
Suppose \( f : \mathbb{R}^n \to \mathbb{R} \) is a given \( G \)-differentiable function. Let \( A \) be a \( m \times n \) matrix with \( m < n \), rank \( A \) is \( m \) and \( b \in \mathbb{R}^m \) be in the range of \( A \). Let \( K \) be the set of all solutions of the equation

\[
Ax = b \quad \text{(LEq)}
\]

This is an underdetermined system of linear equations.

Let \( a^{(j)} \) be the \( j \)-th row vector of \( A \), so that 

\[
A^T := [a^{(1)}, a^{(2)}, \ldots, a^{(m)}].
\]

Then (LEq) is equivalent to the requirement that

\[
\langle a^{(j)}, x \rangle = b_j \quad \text{for} \quad 1 \leq j \leq m.
\]

This is a system of \( m \) linear equality constraints. The rank condition implies that the vectors \( a^{(j)} \) are linearly independent and the null space of \( A \) has dimension \( n - m \). Let \( \tilde{x} \) be a solution of (LEq), then \( K = \{ \tilde{x} + z : z \in N(A) \} \) is an \((n-m)\)-dimensional affine subspace of \( \mathbb{R}^n \). Consider the problem of minimizing \( f \) on \( K \) and finding

\[
\alpha(f, K) := \inf_{x \in K} f(x) \quad \text{(19.1)}
\]

The basic results about this minimization problem are that

(i) If \( f \) is weakly coercive on \( \mathbb{R}^n \), then \( \alpha(f, K) \) is finite and there is an \( \hat{x} \in K \) which minimizes \( f \) on \( K \).

(ii) If \( f \) is (strictly) convex on \( \mathbb{R}^n \) and there is a minimizer of \( f \) on \( K \), then the set of all minimizers is convex (a singleton).

These results follow from earlier theorems in this course.

Theorem 20.1 provides the extremality conditions that hold at a local minimizer of \( f \) on \( K \) and leads to the following.

**Theorem 19.1.** Suppose \( f, K \) as above and \( \tilde{x} \) is a local minimizer of \( f \) on \( K \), then \( \tilde{x} \) satisfies

\[
\langle \nabla f(x), z \rangle = 0 \quad \text{for all} \quad z \in N(A). \quad \text{(19.2)}
\]

Moreover if \( f \) is convex on \( K \) and \( \tilde{x} \) satisfies (19.2), then \( \tilde{x} \) minimizes \( f \) on \( K \). If \( \hat{x} \) is a local maximizer of \( f \) on \( K \), the \( \hat{x} \) also satisfies (19.2).

**Proof.** If \( \tilde{x} \) is a local minimizer, then it satisfies (20.1) for all \( y \in K \). Thus (19.2) holds as \( \tilde{x} \pm z \in K \) for all \( z \in N(A) \). The next sentence follows from corollary 20.2 applied to this case. When \( \hat{x} \) is a local maximizer the signs in (20.1) must be reversed but the above argument still yields (19.2).
This result says that, when m linear equality constraints are imposed, the gradient of f at a constrained local minimizer, or maximizer, must satisfy (n-m) orthogonality conditions. A point \( \tilde{x} \in K \) obeying (19.2) is called a constrained critical point of f on K. A constrained critical point need not be a local minimizer or maximizer; it could be a saddle point of f on K.

Quite often this orthogonality condition is stated more explicitly as follows. The proof depends on the fundamental theorem of linear algebra which says that when \( A \) is an \( m \times n \) matrix as above then \( N(A) \oplus R(A^T) = \mathbb{R}^n \). See Strang [5], Chapter 2 section 4.

**Theorem 19.2. (Linear Lagrange multiplier rule)** Suppose \( f, K \) as above and \( \tilde{x} \) is a local minimizer of f on K, then there is a \( \lambda \in \mathbb{R}^m \) such that

\[
\nabla f(\tilde{x}) = A^T \lambda = \sum_{j=1}^{m} \lambda_j a^{(j)} \tag{19.3}
\]

holds. When f is convex on K and \( \tilde{x} \) satisfies (19.3), then \( \tilde{x} \) minimizes f on K.

**Proof.** Theorem 19.1 says that if \( \tilde{x} \) is a local minimizer of f on K then \( \nabla f(\tilde{x}) \) is orthogonal to \( N(A) \). Thus there is a \( \lambda \in \mathbb{R}^m \) such that \( \nabla f(\tilde{x}) = A^T \lambda \). Thus (19.3) holds and the second part holds just as in the previous theorem. \( \square \)

It is worth noting that existence of solutions to the system (LEq) and (19.3) is completely equivalent to requiring that (LEq) and (19.2) hold. This multiplier rule is a system of (n+m) equations for the (n+m) unknowns \((x, \lambda)\). The system is linear in \( \lambda \) but, equation (19.3) is linear in \( x \) only when \( f \) is a quadratic function.

It is often useful to convert this into an unconstrained optimization problem on \( N(A) \). Let \( \tilde{x} \) be a solution of (LEq) and define \( g : N(A) \to \mathbb{R} \) by

\[
g(z) := f(\tilde{x} + z) \tag{19.4}
\]

This has

\[
\alpha(f, K) = \alpha(g, N(A)) := \inf_{z \in N(A)} g(z).
\]

In practice to use this approach, a basis of \( N(A) \) must first be found - this is an algebraic problem.

**Exercise 19.1** Take \( n = 2 \), \( f(x) := x_1^2 + 2x_2^2 \) and minimize f subject to the constraint \( 2x_1 + x_2 = 2 \). Do this both by elimination and by the Lagrange multiplier rule and verify that you obtain the same minimizer and value for this problem.

### 19.1 Least Norm Solutions of Linear Equations

An important example of minimization on an affine subspace is the problem of minimizing \( f(x) := \|x\|^2 \) on the set K defined by (LEq). If \( x_{LN} \) minimizes f on K, then it will be the solution of (LEq) of least 2-norm - or the point on the affine subspace K that is closest to the origin in the Euclidean metric on \( \mathbb{R}^n \).
The G-derivative of \( f \) is \( \nabla f(x) = 2x \), so \( x_{LN} \) minimizes \( f \) on \( K \) if and only if it is in \( K \) and satisfies \( 2x = A^T \lambda \) for some \( \lambda \in \mathbb{R}^m \). Substitute in (LEq) then \( \lambda \) satisfies
\[
A A^T \lambda = 2b
\] (19.5)
This equation involves a symmetric positive definite matrix and can be solved in a number of different ways. Since \( A^T \) has rank \( m \), equation (19.5) has a unique solution \( \hat{\lambda} \) and then \( x_{LN} := A^T \hat{\lambda} \) is the least norm solution of (LEq). Moreover, from (19.2), \( x_{LN} \) is orthogonal to \( N(A) \), so we can write
\[
K = \{ x_{LN} \} \oplus N(A) \tag{19.6}
\]
where \( \oplus \) implies that this is an orthogonal sum.

A simple way to find \( \hat{\lambda} \) is to minimize the function \( \mathcal{E} \) on \( \mathbb{R}^m \) defined by
\[
\mathcal{E}(\lambda) := \| A^T \lambda \|^2 - 4\langle b, \lambda \rangle \tag{19.7}
\]
This is an \( m \)-dimensional quadratic, unconstrained, convex minimization problem of the type treated in section 14. Thus we find that this problem may be reduced to a (lower dimensional) unconstrained quadratic optimization problem for the Lagrange multiplier.

Example 19.1 Find the point on the hyperplane \( \sum_{j=1}^{n} x_j = 1 \) closest to the origin. That is find the point on this hyperplane that minimizes \( d(x) := \| x \|^2 \). This problem has only one linear constraint equation in \( n \) variables.

From theorem 19.2, there is a real number \( \lambda \) such that the minimizer satisfies
\[
\nabla d(x) = 2x = \lambda(1,1,\ldots,1).
\]
Summing over the entries yields \( n \lambda = 2 \) so the minimizer is \( \hat{x} = n^{-1} (1,1,\ldots,1) \) and \( d(\hat{x}) = 1/\sqrt{n} \). This point is the barycenter of the set of probability vectors \( \Delta'_n \).

19.2. Optimal Portfolio Problems. Optimal portfolio (or asset allocation) problems treat the subject of how to invest a fixed initial capital \( M \) in \( n \) investments (assets or asset classes) where
(a) we know the expected rate of return \( r_j \) on each investment, and
(b) we want to have an overall rate of return of \( R \) on the whole portfolio.

The theory is that a ”rational investor” should choose to distribute his investments in a manner that minimizes the overall variance of the portfolio. Thus we are given the expected rate of return and variance of a particular investment and the covariances of all the different investments. That is you would like to ”most stably” receive the specified return.

Mathematically this can be formulated as follows.

Find \( x \in \mathbb{R}^n \) that minimizes
\[
V(x) := \langle Cx, x \rangle \tag{19.8}
\]
subject to
\[
\sum_{j=1}^{n} x_j = M \quad \text{and} \quad \sum_{j=1}^{n} r_j x_j = R. \tag{19.9}
\]

Here \( x_j \) is the amount invested in the \( j \)-th asset, \( M \) is the initial amount invested and \( R \) is the expected return - both are assumed to be strictly positive. \( C := (c_{jk}) \) is a positive definite symmetric \( n \times n \) matrix which is the variance-covariance matrix of the asset prices.
This is a problem with two linear equality constraints, so the theory of section 13 applies. This function \( V(x) \) is strictly convex from theorem 7.3 and it is also coercive, so \( V \) will attain its infimum on the affine subspace of all vectors which satisfy (18.2)-(18.3).

Let \( e := (1,1,\ldots,1)^T \) and \( r := (r_1,\ldots,r_n)^T \) be the vectors in the constraints. Without loss of generality (wlog), assume that \( e, r \) are linearly independent. (If not then each asset has the same rate of return - so they are indistinguishable from the point of view of return and you just put all your resources into the security with least variance). Henceforth we will also assume that all the \( r_j > 0 \). Why invest in a security where the expected return is negative or zero? Also note that if \( \hat{x} \) is the solution of this problem with \( A = 1, R = R_1 \), then \( Ax \) is the solution of this problem subject to (18.2) and \( R = AR_1 \). So wlog we often take \( A = 1 \) and then adjust \( R \). This is equivalent to just determining the proportion that should be invested in each security.

There are a number of ways to minimize \( V \) subject to these constraints. First we could use the two constraints to eliminate two variables - say \( x_n, x_{n+1} \) and then substitute back into \( V(x) \). This will be a quadratic function of \( n-2 \) variables and one seeks an unconstrained minimizer of this function. This is fine when \( n \) is small and you do not change the vector \( r \).

There are many direct minimization algorithms that are specially designed to find minima of problems such as this. Here the equations and inequalities satisfied at the optimal allocation will be found.

From theorem 19.2, \( \hat{x} \) will be a minimizer of this problem if and only if there is a \( \lambda \in \mathbb{R}^2 \) that satisfies

\[
Cx = \lambda_1 e + \lambda_2 r
\]

(19.10)

\( C \) is invertible as it is positive definite, with \( C^{-1} \) also being symmetric and positive definite. Thus this system has the unique solution

\[
\hat{x} = C^{-1} (\lambda_1 e + \lambda_2 r)
\]

(19.11)

Substitute this back into (19.9)), to obtain 2 linear equations for \( \lambda \). Namely

\[
\langle C^{-1}e, e \rangle \lambda_1 + \langle C^{-1}r, e \rangle \lambda_2 = M
\]

(19.12)

\[
\langle C^{-1}e, r \rangle \lambda_1 + \langle C^{-1}r, r \rangle \lambda_2 = R
\]

(19.13)

When \( e, r \) are linearly independent this system is non-singular and there is a unique solution for \( \lambda_1, \lambda_2 \). Substitute these values into (19.11), to determine the optimal allocation. That is, to find the optimal allocation, we just need the data, the inverse \( C^{-1} \) of the covariance matrix and be able to solve a 2 by 2 symmetric matrix equation for the two Lagrange multipliers.

In practice most investors, or funds, impose further constraints. Many investors are uncomfortable with, or not allowed to, short sell stocks. This means that all the \( x_j \) are required to be positive. Then the allowable allocations are

\[
S := \{ x \in \mathbb{R}_+^n : \text{ and (19.9) holds} \}
\]

(19.14)

This is a bounded closed convex set, with

\[
0 \leq x_j \leq \min (A, R/r_j) \quad \text{for each } j.
\]
provided the vector \( r > 0 \). (What happens when \( r_j \leq 0 \)?) The optimization problem now is to minimize a strictly convex function on a compact convex set, so there is a unique minimizer of \( V \) on \( S \).

The analysis of some of the next sections will be needed to find the conditions obeyed at a local minimizer when this positivity condition is imposed. Other constraints are often imposed also - see section.

Exercises.

Exercise 19.2: Suppose that a line in space is defined by two linear equations
\[
\langle a^{(j)}, x \rangle = 1 \quad \text{with} \quad a^{(1)} = (3, 4, 0) \quad \text{and} \quad a^{(2)} = (2, 1, 2)
\]
Find the point on this line closest to the point \((0, 0, 1)\) by minimizing the Euclidean distance to the point.

Exercise 19.3: Find the equations obeyed by the point on the line in the previous problem that is closest to \((0, 0, 1)\) in the 4-norm.

20. Optimization on Closed Convex Sets

Many important examples of constrained optimization problems are, or may be, posed as problems that involved the minimization of a function on a proper closed convex subset of \( \mathbb{R}^n \). Such problems are called constrained optimization problems. In the next few sections, we shall derive the extremality conditions that must hold at local minimizers, and maximizers, of such problems and obtain the specific results for a number of important classes of problems.

Even when \( n = 1 \) and the domain is an interval the conditions that must hold when the local minimum or maximum occurs at an end-point of the interval involves a number of alternatives. Straightforward analysis shows that constrained local minima generally obey inequalities - not equations. Moreover the the criteria are quite different to those that hold at local minimizers of differentiable convex functions on open convex sets as found in theorem 14.1.

**Theorem 20.1.** Suppose \( C \) is a closed convex subset of \( \mathbb{R}^n \) and \( \bar{x} \) is a local minimizer of \( f \) on \( C \). If \( f \) is \( G \)-differentiable at \( \bar{x} \) then
\[
\langle \nabla f(\bar{x}), y - \bar{x} \rangle \geq 0 \quad \text{for all} \ y \in C.
\] (20.1)

**Proof.** Choose \( y \in C \) and let \( z(t) := (1-t)\bar{x} + y \). Then \( f(z(t)) \geq f(\bar{x}) \) for \( 0 \leq t < \delta \) as \( \bar{x} \) is a local minimizer of \( f \) on \( C \). Thus
\[
t^{-1}[f(x + t(y - x)) - f(x)] \geq 0 \quad \text{for} \quad 0 < t < \delta.
\]
Take the limit of this as \( t \to 0^+ \), then (20.1) holds. \( \square \)

(20.1) is called a variational inequality. It is a necessary condition for \( \bar{x} \) to be a local minimizer of \( f \) on \( C \) - and should be compared with the criterion in theorem 14.1 for the unconstrained case.
Corollary 20.2. Suppose $C$ is a closed convex set in $\mathbb{R}^n$, $f : C \to \mathbb{R}$ is convex and there is a point $\tilde{x} \in C$ where (20.1) holds. Then $\tilde{x}$ minimizes $f$ on $C$.

Proof. If $f$ is convex on $C$, and it is differentiable at a point $\tilde{x}$, then just as in the proof of theorem 15.1, we have

$$f(y) \geq f(\tilde{x}) + \langle \nabla f(\tilde{x}), y - \tilde{x} \rangle \quad \text{for all } y \in C$$

If (20.1) holds then $\tilde{x}$ minimizes $f$ on $C$ as claimed. □

The following result is proved by reversing all the inequalities in the proof of theorem 20.1. It implies that for convex constrained optimization, the (first derivative) criterion for a point to be a local maximizer differs from that for a local minimizer.

Corollary 20.3. Suppose $C$ is a closed convex set in $\mathbb{R}^n$, and $\hat{x}$ is a local maximizer of $f$ on $C$. If $f$ is $G$-differentiable at $\hat{x}$ then

$$\langle \nabla f(\hat{x}), y - \hat{x} \rangle \leq 0 \quad \text{for all } y \in C. \quad (20.2)$$

Unfortunately the functions we want to minimize often are not defined on open subsets of $\mathbb{R}^n$, so a generalization of the concept of the $G$-derivative of a function must be introduced to derive the extremality conditions that hold at minimizers of a function.

Let $C$ be a closed convex set in $\mathbb{R}^n$ and $f : C \to \mathbb{R}$ be a given function on $C$. A vector $v \in \mathbb{R}^n$ is said to be a sub-gradient of $f$ at a point $x(0) \in C$, provided

$$f(x) \geq f(x(0)) + \langle v, x - x(0) \rangle \quad \text{for all } x \in C. \quad (20.3)$$

Note that this is a "global" inequality that must hold for all $x$ in C. The definition does not require any continuity of $f$. You have seen that when $f$ is convex and $G$-differentiable at $x(0)$, then the $G$-derivative $\nabla f(x(0))$ is a subgradient of $f$ at $x(0)$.

The set of all subgradients of a function $f$ at $x(0)$ is called the subdifferential of $f$ at $x(0)$ and is denoted by $\partial f(x(0))$. It is straightforward to verify that $\partial f(x(0))$ is a closed convex subset of $\mathbb{R}^n$.

Example 20.1 The function $f : \mathbb{R} \to [0, \infty)$ defined by $f(x) := |x|$ is not $G$-differentiable at 0 but any number $v \in [-1, 1]$ is a subgradient of $f$ at 0. This may be generalized to the function $f : \mathbb{R}^n \to [0, \infty)$ defined by $f(x) := \|x\|_1$. When $n = 2$, this function is $G$-differentiable at points off the coordinate axes, but not at points on the $x_1, x_2$ axes. Find the possible subgradients there - including at the origin.

Berkowitz chapter 3, theorem 2.1 proves that if $f$ is convex on $C$ and $x(0)$ is an interior point of $C$, then $\partial f(x(0))$ is non-empty. The proof uses "separation theorems " for convex sets in $\mathbb{R}^n$.

Theorem 15.1 shows that if $C$ is an open convex set in $\mathbb{R}^n$ and $f$ is a convex $G$-differentiable function on $C$, then $\nabla f(x)$ is a subgradient of $f$ at each point in $C$. In fact, Berkowitz, chapter 3, theorem 3.1 proves that if $x(0)$ is an interior point of $C$ and $f$ is convex, then $f$ is $G$-differentiable at $x(0)$ if and only if $\partial f(x(0))$ is a singleton and then that singleton is $\nabla f(x(0))$. 
The concept of subgradient allows us to state the following necessary and sufficient condition for the existence of a minimizer of \( f \) on \( C \).

**Theorem 20.4.** Suppose \( C \) is a closed convex set in \( \mathbb{R}^n \) and \( f : C \to \mathbb{R} \) is a given function. Then \( \hat{x} \in C \) minimizes \( f \) on \( C \) if and only if \( 0 \) is a subgradient of \( f \) on \( C \) at \( \hat{x} \).

**Proof.** Suppose \( 0 \) is a subgradient for \( f \) on \( C \) at \( \hat{x} \). Substitute in (20.3), then \( f(x) \geq f(\hat{x}) \) for all \( x \in C \). Conversely if \( \hat{x} \) minimizes \( f \) on \( C \), then \( v = 0 \) is a subgradient of \( f \) at \( \hat{x} \). \( \square \)

It is worth noting that no continuity of \( f \) is required for this result.

### 21. Closed Convex Sets and Cones

The above analysis involved arbitrary closed convex sets. In practice we need to work with closed convex sets that are defined explicitly by systems of equations and inequalities. Here some specific such systems will be described.

The simplest case is a system of linear inequalities. Suppose that \( A \) is an \( m \times n \) matrix with \( a^{(j)} \) being the \( j \)-th row vector of \( A \), so that \( A^T := [a^{(1)}, a^{(2)}, \ldots, a^{(m)}] \). Given \( b \in \mathbb{R}^m \), consider the system of inequalities

\[
\langle a^{(j)}, x \rangle \leq b_j \quad \text{for} \quad 1 \leq j \leq m. \tag{21.1}
\]

Here \( m \) may be either small or much larger than \( n \). This defines a closed convex subset in \( \mathbb{R}^n \) as it is the intersection of a finite number of closed half-spaces. Note that linear equalities are included in this formulation as an equality such as \( \langle c, x \rangle = b_0 \) is equivalent to requiring the two linear inequalities \( \langle c, x \rangle \leq b_0 \) and \( -\langle c, x \rangle \leq b_0 \).

Often this system is written as

\[
Ax \leq b. \tag{LIneq}
\]

The inequality here holds componentwise; we will often write \( x \leq y \) with \( x, y \in \mathbb{R}^m \) provided \( y_j - x_j \geq 0 \) for each \( 1 \leq j \leq m \).

A nonempty set of this type is called a **polytope**. A bounded polytope is called a **polyhedron**. A nontrivial subspace in \( \mathbb{R}^n \), or the positive orthant, are polytopes. The unit simplex and the set of all probability vectors in \( \mathbb{R}^n \) are polyhedra. See section 11 for the specific equalities and inequalities that specify these sets.

A box \( C \) in \( \mathbb{R}^n \) is defined by the \( 2n \) inequalities

\[
c_j \leq x_j \leq d_j \quad \text{for} \quad 1 \leq j \leq n.
\]

More generally when \( \{g_j : 1 \leq j \leq J \} \) is a family of l.s.c convex functions defined on \( \mathbb{R}^n \) and \( c \in \mathbb{R}^J \), then the set of all solutions of the family of inequalities

\[
g_j(x) \leq c_j \quad \text{for} \quad 1 \leq j \leq J \tag{21.2}
\]

is a closed convex set in \( \mathbb{R}^n \). Suppose

\[
C := \{ \ x \in \mathbb{R}^n \ : \ (21.2) \ \text{holds} \} \tag{21.3}
\]
is a non-empty convex set. For a \( \bar{x} \in C \) we say that the j-th inequality in (21.2) is \textit{active} (or is an active constraint) at \( \bar{x} \) provided \( g_j(\bar{x}) = c_j \). Otherwise this is an \textit{inactive} constraint or inequality. The set of indices of the active constraints for \( C \) at \( \bar{x} \) is denoted \( J(\bar{x}) \).

The set \( C \) is called the \textit{feasible set} associated with the inequalities (21.2). Points in the complement \( \mathbb{R}^n \setminus C \) are called infeasible.

The description of the extremality conditions obeyed by solutions of optimization problems on convex sets requires the theory of convex cones in \( \mathbb{R}^n \). Recall that a convex subset \( K \) of \( \mathbb{R}^n \) is said to be a \textit{convex cone} provided that \( cx \in K \) whenever \( x \in K \) and \( c \geq 0 \).

The set \( \{0\} \) is a cone - the trivial cone. Any non-trivial cone is an unbounded convex subset of \( \mathbb{R}^n \). The adjective "convex" will often be omitted; all cones here are assumed to be convex. The standard positive closed cone in \( \mathbb{R}^n \) is \( \mathbb{R}_+ := [0, \infty)^n \). That is \( x \in \mathbb{R}_+ \) iff \( x \geq 0 \) in

When \( K \) is a cone in \( \mathbb{R}^n \), the set \( V := K \cap (-K) \) is a subspace of \( K \) - and is the maximal subspace contained in \( K \). A cone is said to be \textit{strict} if \( K \cap (-K) = \{0\} \). The closed positive orthant \( \mathbb{R}^n_+ := [0, \infty)^n \) is a strict cone.

When \( K \) is a cone, so is \( cK \) for any real number \( c \). Note that \( cK = K \) for \( c > 0 \), and \( cK = -K \) for \( c < 0 \). When \( K_1, K_2 \) are cones, so also are \( K_1 \cap K_2 \) and \( K_1 + K_2 \). The intersection of any family of convex cones is again a convex cone.

Let \( S := \{d^{(1)}, d^{(2)}, \ldots, d^{(m)}\} \) be a finite subset of \( \mathbb{R}^n \). A vector \( x \) is said to be a \textit{positive linear combination (p.l.c.)} of the elements of \( S \) provided

\[
x = \sum_{j=1}^{m} \mu_j d^{(j)} \quad \text{with all the } \mu_j \geq 0.
\]

This is a \textit{strictly positive linear combination (s.p.l.c.)} if all the \( \mu_j > 0 \).

When \( S \) is any nonempty subset of \( \mathbb{R}^n \), define \( K(S) \) to be the smallest convex cone that contains \( S \). It is always well-defined, non-empty and contains the set of all strictly positive linear combinations of finite subsets of \( S \). The closure of \( K(S) \) is called the closed convex cone generated by \( S \).

A convex cone is said to be \textit{polyhedral} if it is the convex cone generated by a finite set. That is there is a finite set of vectors such that every vector in \( K \) has the form (21.4). When \( n \geq 3 \), there are cones which are not polyhedral.

\textbf{Exercises.}

Exercise 21.1: Write the system of inequalities for a box in the forms (LIneq) in a way that involves a simple matrix \( A \).

Exercise 21.2: Give an example of a cone in \( \mathbb{R}^3 \) that is not polyhedral.

Exercise 21.3: Verify that \( K_1 + K_2 \) is a convex cone when \( K_1, K_2 \) are. Suppose \( K \) is the polyhedral cone generated by a finite set \( S \) as above. What is the convex cone \( K + (-K) \)?
22. Tangent and Normal Cones of a Convex Set

Let $C$ be a convex set in $\mathbb{R}^n$ which contains at least 2 points. For each $x \in C$, define the tangent cone $T_x(C)$ of $C$ at $x$ to be the closure of the convex cone generated by $C - \{x\}$. The normal cone $N_x(C)$ of $C$ at $x$ to be the set of all vectors $d \in \mathbb{R}^n$ which satisfy

$$\langle d, y - x \rangle \leq 0 \quad \text{for all } y \in C.$$  

Both of these are closed convex cones.

Example 22.1 Let $\Delta_2$ be the unit simplex in $\mathbb{R}^2$. It is the triangle whose vertices are $(0,0), (1,0), (1,0)$. Equivalently it is the set in $\mathbb{R}^2$ defined by the three inequalities:

$$x_1 \geq 0, \quad x_2 \geq 0, \quad \text{and} \quad x_1 + x_2 \leq 1.$$ 

If $x$ is an interior point of the triangle, then $T_x(C) = \mathbb{R}^2$ and the normal cone is trivial. At the origin the tangent cone is $T_0(C) = \mathbb{R}^2$ and the normal cone is $-\mathbb{R}^2$. At the other two vertices of the triangle the tangent cone is the cone of all directions that point into the triangle. (Write down the algebraic formulae for these sets.)

Example 22.2 Let $B_1$ be the Euclidean unit ball in $\mathbb{R}^n$. It is defined by the single convex inequality:

$$d(x) := \sum_{j=1}^{n} x_j^2 \leq 1$$

When $|x| < 1$, then the tangent cone $T_x(B_1) = \mathbb{R}^n$ and the normal cone is $\{0\}$. When $|x| = 1$, then the tangent cone $T_x(B_1)$ is a half-space and the normal cone is $\{tx : t \geq 0\}$.

Example 22.3 Let $g : \mathbb{R}^n \rightarrow \mathbb{R}$ be a $C^1$-quasi-convex function and $C$ be the closed convex set in $\mathbb{R}^n$ defined by the single convex inequality:

$$g(x) \leq c$$

Assume that the interior of $C$ is non-empty and that the boundary $\partial C$ of $C$ is the level set $L_c(g)$. It is straightforward to verify that when $x \in \text{int}(C)$, the tangent cones $T_x(C) = \mathbb{R}^n$ and the normal cones is $N_x(C) = \{0\}$. When $x \in \partial C$ then the normal cone is $N_x(C) = \{t\nabla g(x) : t \geq 0\}$ provided $\nabla g(x) \neq 0$.

Suppose $C$ is a nonempty closed convex subset of $\mathbb{R}^n$ defined by the system of linear inequalities (LIneq). If $\bar{x} \in C$, then a vector $z$ is in $T_{\bar{x}}(C)$ if and only if it satisfies

$$\langle a^{(j)}, z \rangle \leq 0 \quad \text{for each } j \in J(\bar{x}).$$  

(22.2)

It often requires a lot of algebra to determine this tangent cone from the specification of $C$ via the inequalities (LIneq). However the following provides a simple description of the normal cone to $C$ at $\bar{x}$.

**Theorem 22.1.** Suppose $C$ is defined by the system of linear inequalities (LIneq) and $x$ is in $C$. Then the normal cone $N_x(C)$ of $C$ at $x$ is the closed convex cone generated by the active constraints at $x$.

The proof of this is purely algebraic and quite straightforward - assuming you have worked with convex sets, hyperplanes and linear inequalities. The proof uses "separation"
results and may be found in most texts that treat the theory of linear inequalities and will not be given here. It says that a vector $z$ is in $N_x(C)$ if and only if it is a p.l.c. of the vectors \{$a^{(j)}: j \in J(x)$\}, or that

\[
z = \sum_{j \in J(x)} \mu_j a^{(j)} \quad \text{with all the } \mu_j \geq 0. \quad (22.3)
\]

When we know the representation of the normal cone at a local minimizer (or maximizer) of a function $f$ on $C$, then the variational inequality of theorem 20.1 takes quite a different form. Use of this definition of the normal cone leads to the following restatement of theorem 20.1.

**Theorem 22.2.** Suppose $C$ is a closed convex subset of $\mathbb{R}^n$ and $\tilde{x}$ is a local minimizer of $f$ on $C$. If $f$ is $G$-differentiable at $\tilde{x}$ then

\[
\nabla f(\tilde{x}) + z = 0 \quad \text{for some } z \in N_x(C). \quad (22.4)
\]

When $\tilde{x}$ is a local maximizer, then

\[
\nabla f(\tilde{x}) = z \quad \text{for some } z \in N_x(C). \quad (22.5)
\]

This general result shows that the systems of equations that hold at local minima and maxima can be described explicitly when we have characterizations of the normal cones. The next corollaries provide examples for two important cases.

**Corollary 22.3.** (linear KKT) Suppose $C$ is the closed convex subset of $\mathbb{R}^n$ defined by $\text{(LIneq)}$ and $\tilde{x}$ is a local minimizer of $f$ on $C$. If $f$ is $G$-differentiable at $\tilde{x}$ then

\[
-\nabla f(\tilde{x}) = \sum_{j \in J(\tilde{x})} \mu_j a^{(j)} \quad \text{where each } \mu_j \geq 0. \quad (22.6)
\]

*Proof.* If $\tilde{x}$ is a local minimizer, then from (22.4), $-\nabla f(\tilde{x}) \in N_x(C)$, so the characterization of the normal cone in (22.3) yields (22.6). \qed

The coefficients $\mu_j$ are called the Karush-Kuhn-Tucker (KKT), or inequality, multipliers associated with the active constraints at $\tilde{x}$. Quite often the condition (22.6) for a minimizer of $f$ on $C$ is stated in the form that there is a $\mu \in \mathbb{R}^m_+$, such that $\tilde{x}$ satisfies

\[
\nabla f(x) + \sum_{j=1}^{m} \mu_j a^{(j)} = 0 \quad (22.7)
\]

with $\mu_j = 0$ when the j-th constraint is inactive at $\tilde{x}$. This shows that different conditions hold at a local minimizer of $f$ on $C$ depending on which constraints are ”active” at the point. Inactive constraints are not ”observed” by, or involved in, the extremality equations.

When $J$ inequality constraints are active at a minimizer of this problem, the extremality conditions require $n + J$ equality conditions (from (22.6) and the $J$ active constraints) and $m - J$ inequality conditions (from the inactive linear constraints). There are $J + n$ unknowns $(x, \mu)$. In general one does not try to ”solve” this system directly - but these equations are used as ”stopping criteria” for algorithms for minimizing $f$ on $C$. 
Corollary 22.4. Suppose $f, g$ are $G$-differentiable real valued functions on $\mathbb{R}^n$ with $g$ convex. Define $C$ as in exercise 19.1, assume it has nonempty interior and the boundary $\partial C$ is the level set $L_c(g)$. If \( \tilde{x} \) is a local minimizer of $f$ on $C$ and $\nabla g(\tilde{x}) \neq 0$, then
\[
\nabla f(\tilde{x}) + \mu \nabla g(\tilde{x}) = 0 \quad \text{for some } \mu \geq 0, \quad \text{and} \quad (22.8)
\]
\[
\mu(g(\tilde{x}) - c) = 0. \quad (22.9)
\]
When \( \hat{x} \) is a local maximizer of $f$ on $C$, these again hold save now $\mu \leq 0$.

Proof. Under these assumptions, theorem 22.2 says that \( \tilde{x} \) satisfies (22.4). $g(\tilde{x}) < c$, then (22.8) holds with $\mu = 0$. If $g(\tilde{x}) = c$ then, by assumption, $\tilde{x} \in \partial C$ and example 22.3 gives an explicit characterization of the normal cone which yields (22.8) and (22.9). The proof for a local maximum is similar. \( \square \)

Note that when $\tilde{x} \in \text{int}(C)$, this is just the condition for an unconstrained minimizer of $f$ on $C$. When $\tilde{x} \in \partial C$, then there is an extra unknown ($\mu$) and an extra equation - so (22.4) is now a system of $n + 1$ equations for $n + 1$ unknowns.

Exercises.

Exercise 22.1. Find the normal and tangent cones of points on the axes of the positive quadrant $\mathbb{R}^2_+$ in the plane. Suppose that a differentiable function $f$ is minimized on $\mathbb{R}^2_+$ at the point $(a, 0)$. What conditions are obeyed at this minimizer when $a > 0$? What conditions hold if $a = 0$?

Exercise 22.2. Find the normal and tangent cones of points on the planar faces of the positive octant $\mathbb{R}^3_+$ in space.

Exercise 22.3. Suppose $f$ is a $G$-differential function on $\mathbb{R}^n_+$ that has a local minimizer at a point $\tilde{x}$. Prove that
\[
\nabla f(\tilde{x}) \geq 0 \quad \text{and} \quad \langle \tilde{x}, \nabla f(\tilde{x}) \rangle = 0. \quad (22.10)
\]

23. Programming Problems

The important examples of optimization problems for many economics, finance and business applications are often linear or quadratic programming problems.

A linear programming problem is the problem of minimizing (or maximizing) a linear function on a closed convex set defined by a finite number of linear inequalities. Typically $f(x) := \langle c, x \rangle$ and we want to minimize $f$ on a set $C$ defined by (Lineq). When $C$ is nonempty and bounded, this problem has solutions and a finite value.

A (convex) quadratic programming problem is the problem of minimizing a quadratic function
\[
f(x) := \langle Dx, x \rangle + \langle d, x \rangle \quad (23.1)
\]
on a nonempty closed convex subset $C$ of $\mathbb{R}^n$ defined by a system of equations of the form (Lineq). Here $D$ is assumed to be a p.s.d. $n \times n$ matrix. Again, when $C$ is non-empty and
bounded then \( f \) has a finite minimum and maximum value on \( C \), and the equations obeyed at the minimizers will be (from (22.7))

\[
Dx + d + \sum_{j=1}^{m} \mu_j a^{(j)} = 0
\]

(23.2)

for some \( \mu \in \mathbb{R}_+^m \) that satisfies the complementarity condition

\[
\langle \mu, Ax - b \rangle = 0
\]

(23.3)

These constitute \( n + 1 \) equations for the \( n + m \) unknowns \((x, \mu)\) and the \( m \) inequalities (LIneq) must also hold.

### 24. Optimization for Eigenvalues and Eigenvectors

One of the important classes of constrained optimization problems has been the determination of eigenvalues and eigenvectors of real symmetric matrices. Let \( A \) be an \( n \times n \) real matrix. A number \( \lambda \in \mathbb{C} \) is said to be an eigenvalue of \( A \) provided there is a non-zero vector \( v \in \mathbb{C}^n \) such that

\[
A v = \lambda v
\]

(24.1)

An eigenvector is said to be normalized if \( \|v\|_2 = 1 \). The eigenvalue \( \lambda \) has (geometric) multiplicity \( m \) if there are \( m \) linearly independent eigenvectors corresponding to the eigenvalue \( \lambda \).

In general, real matrices need not have real eigenvalues. In your linear algebra classes, you should have learnt about the properties of the eigenvalues of real symmetric matrices. Strang [5] has a thorough description.

In particular when \( A \) is an \( n \times n \) real symmetric matrix,

(i) if \( \lambda \) is an eigenvalue of \( A \), then \( \lambda \) is real, and there is an associated real eigenvector.

(ii) \( A \) has at most \( n \) real eigenvalues. An elementary result that will be used is that \( \lambda \) is an eigenvalue of \( A \) if and only if \( \lambda + c \) is an eigenvalue of \( A + cI \).

When \( A \) is a real symmetric matrix, let the eigenvalues of \( A \) be denoted \( \sigma(A) = \{\lambda_j : 1 \leq j \leq J\} \) with \( \lambda_j < \lambda_{j+1} \) for all \( j \). Let the eigenvalue \( \lambda_j \) have multiplicity \( m_j \) then,

\[
\sum_{j=1}^{J} m_j = n
\]

That is there will be \( n \) linearly independent eigenvectors of a real symmetric matrix. Also when \( \lambda_j, \lambda_k \) are distinct eigenvalues of \( A \), \( v^{(j)}, v^{(k)} \) are corresponding eigenvectors, then \( \langle v^{(j)}, v^{(k)} \rangle = 0 \) or these eigenvectors are orthogonal.

The determination, and estimation, of eigenvalues of specific matrices has been the focus of a lot of research in recent times. The basic result is the following - which was first stated for related issues for partial differential equations by Lord Rayleigh.
Theorem 24.1. Let $A$ be a real symmetric matrix and $\lambda_1, \lambda_J$ be the smallest and largest eigenvalue of $A$. Then

$$
\lambda_1 = \inf_{x \in S_1} \langle Ax, x \rangle \quad \text{and} \quad \lambda_J = \sup_{x \in S_1} \langle Ax, x \rangle.
$$

(24.2)

Proof. Let $q(x) := \langle Ax, x \rangle$ as before. Then $q$ is continuous and has $\nabla q(x) = 2Ax$ when $A$ is a real symmetric matrix. The set $S_1$ is a compact subset of $\mathbb{R}^n$, so from Weierstrass' theorem $q$ attains its infimum and supremum on $S_1$.

Suppose (w.l.o.g.) that $\lambda_J > 0$ and $\hat{v}$ is a maximizer of $q$ on $S_1$. Then we also have that $\hat{v}$ maximizes $q$ on $B_1$ as if $x \in B_1$ with $x = rd$ and $d \in S_1$ then $q(x) = r^2q(d)$. $B_1$ is a convex set for which the hypotheses of corollary 22.4 hold. Thus the maximizers of $q$ on $B_1$ satisfy the equation

$$
2Ax = \mu x \quad \text{for some} \quad \mu > 0
$$

(24.3)

Substitute $\hat{v}$ here and take inner products with $\hat{v}$, then the definition of $\lambda_J$ implies,

$$
q(\hat{v}) = \frac{1}{2} \mu = \lambda_J
$$

Use this in (24.3), to see that $\hat{v}$ is an eigenvector of $A$ corresponding to the eigenvalue $\lambda_J$. There is a similar proof for $\lambda_1$. \qed

In particular this result shows that the symmetric matrix $A$ is positive definite if and only if $\lambda_1 > 0$ and then (24.2) shows that

$$
\langle Ax, x \rangle \geq \lambda_1 \|x\|_2^2 \quad \text{for all} \quad x \in \mathbb{R}^n
$$

A will be positive-semi-definite if and only if $\lambda_1 \geq 0$. That is, conditions for a symmetric matrix $A$ to be p.s.d or p.d. are often expressed in terms of its eigenvalues. The closed interval $[\lambda_1, \lambda_J]$ is called the numerical range of $A$ and this theorem shows that the set of all eigenvalues of $A$ obeys $\sigma(A) \subseteq [\lambda_1, \lambda_J]$.

Thus the second order conditions given in theorems 12.2 and 12.3 are often expressed in terms of the eigenvalues of the Hessian matrix of $f$ at $\hat{x}$ - rather than in terms of the quadratic forms used there. Namely the necessary condition (12.1) becomes that all the eigenvalues of $D^2f(\hat{x})$ are positive, and the sufficient condition (12.2) says that $\hat{x}$ will be an isolated strict local minimizer of $f$ provided all the eigenvalues of $D^2f(\hat{x})$ are strictly positive.

Similarly the criterion in theorem 13.6 for a function to be convex, or strictly convex, on an open convex set $C$ is often expressed in terms of the eigenvalues of the Hessian of $f$ on $C$.

25. Optimal Portfolio Allocation

Earlier the extremality conditions that hold at a minimizer of an asset allocation problem subject just to two linear equality constraints were found. When the allocations
are all required to be positive, that result must be generalized to include the effects of the inequality constraints. In this case there will be \( \lambda \in \mathbb{R}^2 \) such that the minimizer satisfies
\[
\sum_{k=1}^{n} c_{jk} x_k - \lambda_1 - \lambda_2 r_j \geq 0 \quad \text{for all } j \geq 1, \text{ and}
\]
\[
\sum_{k=1}^{n} c_{jk} x_k - \lambda_1 - \lambda_2 r_j = 0 \quad \text{whenever } x_j > 0.
\]

If each equation here is multiplied by \( x_j \) and added, then the minimizer \( \hat{x} \) will satisfy
\[
\langle Cx, x \rangle = \lambda_1 A + \lambda_2 R
\]
These conditions are not usually sufficient to determine the solution \( \hat{x} \) and \( \lambda \) directly. We have less than \( n + 2 \) equations for the \( n + 2 \) unknowns. (Q. Why have I included "usually" in these sentences?) Nevertheless solutions of this optimization problem may be found very efficiently using standard algorithms for minimizing convex functions on compact convex sets.

Another common requirement is to say that no single security can be more than a fixed percentage of the portfolio. This enforces diversification. In this case, define \( I := [0, b] \) where \( b \) is the maximum investment in a given security. Then the allowable portfolios lie in
\[
S := \{ x \in \mathbb{R}^n : (19.9) \text{ hold and each } x_j \in I \}
\]
This \( S \) is a compact convex set. For it to be non-empty, we must have \( nb \geq A \) and \( b(\sum_{j=1}^{n} r_j) \geq R \) - assuming \( r \geq 0 \). That is
\[
b > \min (A/n, R/|r|) \quad \text{where } |r| := \sum_{j=1}^{n} r_j
\]
In the preceding case this holds with \( b = A \). As \( b \) decreases, this set \( S \) becomes smaller and will be empty when \( b \) is too small. As long as \( S \neq \emptyset \), there will be a unique minimizer of \( V \) on \( S \) and the algorithms for minimizing convex functions on compact convex sets will find the minimizer.

This time the conditions satisfied by a local minimizer of \( V \) on \( S \) will be that there is a \( \lambda \in \mathbb{R}^2 \) such that the minimizer \( \hat{x} \) satisfies
\[
\sum_{k=1}^{n} c_{jk} x_k - \lambda_1 - \lambda_2 r_j \geq 0 \quad \text{whenever } x_j = 0,
\]
\[
\sum_{k=1}^{n} c_{jk} x_k - \lambda_1 - \lambda_2 r_j = 0 \quad \text{whenever } 0 < x_j < b,
\]
\[
\sum_{k=1}^{n} c_{jk} x_k - \lambda_1 - \lambda_2 r_j \leq 0 \quad \text{whenever } x_j = b,
\]
This is a system of \( n \) linear inequalities and 2 equations, for the \( n + 2 \) unknowns. Now we do not even have a simple analog of (25.3).
Still for each of these problems, we have

(i) the existence of a unique minimizer,
(ii) the criteria that must hold at a local minimizer of \( V \) on the set, and
(iii) good algorithms for finding this minimizer.

In each case the criteria that hold at the minimizer has a different form. It is

(i) a system of \( n + 2 \) linear equations for \( n + 2 \) unknowns when there are no inequality constraints,
(ii) a system of at least 3 equations and at most \( n - 1 \) inequalities when we require the components to be positive
(iii) a system of at least two equations and up to \( n \) inequalities when the components must lie in a fixed interval.

There are other variations on these problems. For example, one could allow the components \( x_j \) to lie in different intervals \( I_j \subset \mathbb{R} \), or the intervals could have the form \( I = [-b_1, b_2] \) where \( b_1, b_2 \) are both positive. For each such problem, the theory we have described allows us to conclude that there is a unique minimizer of \( V \) on \( S \) and to find the extremality conditions that such a minimizer must satisfy.

26. THE LAGRANGE MULTIPLIER RULE

In your first course on multivariable calculus, you probably learnt about the Lagrange multiplier rule. It provides the equations that holds at the local minimizers, or maximizers, of a differentiable function subject to an equality constraint. Generally no proofs are provided and, in fact, extra conditions must hold for the usual criterion to hold. Here the rule will be proved for a case of \( L \) equality constraints in \( \mathbb{R}^n \) subject to a "non-degeneracy" condition.

Let \( U \) be an open set in \( \mathbb{R}^n \) and \( f, h_1, \ldots, h_L \) be differentiable real valued functions defined on \( U \) with \( 1 \leq L < n \). Define \( H : U \to \mathbb{R}^L \) by

\[
H(x) := (h_1(x), \ldots, h_L(x))^T
\]

Given a vector \( b \in \mathbb{R}^L \), let

\[
S(b) := \{ x \in U : H(x) = b \}
\]  
(26.1)

That is, \( S(b) \) is the set of all solutions of \( L \) equalities in the \( n \) variables \( x_j \). We will assume that \( S(b) \) is a nonempty closed subset of both \( U \) and \( \mathbb{R}^n \). Note that

\[
DH(x) = (D_j h(x)) = [\nabla h_1(x), \ldots, \nabla h_L(x)]^T
\]

is a \( L \times n \) real matrix. A point \( x \in S(b) \) is said to be a regular point for (this representation of) the set \( S(b) \) provided the rank of \( DH(x) \) is \( L \). That is provided the vectors \( \{\nabla h_1(x), \ldots, \nabla h_L(x)\} \) are linearly independent.

Often we take \( b = 0 \) by changing the definition of \( H \). In this case \( S \) will be used for \( S(0) \). The Lagrange multiplier rule holds provided at a local minimizer (or maximizer) of \( f \) on \( S \) whenever the minimizer (or maximizer) is a regular point of \( S \). This may be stated as follows.
Theorem 26.1. (Lagrange multiplier rule) Suppose \( f, S(b) \) are continuously differentiable functions defined on an open set \( U \) in \( \mathbb{R}^n \). If \( \tilde{x} \) is a local minimizer of \( f \) on \( S(b) \) and is a regular point for \( S(b) \), then there is a unique vector \( \lambda \in \mathbb{R}^L \) such that \( \tilde{x} \) satisfies

\[
\nabla f(x) + \sum_{l=1}^{L} \lambda_l \nabla h_l(x) = 0 \tag{26.2}
\]

(26.1) and (26.2) constitute a system of \( n + L \) equations for \( n + L \) unknowns; namely the \( n \) entries in \( \tilde{x} \) and the \( L \) Lagrange multipliers \( \lambda_l \).

This result will be proved via a "penalty function" formulation - as is common in numerical analysis. Given the local minimizer \( \tilde{x} \) of \( f \) on \( S \) and an \( \epsilon > 0 \), consider the function

\[
F_k(x) := f(x) + \frac{k}{2} \sum_{l=1}^{L} h_l(x)^2 + \frac{\epsilon}{2} \|x - \tilde{x}\|^2_2 \tag{26.3}
\]

For \( \delta \) small enough, let \( B := \{ x \in U : \|x - \tilde{x}\|_2 \leq \delta \} \) be a closed ball of radius \( \delta \) and center \( \tilde{x} \).

Consider the penalized problem of minimizing \( F_k \) on \( B \). This is a problem of minimizing a continuously differentiable function on a closed ball in \( \mathbb{R}^n \); so the constraints are easy to work with. This minimization problem has a solution from Weierstrass’ theorem. Let \( x^{(k)} \) be a minimizer of this problem and \( \Gamma := \{ x^{(k)} : k \geq 1 \} \) be a corresponding sequence of minimizers.

Lemma 26.2. (Convergence) Suppose \( f, S, F_k, B \) as above and \( \tilde{x} \) is a local minimizer of \( f \) on \( S \). If \( \Gamma \) is a sequence of minimizers of \( F_k \) on \( B \), then \( x^{(k)} \) converges to \( \tilde{x} \) as \( k \to \infty \).

Proof. Note that \( F_k(\tilde{x}) = f(\tilde{x}) \) for all \( k \), so

\[
F_k(x^{(k)}) \leq f(\tilde{x}) \quad \text{for all } k \geq 1. \tag{26.4}
\]

Define \( \alpha_1 := \inf_{x \in B_1} f(x) \). Then the last inequality and the definition of \( F_k \) show that

\[
\alpha_1 + \frac{k}{2} \sum_{l=1}^{L} h_l(x^{(k)})^2 \leq f(\tilde{x}) \quad \text{for all } k \geq 1
\]

Since \( k \not\to \infty \), this implies that

\[
\sum_{l=1}^{L} h_l(x^{(k)})^2 \to 0^+ \quad \text{as } k \to \infty.
\]

Thus each \( h_l(x^{(k)}) \to 0 \) as \( k \to \infty \). Let \( \hat{x} \) be a limit point of the sequence \( \Gamma \). Then since each \( h_l \) is continuous, we have \( h_l(x^{(k)}) \to h_l(\hat{x}) \). Thus \( \hat{x} \in S \). Take limits in (26.4), then

\[
f(\hat{x}) + \frac{\epsilon}{2} \|\hat{x} - \tilde{x}\|^2_2 \leq f(\tilde{x}) = \alpha(f, S).
\]

This can only happen if \( \hat{x} = \tilde{x} \), so the result holds. \( \square \)
Proof of theorem 18.1:
A consequence of the preceding lemma is that, for sufficiently large \( k \), \( x^{(k)} \) will be in the interior of \( B_1 \). When this holds, then \( x^{(k)} \) will satisfy

\[
\nabla f(x) + k \sum_{l=1}^{L} h_l(x) \nabla h_l(x) + \epsilon(x - \tilde{x}) = 0 \tag{26.5}
\]

If \( x^{(k)} \in S \) then we must have \( x^{(k)} = \tilde{x} \) and this formula implies that \( \nabla f(\tilde{x}) = 0 \) which has the form (26.2).

Otherwise, define \( \lambda^k_l := kh_l(x^{(k)}) \) then (26.5) implies that

\[
\nabla f(x^{(k)}) + \sum_{l=1}^{L} \lambda^k_l \nabla h_l(x^{(k)}) = \epsilon(\tilde{x} - x^{(k)}) \tag{26.6}
\]

This is a linear equation of the form \( D_k \lambda^k = c^k \) where \( D_k := DH(x^{(k)}) := [\nabla h_1((x^{(k)}), \ldots, \nabla h_L((x^{(k)}))] \) is an \( n \times L \) matrix. The matrix \( D := DH(\tilde{x}) \) has rank \( L \) by assumption. Thus for large \( k \), \( D_k \) has rank \( L \) as the columns are continuous functions of \( x \) on \( B_1 \). \( \lambda^k \) is a column vector with \( L \) components and \( c^k := \epsilon(\tilde{x} - x^{(k)}) - \nabla f(x^{(k)}) \). Premultiply this equation by \( D_k^T \), to obtain

\[
D_k^T D_k \lambda^k = D_k^T c^k
\]

When \( k \) is large enough, the matrix on this left hand side is an \( L \times L \) matrix of rank \( L \). This equation has a unique solution for the vector \( \lambda^k \). Let \( k \to \infty \), then since the functions are continuously differentiable, \( D_k \) converges to \( \tilde{D} \) and \( c^k \to \tilde{c} \). Thus the \( \lambda^k \) converge to

\[
\tilde{\lambda} := (D^T D)^{-1} D^T \tilde{c}
\]

Now take limits as \( k \to \infty \) in (26.6), each term converges to a limit and the RHS goes to zero, so the Lagrange multiplier rule (26.2) holds at the local minimizer \( \tilde{x} \) of \( f \) on \( S \) and \( \lambda \) is unique.

This theorem is Corollary 2 of theorem 5.2 in Chapter 4 of Berkovitz, page 156. He gives a number of examples of its use.