(x_i) - f(y_{i+1}) + (\epsilon - 1)t_i \tau^{-1} df(x_{i+1})(X_i)+$$

$$+\frac{1}{2}t_i^2 \parallel X_i \parallel^2 [(1-\epsilon)^2 a - \frac{a}{b}(1-\epsilon)^2 b] =$$

$$= f(x_i) - f(y_{i+1}) + (\epsilon - 1)\mu t_i df(x_i)(X_i) > 0.$$

Consequently $\gamma([0, \bar{r}t_i]) \subset D_{10}$ and particularly $x_{i+1} \in D_{\cdot 0}$.

We remark that

$$(\mu - 1) df(x_i)(X_i) = \tau^{-1} df(x_{i+1})(X_i) - df(x_i)(X_i) =$$

$$= t_i \int_0^1 \tau^{-1}\text{Hess}(\gamma(\text{stX}_i))(X_i, X_i)ds,$$

and hence

$$t_i b \parallel X_i \parallel \geq (\mu - 1) df(x_i)(e_i) \geq a t_i \parallel X_i \parallel \geq a d(x_i, x_{i+1}).$$

If $\mu > 0$, we find

$$f(x_i) - f(x_{i+1}) \geq (1-\epsilon)\mu t_i \parallel X_i \parallel\parallel df(x_i)(e_i) \geq$$

$$\geq (1-\epsilon)\mu[(1-\mu)/b][df(x_i)(e_i)]^2$$

and hence

$$\lim_{i\to\infty} df(x_i)(e_i) = 0, \quad \lim_{i\to\infty} d(x_i, x_{i+1}) = 0.$$

For $\mu = 0$, we can apply the comparison principle.

3.5. Theorem. *Let $D \subset M$ be an open set, $f : D \to R$ be of class C^1, and D_{10} be compact. If the sequence*

$$x_{i+1} = \exp_{x_i}(t_i X_i), \quad i = 1, 2, ...$$

is fixed by $X_i \neq 0$ and

(13) $$f(x_{i+1}) = \min_t\{f(\gamma(t)) \mid \gamma(t) = \exp_{x_i}(tX_i) \in D_{10}\},$$

then $\{x_i\}$ is included in D_{10}, $\{x_i\}$ is strongly downward and the proposition (4) is satisfied.

Proof. If $x_i \in D_{10}$, then the connected part D_{10} of the set $D_i : f(x) \leq f(x_i)$, is compact and there exist t_i satisfying (13). Hence $x_{i+1} \in D_{i0}$. Suppose $df(x_i)(X_i) \leq 0$, and $\bar{x}_{i+1}$ is obtained as in Theorem 3.3 with $\mu = 0$. Then $f(x_{i+1}) \leq f(\bar{x}_{i+1})$ and hence

$$f(x_i) - f(x_{i+1}) \geq f(x_i) - f(\bar{x}_{i+1}) \geq \sigma(-df(x_i)(e_i)), e_i = X_i / \parallel X_i \parallel,$$

with $\sigma(t) = \frac{1}{2}\epsilon t\delta\left(\frac{t}{2}\right)$. Consequently $\lim\limits_{i\to\infty} df(x_i)(e_i) = 0$.

As $\sigma([0, t_i]) \subset D_{i0}$, we have

$$f(x_i) \geq f(\gamma(t)) \geq f(x_{i+1}), \quad \forall t \in [0, 1],$$

i.e., the sequence $\{x_i\}$ is strongly downward.

4 Influence of the sectional curvature on descent algorithms

Denote by K the sectional curvature of the Riemannian manifold (M, g). If $K > 0$, then the adjacent geodesic starting at the some point tend to approximate one each other and consequently we have a liberality in selecting the decrease vector X_i and the steplength t_i along the corresponding geodesic, without distancing ourselves essentially from the critical point x_* of the function f. If $K < 0$, then the behaviour of geodesics is contrary, namely, the adjacent geodesics of a given geodesic, all starting at a given point, will go away exponentially from it; consequently, on manifolds with negative curvature we are forced to select carefully the decrease vector X_i and the steplength t_i(either enough small number or enough great number for each i).

4.1. Topogonov Theorem. *Let (M, g) be a complete Riemannian manifold with $K \geq H$, and γ_1, γ_2 be segments of normal geodesics in M with $\gamma_1(0) = \gamma_2(0)$. Let $\sum(H)$ be a 2–dimensional manifold with constant curvature H. Suppose that γ_1 is a minimal geodesic and $L(\gamma_2) \leq \frac{\pi}{\sqrt{H}}$, when $H > 0$. If $\bar{\gamma}_1, \bar{\gamma}_2$ are two geodesics in $\sum(H)$ satisfying $\bar{\gamma}_1(0) = \bar{\gamma}_2(0)$, $L(\gamma_i) = L(\bar{\gamma}_i) = L_i$, $ang(\bar{\gamma}_1'(0),\ \bar{\gamma}_2'(0)) = ang\ (\gamma_1'(0),\ \gamma_2'(0))$, then*

$$d(\gamma_1(L_1), \gamma_2(L_2)) \leq d(\bar{\gamma}_1(L_1),\ \bar{\gamma}_2(L_2)).$$

4.2. Corollary. *Let (M, g) be a complete Riemannian manifold with $K \geq 0$. If $\gamma_1(t) = \exp_x(tX_1), \gamma_2(t) = \exp_x(tX_2)$ are normal geodesics (i.e., $\parallel X_1 \parallel = \parallel X_2 \parallel = 1$), then*

$$d(\gamma_1(t_1),\ \gamma_2(t_2)) \leq \parallel t_2 X_2 - t_1 X_1 \parallel .$$

4.3. Lemma. *If $f : M \to R$ is a C^1 convex function, $\{x_i\}$ is generated by (1), with $X_i = -grad f(x_i)$ and $K \geq 0$, then*

$$d^2(x_{i+1}, y) \leq d^2(x_i, y) + t_i^2 + 2\frac{t_i}{\| X_i \|}(f(y) - f(x_i)), \quad \forall y \in M.$$

Proof. Let $\gamma_1(t) = \exp_{x_i}(tX_1)$ be a minimal geodesic, with

$$\gamma_1(0) = x_i, \quad \gamma_1(t_1) = y, \quad t_1 = d(x_i, y)$$

and let

$$\gamma_2(t) = \exp_{x_i}(tX_2), \quad df(X_2) < 0, \quad X_2 = X_i = - \text{ grad } f(x_i),$$

$$\gamma_2(0) = x_i, \quad \gamma_2(t_i) = x_{i+1}, \quad t_i = t_2.$$

From Corollary 4.2 and the convexity of C^1 functions, i.e.,

$$f(x_i) + t_1 df(X_1)(x_i) \leq f(\gamma_1(t_1)),$$

it follows

$$d^2(x_{i+1}, y) \leq \left\| t_i \frac{X_i}{\| X_i \|} - t_1 X_1 \right\|^2 = t_1^2 + t_i^2 - 2\frac{t_i}{\| X_i \|} g(X_i, t_1 X_1) \leq$$

$$\leq d^2(x_i, y) + t_i^2 + \frac{2t_i}{\| df(x_i) \|}(f(y) - f(x_i)).$$

In the hypotheses of Lemma 4.3, we have

$$d^2(x_{i+1}, z) \leq d^2(x, z) + t_i^2, \quad \forall i \in N, \quad \forall z \in \mathcal{O},$$

where $\mathcal{O} = \{z \in M \mid f(z) \leq \inf_i f(x_i)\}$.

4.4. Theorem. *Same hypotheses as in Lemma 4.3. Let O_* be the set of all minimizers of f. If $x_* \in O_*$ and $x_i \notin O_*$, then $d(x_{i+1}, x_*) < d(x_i, x_*)$ for all t_i satisfying*

$$0 < t_i < \frac{2}{\| df(x_i) \|}(f(x_i) - f(x_*)).$$

Proof. Lemma 4.3 with $y = x_*$ gives

$$d^2(x_{i+1}, x_*) \leq d^2(x_i, x_*) + t_i^2 + 2\frac{t_i}{\| df(x_i) \|}(f(x_*) - f(x_i)).$$

Since $x_i \neq x_*$, the inequality $0 < t_i < \frac{2}{\|df(x_i)\|}(f(x_i) - f(x_*))$ implies $t_i^2 + 2\frac{t_i}{\|df(x_i)\|}(f(x_*) - f(x_i)) < 0$.

In the hypotheses of Lemma 4.3, we can select a suitable steplength t_i for which the sequence (1) converges to x_* and the sequence $f(x_i)$ has an infimum [3], [8].

5 Central path of a convex program like minus gradient line

Let (M, g) be a complete n–dimensional Riemannian manifold. We consider the convex programming problem

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

The interior of the feasible region $F : f_\alpha(x) \leq 0$ is denoted by F^0, and we accept the following assumptions: 1) F^0 is nonemtpy; 2) F^0 is bounded; 3) 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.

The logarithmic barrier function associated to the preceding convex program is defined by

$$\phi(x, \mu) = -\frac{f_0(x)}{\mu} - \sum_{\alpha=1}^{m} \ln(-f_\alpha(x)),$$

where μ is the barrier strictly positive parameter. The first and the second covariant derivatives of ϕ with respect to the Riemannian connection induced by the metric g are

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

$$H(x, \mu) = \text{Hess}\phi(x, \mu) = -\frac{\text{Hess}f_0(x)}{\mu} + \sum_{\alpha=1}^{m} \left[\frac{\text{Hess}f_\alpha(x)}{-f_\alpha(x)} + \frac{df_\alpha(x) \otimes df_\alpha(x)}{f_\alpha(x)^2} \right].$$

The Hessian H is positive semidefinite since ϕ is a convex function.

Suppose H is positive definite, and we use alternatively the Riemannian manifolds (M, g) and (M, H). The function ϕ is strictly convex on F^0 in (M, g), and takes infinite values on the boundary ∂F. Consequently ϕ achieves the minimal value at a unique critical point $x = x(\mu)$, called the *μ–center*, solution of the system

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

5.1.Definition. The set of all μ–centers, when μ runs from ∞ to 0, is called the *primal central path.*

5.2. Theorem. *On the Riemannian manifold (M, g), the primal central path is a reparametrized integral curve of the vector field*

$$-H^{-1} df_0 = -\text{grad} f_0.$$

Proof. Deriving with respect to μ in the system which describes $x(\mu)$, via the covariant derivative induced by the Riemannian metric g, we obtain

$$\mu^2 H(x, \mu) \left(\frac{dx}{d\mu} \right) + df_0(x) = 0 \quad \text{or} \quad \frac{dx}{d\mu} = -\mu^{-2} H^{-1} \circ df_0.$$

Consequently, the central path is a minus gradient line for the Riemannian metric $\mu^2 H$. By the substitution $\mu = -\frac{1}{u},\ u \in (-\infty, 0)$ we find $\frac{dx}{du} = -H^{-1} \circ f_0$.

In other words, the central path is a reparametrized minus gradient line for the Riemannian metric H.

References

[1] J.Cheeger, D.Ebin, *Comparison theorems in Riemannian geometry,* North–Holland, Amsterdam, 1975.

[2] M.P.do Carmo, *Riemannian geometry*, Boston, Birkhauser, 1992.

[3] O.P.Ferreira, P.R.Oliveira, *Subgradient algorithm on Riemannian manifolds*, Publicacoes Tecnicas, COPPE, Universidade Federal do Rio de Janeiro, ES–354, 1995.

[4] S.T.Smith, *Optimization techniques on Riemannian manifolds,* Fields Institute Communications, 3(1994), 113–136.

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

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

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

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

[9] C.Udrişte, *Optimization methods on Riemannian manifolds*, IRB International Workshop, Monteroduni, Italy, August 8–12, 1995, (to appear).

[10] C.Udrişte, *Riemannian convexity in programming (I)*, The 25-th National Conference on Geometry and Topology, Al.I.Cuza University of Iassy, Romania, September 18–23, 1995; Analele Stiintifice ale Universitatii Al.I.Cuza, Iasi, 42(1996), 123–136.

[11] C.Udrişte, *Riemannian convexity in programming (II)*, The Second International Workshop on Differential Geometry and Its Applications, Ovidius Univ. of Constantza, Romania, Sept. 25–28, 1995; Balkan Journal of Geometry and Its Applications, 1, 1(1996), 99–109.

[12] C.Udrişte, *Riemannian convexity*, The Second International Workshop on Differential Geometry and Its Applications, Ovidius Univ. of Constantza, Romania, Sept.25–28, 1995; Balkan Journal of Geometry and Its Applications, 1, 1(1996), 111–116.

University Politehnica of Bucharest,
Department of Mathematics I,
Splaiul Independentei 313,
77206 Bucharest, Romania;
e–mail:udriste@mathem.pub.ro;
Fax:401/411.53.65