EXERCISES

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?

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.

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^*.

2.4

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

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

Hint: Use the relation

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

2.5

Suppose that the Lipschitz condition

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

[cf. Eq. (2.16)] is replaced by the following two conditions:

(i) For every bounded set $A \subset \Re^n$, there exists some constant L such that

$$\|\nabla f(x) - \nabla f(y)\| \le L\|x - y\|, \qquad \forall\, x, y \in A.$$

(ii) The set $\big\{x \mid f(x) \le c\big\}$ is bounded for every $c \in \Re$.

Show that:

(a) Condition (i) is always satisfied if f is twice continuously differentiable.

(b) The convergence result of Prop. 1.2.3 remains valid provided that the constant stepsize s is allowed to depend on the choice of the initial vector x^0. *Hint*: Choose a stepsize that guarantees that x^k stays within the level set $\big\{x \mid f(x) \le f(x^0)\big\}$ for all k.

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 $x^* = Q^{-1}b$ if $s \in (0, 2/\bar{L})$, where $\bar{L}$ is the maximum eigenvalue of $D^{1/2}QD^{1/2}$.

$) \le \frac{\|\nabla f(x)\|^2}{m}.$

$x)\big)(y - x)dt.$

$\forall\, x, y \in \Re^n,$

onditions:

s some constant L such that

$\forall\, x, y \in A.$

y $c \in \Re$.

e continuously differentiable.

ains valid provided that the he choice of the initial vector ees that x^k stays within the

$) = \frac{1}{2}x'Qx - b'x$, where Q is

$x) - \nabla f(y)\| \le L\|x - y\|$ is nvalue of Q.

$^k - sD\nabla f(x^k)$, where D is hat the method converges to the maximum eigenvalue of

2.7

Apply the steepest descent method with constant stepsize α to the function f of Exercise 1.11 of Section 1.1. Show that the gradient ∇f satisfies the Lipschitz condition

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

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 $x^* = 0$?

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.

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

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

2.10 (Alternative Assumptions for Convergence)

Consider the gradient method $x^{k+1} = x^k + \alpha^k d^k$. Instead of $\{d^k\}$ being gradient related, assume one of the following two conditions:

(1) It can be shown that for any subsequence $\{x^k\}_{k\in\mathcal{K}}$ that converges to a nonstationary point, the corresponding subsequence $\{d^k\}_{k\in\mathcal{K}}$ is bounded and satisfies

$$\liminf_{k\to\infty,\, k\in\mathcal{K}} \nabla f(x^k)'d^k < 0.$$

(2) α^k is chosen by the minimization rule, and for some $c > 0$ and all k, we have

$$|\nabla f(x^k)'d^k| \geq c\|\nabla f(x^k)\|\,\|d^k\|.$$

Show that the result of Prop. 1.2.1 holds.

2.11

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 $0 \leq x_1 \leq R_1$, $0 \leq x_2 \leq R_2$, 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.

2.12 (Behavior of Steepest Descent Near a Saddle Point)

Let $f = (1/2)x'Qx$, where Q is invertible and has at least one negative eigenvalue. Consider the steepest descent method with constant stepsize and show that unless the starting point x^0 belongs to the subspace spanned by the eigenvectors of Q corresponding to the nonnegative eigenvalues, the generated sequence $\{x^k\}$ diverges.

2.13 (Convergence to a Single Limit)

Consider the steepest descent method $x^{k+1} = x^k - \alpha^k \nabla f(x^k)$ and assume that for all x, y, we have

$$\frac{\|\nabla f(x) - \nabla f(y)\|^2}{L} \leq \big(\nabla f(x) - \nabla f(y)\big)'(x - y).$$

(It can be shown that this condition holds if f is convex, twice continuously differentiable and its Hessian matrix has eigenvalues that are less than or equal to L.) Assume also that f has at least one stationary point. Show that $\{x^k\}$ converges to a stationary point of f under one of the following two conditions:

(i) For some $\epsilon > 0$, we have

$$\epsilon \leq \alpha^k \leq \frac{2 - \epsilon}{L}, \qquad \forall\, k.$$

;, and for some $c > 0$ and all k, we

$f(x^k)\| \, \|d^k\|.$

he current I between two points A
he values x_1 and x_2 of two variable
: R_2, and R_1, R_2 are given. The
ematical model of the network and
she keeps the value x_2 of the second
first resistor until the current I is
f the first resistor fixed and adjusts
current I is maximized. She then
ogress can be made. She knows *a*
ıes x_1 and x_2 can never reach their
hether there is a sound theoretical
Consider how the steepest descent
ms.

Near a Saddle Point)

and has at least one negative eigen-
ıod with constant stepsize and show
;s to the subspace spanned by the
ıonnegative eigenvalues, the gener-

·)

$^{k+1} = x^k - \alpha^k \nabla f(x^k)$ and assume

$) - \nabla f(y))'(x - y).$

ds if f is convex, twice continuously
ıs eigenvalues that are less than or
ıt least one stationary point. Show
t of f under one of the following two

$-\epsilon$, $\quad \forall\, k.$

(ii) $\alpha^k \to 0$ and $\sum_{k=0}^{\infty} \alpha^k = \infty$.

Hint: Show that for any stationary point $\overline{x}$ we have

$$\|x^{k+1} - \overline{x}\|^2 \le \|x^k - \overline{x}\|^2 - \alpha^k \left(\frac{2}{L} - \alpha^k\right) \|\nabla f(x^k)\|^2.$$

2.14 (Steepest Descent with Diminishing Stepsize [CoL94])

Consider the steepest descent method

$$x^{k+1} = x^k - \alpha^k \nabla f(x^k),$$

assuming that the function f is convex.

(a) Use the convexity of f to show that for any $y \in \Re^n$, we have

$$\|x^{k+1} - y\|^2 \le \|x^k - y\|^2 - 2\alpha^k \big(f(x^k) - f(y)\big) + \big(\alpha^k \|\nabla f(x^k)\|\big)^2.$$

(b) Assume that

$$\sum_{k=0}^{\infty} \alpha^k = \infty, \qquad \alpha^k \|\nabla f(x^k)\|^2 \to 0.$$

Show that $\liminf_{k\to\infty} f(x^k) = \inf_{x\in\Re^n} f(x)$. *Hint*: Argue by contradiction. Assume that for some $\delta > 0$, there exists y with $f(y) < f(x^k) - \delta$ for all k sufficiently large. Use part (a).

(c) Assume that

$$\alpha^k = \frac{s^k}{\|\nabla f(x^k)\|},$$

where

$$\sum_{k=0}^{\infty} s^k = \infty, \qquad \sum_{k=0}^{\infty} (s^k)^2 < \infty.$$

Show that $\liminf_{k\to\infty} f(x^k) = \inf_{x\in\Re^n} f(x)$, and that if f has at least one global minimum, then $\{x^k\}$ converges to some global minimum. *Hint*: In part (a), set y to some x^* such that $f(x^*) < f(x^k)$ for all k (if no such x^* exists, we are done). Show that the relation

$$\|x^{k+1} - x^*\|^2 \le \|x^k - x^*\|^2 + (s^k)^2$$

implies that $\{x^k\}$ is bounded and hence also that $\{\nabla f(x^k)\}$ is bounded. Use part (b).

2.15 (Divergence with Diminishing Stepsize)

Consider the one-dimensional function

$$f(x) = \frac{2}{3}|x|^3 + \frac{1}{2}x^2$$

and the method of steepest descent with stepsize $\alpha^k = \gamma/(k+1)$, where γ is a positive scalar.

(a) Show that for $\gamma = 1$ and $|x^0| \geq 1$ the method diverges. In particular, show that $|x^k| \geq k+1$ for all k.

(b) Characterize as best as you can the set of pairs (γ, x^0) for which the method converges to $x^* = 0$.

(c) How do you reconcile the results of (a) and (b) with Prop. 1.2.4.

2.16

There are several criteria for implementing approximately the minimization rule in a gradient method. An example of such a criterion is that α^k satisfies simultaneously

$$f(x^k) - f(x^k + \alpha^k d^k) \geq -\sigma\alpha^k \nabla f(x^k)'d^k, \tag{2.26}$$

$$\nabla f(x^k + \alpha^k d^k)'d^k \geq \beta \nabla f(x^k)'d^k, \tag{2.27}$$

where α and β are some scalars with $\sigma \in (0, 1/2)$ and $\beta \in (\sigma, 1)$. If α^k is indeed a minimizing stepsize, then $\nabla f(x^k + \alpha^k d^k)'d^k = dg(\alpha^k)/d\alpha = 0$, where g is the function $g(\alpha) = f(x^k + \alpha d^k)$, so Eq. (2.27) is in effect a test on the accuracy of the minimization (see Fig. 1.2.11).

(a) Show that if conditions (2.26) and (2.27) are satisfied by a gradient method at each iteration and the direction sequence is gradient related, then all limit points of the generated sequence $\{x^k\}$ are stationary points of f.

(b) Assume that there is a scalar M such that $f(x) \geq M$. Show that there exists an interval $[c_1, c_2]$ with $0 < c_1 < c_2$, such that every $\alpha \in [c_1, c_2]$ satisfies Eqs. (2.26) and (2.27).

2.17 (Steepest Descent with Errors)

Consider the steepest descent method $x^{k+1} = x^k - \alpha^k\big(\nabla f(x^k) + e^k\big)$, where e^k is an error satisfying $\|e^k\| \leq \delta$ for all k. Show that for any $\delta' > \delta$, there exists a range of positive stepsizes $[\underline{\alpha}, \overline{\alpha}]$ such that if $\alpha^k \in [\underline{\alpha}, \overline{\alpha}]$ for all k, then either $f(x^k) \to -\infty$ or there is at least one limit point of $\{x^k\}$ in the set $\big\{x \mid \|\nabla f(x^k)\| < \delta'\big\}$. *Hint*: Use Prop. 1.2.3.

ze)

$\alpha^k = \gamma/(k+1)$, where γ is

hod diverges. In particular,

f pairs (γ, x^0) for which the

l (b) with Prop. 1.2.4.

oximately the minimization
criterion is that α^k satisfies

$f(x^k)'d^k,$ (2.26)

$)'d^k,$ (2.27)

/2) and $\beta \in (\sigma, 1)$. If α^k is
$'d^k = dg(\alpha^k)/d\alpha = 0$, where
27) is in effect a test on the

are satisfied by a gradient
on sequence is gradient re-
sequence $\{x^k\}$ are stationary

$f(x) \geq M$. Show that there
such that every $\alpha \in [c_1, c_2]$

$- \alpha^k\big(\nabla f(x^k) + e^k\big)$, where
w that for any $\delta' > \delta$, there
hat if $\alpha^k \in [\underline{\alpha}, \overline{\alpha}]$ for all k,
limit point of $\{x^k\}$ in the set

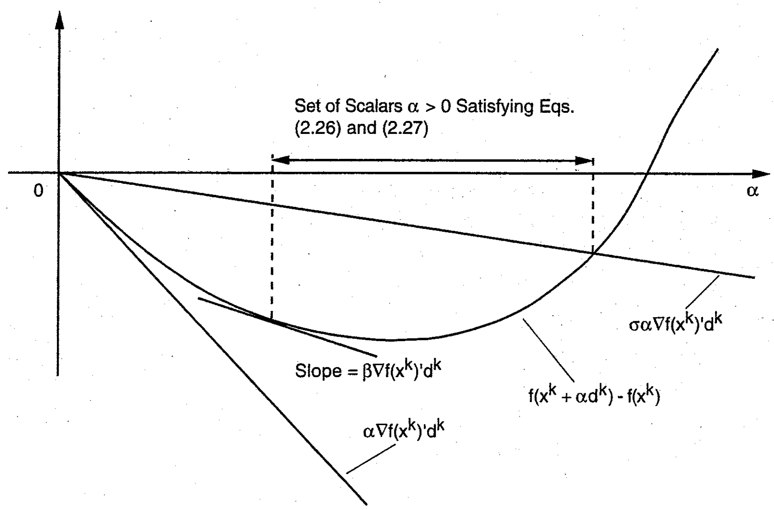

Figure 1.2.11. Illustration of the stepsize selection criterion based on Eqs. (2.26) and (2.27).

2.18 (A Continuum of Limit Points for Steepest Descent [Zou76])

Consider the two-dimensional function

$$f(x) = \begin{cases} (r-1)^2 - \frac{1}{2}(r-1)^2 \cos\left(\frac{1}{r-1} - \phi\right) & \text{if } r \neq 1, \\ 0 & \text{if } r = 1, \end{cases}$$

where

$$r = \sqrt{x_1^2 + x_2^2}, \qquad \phi = \arctan(x_1/x_2).$$

This function is minimized at each point of the circle where $r = 1$. Consider a nonoptimal starting point and the method of steepest descent where x^{k+1} is set equal to the first local minimum along the line $\{x^k - \alpha\nabla f(x^k) \mid \alpha \geq 0\}$. Show that this method follows a spiral path that comes arbitrarily close to every point of the circle of optimal points.

2.19 (Simplified Steepest Descent)

(a) Consider the unconstrained minimization of a function f of the form

$$f(x) = F\big(x, g(x)\big),$$

where $g : \Re^m \to \Re^n$ is continuously differentiable and $F(x, y)$ is a continuously differentiable function of the two arguments $x \in \Re^n$ and $y \in \Re^m$. It is sometimes convenient to approximate the gradient of $F\big(x, g(x)\big)$ by neglecting the dependence on g. This leads to the method

$$x^{k+1} = x^k - \alpha^k \nabla_x F\big(x^k, g(x^k)\big),$$

where α^k is chosen by the minimization rule or the Armijo rule on the function f. (Such a method makes sense when $\nabla_x F$ is much easier to compute than $\nabla g \nabla_y F$.) Show that if there exists $\gamma \in (0,1)$ such that

$$\left\|\nabla g(x)\nabla_y F\big(x, g(x)\big)\right\| \leq \gamma \left\|\nabla_x F\big(x, g(x)\big)\right\|, \qquad \forall\, x \in \Re^n,$$

then the method is convergent in the sense that all limit points of the sequences that it generates are stationary points of f.

(b) Consider the constrained minimization problem

$$\begin{aligned} &\text{minimize} \;\; f(x,y) \\ &\text{subject to} \;\; h(x,y) = 0 \end{aligned}$$

where $f : \Re^{n+m} \to \Re$ and $h : \Re^{n+m} \to \Re^m$ are continuously differentiable functions of the two arguments $x \in \Re^n$ and $y \in \Re^m$. Consider also a method of the form

$$x^{k+1} = x^k - \alpha^k \nabla_x f\big(x^k, y^k\big),$$

where y^k is a solution of $h(x^k, y) = 0$, viewed as a system of m equations in the unknown vector y, and α^k is chosen by the minimization rule or the Armijo rule. Formulate conditions that guarantee that this method is convergent.

1.3 GRADIENT METHODS – RATE OF CONVERGENCE

The second major issue regarding gradient methods relates to the rate (or speed) of convergence of the generated sequences $\{x^k\}$. The mere fact that $\{x^k\}$ converges to a stationary point x^* will be of little practical value unless the points x^k are reasonably close to x^* after relatively few iterations. Thus, the study of the rate of convergence provides what are often the dominant criteria for selecting one algorithm in favor of others for solving a particular problem.

Approaches for Rate of Convergence Analysis

There are several approaches towards quantifying the rate of convergence of nonlinear programming algorithms. We will discuss briefly three possibilities and then concentrate on the third.

(a) *Computational complexity approach*: Here we try to estimate the number of elementary operations needed by a given method to find