Proposition 1.2.6: (Convergence for Spacer Steps) Consider a sequence $\{x^k\}$ such that

$$f(x^{k+1}) \leq f(x^k), \qquad k = 0, 1, \ldots$$

Assume that there exists an infinite set $\mathcal{K}$ of integers for which

$$x^{k+1} = x^k + \alpha^k d^k, \qquad \forall\, k \in \mathcal{K},$$

where $\{d^k\}_\mathcal{K}$ is gradient related and α^k is chosen by the minimization rule, or the limited minimization rule, or the Armijo rule. Then every limit point of the subsequence $\{x^k\}_\mathcal{K}$ is a stationary point.

EXERCISES

1.2.1

Consider the problem of minimizing the function of two variables $f(x, y) = 3x^2 + y^4$.

(a) Apply one iteration of the steepest descent method with $(1, -2)$ as the starting point and with the stepsize chosen by the Armijo rule with $s = 1$, $\sigma = 0.1$, and $\beta = 0.5$.

(b) Repeat (a) using $s = 1$, $\sigma = 0.1$, $\beta = 0.1$ instead. How does the cost of the new iterate compare to that obtained in (a)? Comment on the tradeoffs involved in the choice of β.

(c) Apply one iteration of Newton's method with the same starting point and stepsize rule as in (a). How does the cost of the new iterate compare to that obtained in (a)? How about the amount of work involved in finding the new iterate?

1.2.2

Describe the behavior of the steepest descent method with constant stepsize s for the function $f(x) = \|x\|^{2+\beta}$, where $\beta \geq 0$. For which values of s and x^0 does the method converge to $x^* = 0$. Relate your answer to the assumptions of Prop. 1.2.3.

1.2.3

Consider the function $f : \Re^n \mapsto \Re$ given by

$$f(x) = \|x\|^{3/2},$$

and the method of steepest descent with a constant stepsize. Show that for this function, the Lipschitz condition $\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|$ for all x and y is not satisfied for any L. Furthermore, for any value of constant stepsize, the method either converges in a *finite* number of iterations to the minimizing point $x^* = 0$ or else it does not converge to x^*.

1.2.4

Apply the steepest descent method with constant stepsize α to the function f of Exercise 1.1.11. Show that the gradient ∇f satisfies the Lipschitz condition

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \qquad \forall\, x, y \in \Re,$$

for some constant L. Write a computer program to verify that the method is a descent method for $\alpha \in (0, 2/L)$. Do you expect to get in the limit the global minimum $x^* = 0$?

1.2.5 (www)

Suppose that the Lipschitz condition

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \qquad \forall\, x, y \in \Re^n,$$

[cf. Eq. (1.20)] is replaced by the condition that for every bounded set $A \subset \Re^n$, there exists some constant L such that

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|, \qquad \forall\, x, y \in A. \tag{1.33}$$

Show that:

(a) Condition (1.33) is always satisfied if the level sets $\{x \mid f(x) \leq c\}$, $c \in \Re$ are bounded, and f is twice continuously differentiable.

(b) The convergence result of Prop. 1.2.3 remains valid provided that the level set

$$A = \{x \mid f(x) \leq f(x^0)\}$$

is bounded and the stepsize is allowed to depend on the choice of the initial vector x^0. *Hint*: The key idea is to show that x^k stays in the set A, and to use a stepsize α^k that depends on the constant L corresponding to this set. Let

$$R = \max\{\|x\| \mid x \in A\},$$

$$G = \max\{\|\nabla f(x)\| \mid x \in A\},$$

and

$$B = \{x \mid \|x\| \le R + 2G\}.$$

Using condition (1.33), there exists some constant L such that $\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|$, for all $x, y \in B$. Suppose the stepsize α^k satisfies

$$0 < \epsilon \le \alpha^k \le (2 - \epsilon)\gamma^k \min\{1, 1/L\},$$

where

$$\gamma^k = \frac{|\nabla f(x^k)'d^k|}{\|d^k\|^2}.$$

Let $\beta^k = \alpha^k(\gamma^k - L\alpha^k/2)$, which can be seen to satisfy $\beta^k \ge \epsilon^2\gamma^k/2$ by our choice of α^k. Show by induction on k that with such a choice of stepsize, we have $x^k \in A$ and

$$f(x^{k+1}) \le f(x^k) - \beta^k\|d^k\|^2, \qquad \forall\, k \ge 0.$$

1.2.6

Suppose that f is quadratic and of the form $f(x) = \frac{1}{2}x'Qx - b'x$, where Q is positive definite and symmetric.

(a) Show that the Lipschitz condition $\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|$ is satisfied with L equal to the maximal eigenvalue of Q.

(b) Consider the gradient method $x^{k+1} = x^k - sD\nabla f(x^k)$, where D is positive definite and symmetric. Show that the method converges to $x^* = Q^{-1}b$ for every starting point x^0 if and only if $s \in (0, 2/\bar{L})$, where $\bar{L}$ is the maximum eigenvalue of $D^{1/2}QD^{1/2}$.

1.2.7

An electrical engineer wants to maximize the current I between two points A and B of a complex network by adjusting the values x_1 and x_2 of two variable resistors, where $0 \le x_1 \le R_1$, $0 \le x_2 \le R_2$, and R_1, R_2 are given. The engineer does not have an adequate mathematical model of the network and decides to adopt the following procedure. She keeps the value x_2 of the second resistor fixed and adjusts the value of the first resistor until the current I is maximized. She then keeps the value x_1 of the first resistor fixed and adjusts the value of the second resistor until the current I is maximized. She then repeats the procedure until no further progress can be made. She knows *a priori* that during this procedure, the values x_1 and x_2 can never reach their extreme values 0, R_1, and R_2. Explain whether there is a sound theoretical basis for the engineer's procedure. *Hint*: Consider how the steepest descent method works for two-dimensional problems.

1.2.8

Consider the gradient method $x^{k+1} = x^k + \alpha^k d^k$, where α^k is chosen by the Armijo rule or the line minimization rule and

$$d^k = -\begin{bmatrix} 0 \\ \vdots \\ 0 \\ \frac{\partial f(x^k)}{\partial x_i} \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

where i is the index for which $\left|\partial f(x^k)/\partial x_j\right|$ is maximized over $j = 1, \ldots, n$. Show that every limit point of $\{x^k\}$ is stationary.

1.2.9

Consider the gradient method $x^{k+1} = x^k + \alpha^k d^k$ for the case where f is positive definite quadratic, and let $\bar{\alpha}^k$ be the stepsize corresponding to the line minimization rule. Show that a stepsize α^k satisfies the inequalities of the Goldstein rule if and only if

$$2\sigma\bar{\alpha}^k \leq \alpha^k \leq 2(1-\sigma)\bar{\alpha}^k.$$

1.2.10 (www)

Let f be twice continuously differentiable. Suppose that x^* is a local minimum such that for all x in an open sphere S centered at x^*, we have, for some $m > 0$,

$$m\|d\|^2 \leq d'\nabla^2 f(x) d, \qquad \forall\, d \in \Re^n.$$

Show that for every $x \in S$, we have

$$\|x - x^*\| \leq \frac{\|\nabla f(x)\|}{m}, \qquad f(x) - f(x^*) \leq \frac{\|\nabla f(x)\|^2}{2m}.$$

Hint: Use the relation

$$\nabla f(y) = \nabla f(x) + \int_0^1 \nabla^2 f\big(x + t(y-x)\big)(y-x)\,dt.$$

See also Exercise 1.1.9.