GENERAL DESCENT ALGORITHM ON AFFINE MANIFOLDS

C.L. Pripoae and G.T. Pripoae

Abstract

We study stopping and convergence criteria for the general descent algorithm, with affine differential notions and techniques only. An example is provided which shows that this framework is more natural for optimization problems than the Riemannian one.

Mathematical Subject Classification: 49M10, 90C30, 53B05 .
Key words : steepest descent algorithm, affine differential geometry, convergence.

1 Introduction

Given a differentiable manifold M and a C^2- differentiable function on M, one looks for

$$min\{f(x)|x \in M\}. \tag{1}$$

This is a modern formulation of some classical constrained optimization problems. Since the 70-ties, the main tool in building algorithms for solving it was Riemannian geometry (see [5] and [7] for reviews and details); but, as the classical setting (of subsets in affine $\mathbf{R}^n$) also suggests, the essence of the problem and the main notions (as convexity...) are affine ones. This is why we have extended the framework of function optimization to Affine differential geometry ([3], [4]). In [4], a general descent algorithm was described, involving only affine notions; however, the stopping and the convergence (to a critical point) criteria were expressed through metric (even not necessary Riemannian) conditions.

In this paper we give to the general descent algorithm a *purely affine* description (§2): at each step, the descent curves are auto-parallel with respect to a (step-dependent) linear connection; the stopping and the convergence criteria are expressed

Editor Gr.Tsagas *Proceedings of The Conference of Geometry and Its Applications in Technology and The Workshop on Global Analysis, Differential Geometry and Lie Algebras, 1999,* 165-169

by estimates of the values the differential of f takes on a properly choosed and parallel transported vector base. Finally, an example is provided (§3), which reflects the advantages of our algorithm comparatively to the Riemannian ones.

2 General descent algorithm - an affine approach

Let M be a n-dimensional differentiable manifold and a C^2- real differentiable function $f \in \mathcal{F}(M)$. (Usually, M proceeds from a -"smooth enough"- constrained optimization problem on $\mathbf{R}^n$.)

Denote by $\mathcal{C}(\mathcal{M})$ the affine module of linear connections on M. For each linear connection $\nabla \in \mathcal{C}(\mathcal{M})$, the couple (M, ∇) is called an affine (differential) manifold. Denote by exp^{∇} the exponential application determined by ∇, which carries rays from tangent spaces T_pM to ∇- autoparallel curves through $p \in M$. When $M = \mathbf{R}^n$ and ∇ is the canonical connection, then its parallelism coincides with the one provided by the natural affine structure of $\mathbf{R}^n$.

The general descent algorithm working on Riemannian manifolds ([7]) was extended for the affine (differential) setting ([4]):

Step 1. Set $i = 1$. Choose $x_i \in M$.

Step 2. If $df_{x_i} = 0$, then stop !

Step 3. Choose a tangent vector $X_i \in T_{x_i}M$ such that $df(X_i) < 0$

Step 4. Choose a linear connection ∇ . Determine a real number t_i such that

$$f(exp^{\nabla}_{x_i}(t_i X_i)) < f(x_i)$$

Step 5. Set $x_{i+1} = exp^{\nabla}_{x_i}(t_i X_i)$

Step 6. If x_{i+1} satisfies a given stopping criterion, then stop !

Step 6. Set $i = i + 1$ and go to Step 2.

Remarks. (i) The linear connection ∇ in Step 4 and its exponential depend on i; so, instead of choosing an apriori "absolute" Riemannian metric (with its Levi-Civita connection and geodesics) as in the Riemannian setting, we "relativize" and adapt the parallelism, requiring only step- dependent (local) affine structures.

(ii) In the *pure* affine setting, we replace the gradient of f (as a Riemannian notion) by the differential of f; then, at Step 6, we choose a basis $\{e_1^{(1)}, ..., e_n^{(1)}\}$ in $T_{x_1}M$, such that $df_{x_1}(e_1^{(j)})$ be 0 or 1 (a kind of "normalization"); then we ∇ -transport it along the ∇ -autoparallel curve $t \to exp^{\nabla}_{x_1}(tX_1)$ to a basis $\{e_1^{(2)}, ..., e_n^{(2)}\}$ of $T_{x_2}M$, and so on. If for counter i we have $|df(e_j^{(i+1)})| < \epsilon$, for every $j \in \{1, ..., n\}$, then STOP. (The procedure is coordinate free.)

(iii) Denote by γ the piecewise differentiable curve constructed from all the (step-dependent) autoparallel curves in the algorithm; we reparametrize each piece such that γ is defined for parameter values $t \geq t_1$ and on each interval $[t_i, t_{i+1}]$ differs from the respective autoparallel curve by a translation. For each natural number k, the tangent vectors $X_1, ..., X_k$ prolong to a (unique) vector field X along γ (restricted to $[t_1, t_{k+1}]$), which is (but in a finite number of points) the tangent one. Denote by X the "maximal" prolongation of all X_i. Analogously, denote $E_1, ..., E_n$ the parallel vector fields along γ modelling a moving frame like in Remark (ii).

We say f is convex along γ if at each step the Hessian of f calculated from the respective step-dependent connection is semi-positively definite.

Theorem. *Let M be a differentiable manifold, $f \in \mathcal{F}(M)$ be a C^2 differentiable function and the sequence of points $\{x_i\}_i$ be generated by the affine general descent algoritm. Suppose the "subgraph" $S = \{x \in M | f(x) < f(x_1)\}$ is compact and the function f is bounded from below.*

(i) If there exists a moving frame $E_1, ..., E_n$ along γ as above, such that for every $\epsilon > 0$, the algorithm stops according to the stopping criterion in Remark (ii), then there exists a sub- sequence $\{x_{i_\alpha}\}_{i_\alpha}$ converging to a critical point x_* of f.

(ii) If f is convex along γ, then the sequence $\{x_i\}_i$ converges to a (unique) minimum point x_*.

Proof. The Step 3 ensures us that the sequence $\{f(x_i)\}_i$ is decreasing. Since f is lower bounded, it follows that the previous sequence is convergent to a real value t^*.

The set S is compact and contains the sequence $\{x_i\}$, so there exists a sub- sequence $\{x_{i_\alpha}\}_{i_\alpha}$ convergent to a point $x_* \in S$. By continuity of *df* and by the property of the moving frame , we obtain $df_{x_*} = 0$, hence (i) is proved.

The assertion (ii) follows from general arguments about convex functions. We stress ([2]) that convexity with respect to affine connections is very analogous to Riemannian convexity (expressed by semi-positive definition of the Riemannian Hessian). □

3 An example

We consider here a comparative case study. Let $M := (0, \infty)$ and $f : M \to \mathbf{R}$, given by:

$$f(x^1, x^2) = e^{(x^1-10)^2} + (x^2 - 20)^2.$$

Obviously, the only critical point for f is $x = (10, 20)$, which is a minimum one.

How can we approach it through a steepest descent algorithm starting from $x_1 = (1, 1)$?

I. The classical case imposes movement along straight (half) lines in M, which are in fact geodesics of the (very particular) Euclidean metric of $\mathbf{R}^2$ restricted to

M. If "lucky", we may start with the best direction, namely the (unique) half-line through (1,1) and (10,20). In exchange, a Newton variant moving along $(-gradf)$ is much worse, starting with an almost horizontal deplacement, since $(-gradf)_{(1,1)} = 2(9e^{81}, 19)$.

II. The (general) Riemannian case allows to choose an apriori metric on M. As in [7], let take the Poincaré metric:

$$g_{ij} = \frac{1}{(x^2)^2}\delta_{ij}.$$

The Riemannian steepest descent algorithm ([7], [5], [1]) moves along geodesics of g, which are (half) lines and (segments of) semicircles in M. Between x_1 and x_{min} there exists a (unique) semicircle arc, which represents a best (admissible) direction corresponding to $X_1 = (3, 80)$; however, moving from x_1 along other admissible directions (such as $(-grad_P f)_{(1,1)} = (-gradf)_{(1,1)}$) may "diverge badly" from the previous one.

III. The Affine differential case allows a large choice of affine connections. We adopt a very particular one, and take the (globally defined) family of affine connections $\nabla = \nabla^{\alpha,\beta}$ on M (indexed after real parameters α and β), which (in the canonical coordinates) has vanishing coefficients but

$$\Gamma^1_{11} = \frac{\alpha}{x^1} \quad , \quad \Gamma^2_{22} = \frac{\beta}{x^2}.$$

Integrating the equations of the auto-parallel curves of ∇, we obtain that:

$$x^1(t) = \begin{cases} b_1 e^{a_1 t} & \text{if} \quad \alpha = -1 \\ (a_1 t + b_1)^{\frac{1}{\alpha+1}} & \text{if} \quad \alpha \neq -1 \end{cases} \quad , \quad x^2(t) = \begin{cases} b_2 e^{a_2 t} & \text{if} \quad \beta = -1 \\ (a_2 t + b_2)^{\frac{1}{\beta+1}} & \text{if} \quad \beta \neq -1 \end{cases} ,$$

where a_1, a_2, b_1, b_2 are arbitrary real constants. Imposing the curves start from x_1, we deduce $b_1 = b_2 = 1$. For each couple (α, β), we have a (unique) auto-parallel curve from x_1, passing through the minimum point $(10, 20)$. Their behaviour depends on the parameters values.

Conclusion . In the first two cases, the starting curve has only one "best chance" to reach fastest the minimum point. In the affine setting, we obtain (at least) a "double infinity" of such "best choices", in a thick(ly) spray of $\nabla^{\alpha,\beta}$- auto-parallel curves. (Of course, considering ALL linear connections in M, the chance to get a "best" curve, since the beginning, increases.)

References

[1] J.X. da Cruz Neto, L.L. de Lima, P.R. Oliveira, *Geodesic algorithms in Riemannian geometry*, Balkan J. Geom. Appl. 3, 2 (1998), 89- 100.

[2] J.R. Munkres, *Elementary Differential Topology*, Princeton Univ. Press, 1966.

[3] Pripoae C.L., *Affine differential invariants associated to convex functions,* Proc. Conf. SSM, Cluj, 1998 (to appear).

[4] Pripoae C.L., *General descent algorithm in affine and metric frameworks*, Proc. SSM, Craiova 1999 (to appear).

[5] Rapcsak T., *Smooth Nonlinear Optimization in* $\mathbf{R}^n$, Kluwer Acad. Publ., Dordrecht, 1997.

[6] Smith S.T., *Optimization techniques on Riemannian manifolds*, Fields Instit. Comm. AMS, 3 (1994), 113-146

[7] Udriste C., *Convex Functions and Optimization Methods on Riemannian Manifolds*, Kluwer Acad. Publ., 1994.

Authors' addresses:

C.L. Pripoae
Academy of Economic Studies,
Department of Mathematics,
6 Piata Romana, 70167, Bucharest, Romania

G.T. Pripoae
University of Bucharest,
Faculty of Mathematics,
14 Academiei Str., 70109, Bucharest, Romania
e-mail: gpripoae@geo.math.unibuc.ro