APPENDIX
RESULTS FROM CALCULUS AND LINEAR ALGEBRA

Several useful results from real analysis and basic calculus are summarized here.

Limit point of a sequence in $\Re^n$. Let $\{x_k\}$ be a sequence in $\Re^n$. A point x^* is called a *limit point* of $\{x_k\}$ if there exists a subsequence $\{x_{k_j}\}$ that converges to x^*. If all subsequences converge to x^* them we write $\lim_{k\to\infty} x_k = x^*$.

Integral Mean-value Theorem. Let $f(x) \in C^1$ in an open interval $D \subset \Re^1$. For any two points x and $x + h \in D$,

$$f(x+h) = f(x) + \int_x^{x+h} f'(z)\,dz.$$

This relationship will often be used in one of the equivalent forms:

$$f(x+h) - f(x) - hf'(x) = \int_x^{x+h} \big(f'(z) - f'(x)\big)\,dz$$

or

$$f(x+h) - f(x) - hf'(x) = h\int_0^1 \big(f'(x+\xi h) - f'(x)\big)\,d\xi. \qquad (A.1)$$

Second-Order Integral Mean-value Theorem. Let $f(x) \in C^2$ in an open interval $D \subset \Re^1$. For any two points x and $x + h \in D$,

$$f(x+h) = f(x) + hf'(x) + h^2\int_0^1 (1-\xi)f''(x+\xi h)\,d\xi.$$

This relationship will often be expressed in the following form:

$$f(x+h) = f(x) + hf'(x) + \tfrac{1}{2}h^2 f''(x) + h^2\int_0^1 (1-\xi)\big(f''(x+\xi h) - f''(x)\big)\,d\xi. \qquad (A.2)$$

Second-Order Mean-value Theorem. Let $f : D \to \Re$ and $f(x) \in C^2$ in an open set $D \subset \Re^n$. For any two points x and $x + h \in D$ for which all points on the line segment are in D there exists a scalar $0 < \xi < 1$ such that

$$f(x+h) = f(x) + h^T\nabla f(x) + \tfrac{1}{2}h^T\nabla^2 f(x+\xi h)h. \qquad (A.3)$$

Bounds on the definite integral. If $m \le f(x) \le M$ when $a \le x \le b$, then

$$m(b-a) \le \int_a^b f(z)\,dz \le M(b-a). \qquad (A.4)$$

Lipschitz continuity. A function $h(x)$ is said to be *Lipschitz continuous* with constant γ on an open interval D, denoted $h(x) \in L_\gamma(D)$, if for every x, $y \in D$,

$$|h(x) - h(y)| \leq \gamma |x - y|.$$

Rates of convergence. Let $x^* \in \Re$, $x_k \in \Re$, $k = 0, 1, 2, \ldots$. The sequence $\{x_k\}$ is said to *converge to* x^* if

$$\lim_{k \to \infty} |x_k - x^*| = 0.$$

If, in addition, there exists a constant β $(0 \leq \beta < 1)$ and a positive integer $\bar{k}$ such that, for all $k \geq \bar{k}$,

$$|x_{k+1} - x^*| \leq \beta |x_k - x^*|,$$

then $\{x_k\}$ is said to be *q-linearly convergent* to x^*. The "q" stands for quotient and is frequently dropped. If for some sequence $\{\beta_k\}$ that converges to zero,

$$|x_{k+1} - x^*| \leq \beta_k |x_k - x^*|, \tag{A.5}$$

then $\{x_k\}$ is said to *converge q-superlinearly* to x^*. If there exist constants r $(r > 1)$, β $(\beta \geq 0)$ and $\bar{k}$ $(\bar{k} \geq 0)$ such that $\{x_k\}$ converges to x^*, and

$$|x_{k+1} - x^*| \leq \beta |x_k - x^*|^r, \tag{A.6}$$

for all $k \geq \bar{k}$, then $\{x_k\}$ is said to converge to x^* with *q-order at least* r. If $r = 2$ or 3, the convergence is said to be *q-quadratic* or *q-cubic*, respectively.

If $\{x_k\}$ converges to x^* and in place of (A.5),

$$|x_{k+j} - x^*| \leq \beta_k |x_k - x^*|,$$

for some fixed integer j and all $k \geq \bar{k}$, then $\{x_k\}$ is said to be *j-step q-superlinearly convergent* to x^*. If $\{x_k\}$ converges to x^* and in place of (A.6), for $k \geq \bar{k}$,

$$|x_{k+j} - x^*| \leq \beta |x_k - x^*|^r,$$

for some fixed integer j, then $\{x_k\}$ is said to have *j-step q-order convergence of order at least* r.

Modulus of continuity. Let D be a subset of $\Re^n$ and let $g(x) \in \Re^n$ denote a vector function defined over a set $D_0 \subset D$. The quantity

$$\omega(t) = \max\{\, \|g(x) - g(y)\| : x, y \in D_0, \|x - y\| \leq t \,\},$$

is known as the *modulus of continuity* of g on D_0. The modulus of continuity is well-defined and bounded for all $t \in [0, \infty)$ if any one of the following conditions holds:

(a) D_0 is a convex subset of D and g is uniformly continuous;
(b) D_0 is a closed bounded subset of D and g is continuous;
(c) $g(x) \in L_\gamma(D_0)$.

The singular-value decomposition.

Any matrix may be factorized in the following form:

$$A = USV^T,$$

where A is an $m \times n$ matrix, U and V are orthogonal matrices and S is a diagonal matrix. There is no loss in generality if we assume $m \geq n$. Given such a factorization it is clearly possible to ensure the diagonal elements of S are non-negative and ordered such that $\sigma_i \geq \sigma_{i-1}$, where $\sigma_i = S_{ii}$. This factorization is known as the singular-value decomposition. It is usually abreviated to SVD. The elements of σ are the singular values.

The SVD is an important computational and theoretical tool.

Spectral norm of a matrix and singular values. Consider any vector norm $\|\cdot\|$. For a given matrix A, we may define $\|A\|$ to be

$$\|A\| = \sup_{x \neq 0} \frac{\|Ax\|}{\|x\|}. \tag{A.7}$$

The non-negative quantity $\|A\|$ is known as the matrix norm of A *subordinate* to the vector norm. The matrix norm subordinate to $\|x\|_2$ is

$$\|A\|_2 = (\text{maximum eigenvalue of } A^T A)^{\frac{1}{2}}.$$

This may be proved as follows. The matrix $A^T A$ is symmetric and its eigenvalues are all non-negative because $x^T A^T A x = (Ax)^T (Ax) \geq 0$. We denote its eigenvalues by $\{\sigma_i^2\}$, where $\sigma_1^2 \leq \sigma_2^2 \leq \cdots \leq \sigma_n^2$. From its definition,

$$\|A\|_2^2 = \max \frac{\|Ax\|_2^2}{\|x\|_2^2} = \max \frac{x^T A^T A x}{x^T x}.$$

If $v_1, v_2, \ldots, v_n$ is an orthonormal set of eigenvectors of $A^T A$, we may write $x = \sum_i d_i v_i$. Hence

$$\max \frac{x^T A^T A x}{x^T x} = \frac{\sum d_i^2 \sigma_i^2}{\sum d_i^2} \leq \sigma_n^2.$$

The value σ_n^2 is attained when $x = v_i$. The non-negative quantities σ_i are the *singular values* of A. The 2-norm of a matrix is often called the *spectral norm*.

If B is a symmetric matrix, the singular values of B are the absolute values of the eigenvalues of B. Therefore

$$\|A\|_2 = \text{eigenvalue of maximum modulus of } B.$$