W. E. Langlois

Conditions for Termination of the Method of Steepest Descent after a Finite Number of Iterations

Certain problems in mathematical physics can be recast as extremum problems of quadratic functionals. In particular, considerable attention has been given to the minimization of a quadratic polynomial associated with a bounded and positive semidefinite linear operator T defined on a Hilbert space H, viz., the quadratic polynomial $Q(x) = [Tx, x] - [x, g] - [g, x]$, where g is a fixed element of H and [,] denotes dot product.

In principle, the complete solution to this problem is embodied in the well-known result:

If the projection g_0 of g on the null-space H_0 of T is the zero-vector, then

$$Q(T_1^{-1}g + v) = \min_{x \varepsilon H} Q(x) = -||T_1^{-1/2}g||^2,$$

where T_1 is the restriction of T to $H_0^{\perp}$ and v is any vector in H_0; if g_0 is not zero, then $Q(x)$ is not bounded from below.[1] However, this solution is not constructive, for it involves the operator T_1^{-1}, which cannot always be calculated explicitly.

Kantorovich[2] has recast the classical Method of Steepest Descent in a Hilbert space setting, thus providing an algorithm which determines a minimizing sequence for $Q(x)$. If $g \varepsilon H_0^{\perp}$, so that $Q(x)$ has a finite minimum, the method proceeds as follows:

A first approximation x_0 is chosen. If $Tx_0 = g$, we have made a lucky guess, for $Q(x_0) = Q(T_1^{-1}g)$, which is the minimum. Otherwise, take

$$z_0 = Tx_0 - g,$$

$$\epsilon_0 = [z_0, z_0]/[Tz_0, z_0],$$

$$x_1 = x_0 - \epsilon_0 z_0.$$

If $Tx_1 \neq g$, the process continues. Thus, at the n^{th} step, one takes

$$z_n = Tx_n - g,$$

$$\epsilon_n = \begin{cases} [z_n, z_n]/[Tz_n, z_n] & z_n \neq 0 \\ 0 & z_n = 0 \end{cases},$$

$$x_{n+1} = x_n - \epsilon_n z_n.$$

Thus, the corrections at successive steps are related by the nonlinear transformation $z_{n+1} = z_n - \epsilon_n T z_n$.

Kantorovich[2] not only established that $\{x_n\}$ is a minimizing sequence, in the sense that

$$\lim_{n\to\infty} Q(x_n) = \min_{x \varepsilon H} Q(x),$$

but showed further that, if T is positive-definite or if zero is an isolated point of the spectrum, the sequence $\{x_n\}$ itself converges in the strong topology with the speed of a geometrical progression. Balakrishnan[3] has recently provided a short proof of the convergence of $Q(x_n)$ to the minimum value.[4]

In the present study, we investigate whether it is possible for Steepest Descent to terminate after a finite number of of iterations. We find that this happens if and only if z_0 is an eigenvector, in which case the method terminates with the first iteration.

A preliminary result is elementary enough to be proved quite generally. We need not even assume T is linear, provided we take "eigenvector" to mean any nonzero vector v for which there exists a complex number λ such that $Tv = \lambda v$. We have:

Lemma: Let T be a transformation defined on a Hilbert space H. For those nonzero z in H such that $[Tz, z] \neq 0$, define

$$\epsilon = \frac{[z, z]}{[Tz, z]},$$

$$w = z - \epsilon Tz.$$

Then $w = 0$ if and only if z is an eigenvector of T.

Proof: The necessity is trivial. To establish the sufficiency, assume that z is an eigenvector, so that $Tz = \lambda z$ for some $\lambda \neq 0$. Then $[Tz, z] = \lambda[z, z]$, so that $\epsilon\lambda = 1$. Hence,

$$w = z - \epsilon Tz = z - \epsilon\lambda z = 0. \qquad Q.E.D.$$

It follows immediately that the Method of Steepest Descent terminates at step $(n + 1)$ if and only if z_n is an eigenvector of T. Thus, it is natural to enquire when z_n is an eigenvector. The answer is provided by the following:

Theorem: *Let T be a bounded and self-adjoint linear operator defined on a Hilbert space H. For z not in the null-space of T define*

$$\epsilon = \frac{[z, z]}{[Tz, z]},$$

$$w = z - \epsilon Tz.$$

Then w is not an eigenvector of T, no matter what choice is made for z.

Proof: If T has no eigenvectors, the conclusion follows trivially. If z is an eigenvector, the lemma shows that $w = 0$, which is not an eigenvector if only by definition. Thus we need only rule out the possibility that T has eigenvectors which include w but not z.

Assume there exists an eigenvalue λ such that $w = z - \epsilon Tz$ is in the eigenmanifold M_λ and take

$$z = x + y, \qquad x \varepsilon M_\lambda, \qquad y \varepsilon M_\lambda^\perp.$$

Then

$$w = (1 - \epsilon\lambda)x + v,$$

where

$$v \overset{\text{def.}}{=} y - \epsilon Ty.$$

Since T is self-adjoint and $y \varepsilon M_\lambda^\perp$, we have $Ty \varepsilon M_\lambda^\perp$ and hence $v \varepsilon M_\lambda^\perp$.

On the other hand, $w \varepsilon M_\lambda$ and $x \varepsilon M_\lambda$, therefore $v \equiv w - (1 - \epsilon\lambda)x \varepsilon M_\lambda$, and consequently $v = 0$, i.e., $Ty = \epsilon^{-1}y$. Thus y is an eigenvector with eigenvalue ϵ^{-1}, so that $[Ty, y] = \epsilon^{-1} ||y||^2$. From the definition of ϵ we then have

$$\epsilon = \frac{||x||^2 + ||y||^2}{\lambda\, ||x||^2 + \epsilon^{-1}\, ||y||^2},$$

so that

$$(1 - \epsilon\lambda)\, ||x||^2 = 0.$$

If $\epsilon\lambda \neq 1$ then $x = 0$, so that $w = v = 0$. On the other hand, if $\epsilon\lambda = 1$ then $y \varepsilon M_\lambda \cap M_\lambda^\perp = \{0\}$. In this case z is an eigenvector of T; by the lemma of Ref. 2 we again have $w = 0$. *Q.E.D.*

The Method of Steepest Descent applies when the bounded linear operator T is positive semidefinite, hence self-adjoint, so that the hypotheses of the theorem are satisfied. Hence, if x_0 is not itself an exact solution to the minimum problem, we have the alternative:

Either: z_0 is an eigenvector and $z_n = 0$, $n = 1, 2, 3, \cdots$

Or: No z_n is an eigenvector and the method converges only in the limit as $n \to \infty$.

It seems appropriate to point out that a lucid discussion of Steepest Descent has recently become available in English translation (Chapter XV of Ref. 5).

References and footnotes

1. *A short proof;* Set $x = x_0 + x_i$, where $x_0 \varepsilon H_0$ and $x_i \varepsilon H_0^\perp$. Then, since a positive semidefinite linear operator is by definition self-adjoint,

$$\begin{aligned} Q(x) &= [Tx_1, x_1] - [x, g_1] - [g_1, x_1] - [x_0, g_0] - [g_0, x_0] \\ &= [\sqrt{T}x_1, \sqrt{T}x_1] - [\sqrt{T}x_1, T_1^{-1/2}g_1] \\ &\qquad - [T_1^{-1/2}g_1, \sqrt{T}x_1] - [x_0, g_0] - [g_0, x_0] \\ &= ||\sqrt{T}x_1 - T_1^{-1/2}g_1||^2 - ||T_1^{-1/2}g_1||^2 \\ &\qquad - [x_0, g_0] - [g_0, x_0]. \end{aligned}$$

If $g_0 = 0$, then

$$\begin{aligned} Q(x) &= ||\sqrt{T}x_1 - T_1^{-1/2}g||^2 - ||T_1^{-1/2}g||^2 \\ &\geq -||T_1^{-1/2}g||^2 = Q(T_1^{-1}g) = Q(T_1^{-1}g + v). \end{aligned}$$

If $g_0 \neq 0$, then $\inf_{x \varepsilon H} Q(x) \leq \inf_{\lambda \text{real}} Q(\lambda g_0)$

$$\begin{aligned} &= -\sup_{\lambda \text{real}} ([\lambda g_0, g_0] + [g_0, \lambda g_0]) \\ &= -2\, ||g_0||^2 \sup_{\lambda \text{real}} \lambda = -\infty. \end{aligned}$$

2. L. V. Kantorovich, *Functional Analysis and Applied Mathematics.* Translated from the Russian by C. D. Benster, National Bureau of Standards Report 1509, 1952.
3. A. V. Balakrishnan, "A General Theory of Nonlinear Estimation Problems in Control Systems," *J. Math. Anal. and Applications* 8, 4–30 (1964).
4. *Balakrishnan's proof:* Assume no z_n is zero, for otherwise the result is trivial. Since $g \varepsilon H_0^\perp$,

$$-\infty < \min_{x \varepsilon H} Q(x) \leq Q(x_{n+1}) = Q(x_n) - \epsilon_n\, ||z_n||^2 \leq Q(x_n),$$

so that $\{Q(x_n)\}$ converges. Thus, $\epsilon_n\, ||z_n||^2 \to 0$. But

$$\epsilon_n = \frac{[z_n, z_n]}{[Tz_n, z_n]} = \frac{||z_n||^2}{||\sqrt{T}z_n||^2} \geq \frac{1}{||\sqrt{T}||^2} > 0,$$

so that

$$||z_n|| = ||Tx_n - g|| = ||\sqrt{T}(\sqrt{T}x_n - T_1^{-1/2}g)|| \to 0.$$

Since $\sqrt{T}x_n$ and $T_1^{-1/2}\, g$ are both in $H_0^\perp$, this implies

$$\begin{aligned} Q(x_n) - \min Q(x) &= Q(x_n) + ||T_1^{-1/2}g||^2 \\ &= ||\sqrt{T}x_n - T_1^{-1/2}g|| \to 0. \end{aligned}$$

5. L. V. Kantorovich and G. P. Akilov, *Functional Analysis in Normed Spaces*, Pergamon Press, 1964.

Received July 7, 1965.