Chapter 1

Mathematical Preliminaries

Definition (Limit)**.** The limit L of a function (if it exists) at c is given by $\lim_{x\to c} f(x) = L$ which means that to each $\epsilon > 0$ there is a corresponding $\delta > 0$ such that $\mid f(x) - L \mid < \epsilon$ whenever $0 <\mid x - c \mid < \delta$.

Definition (Continuity)**.** A function f is said to be continuous at c if $\lim_{x\to c} f(x) = f(c)$.

Definition (Derivative)**.** The derivative of a function f at c (if it exists) is defined by $f'(c) = \lim_{x\to c} \frac{f(x)-f(c)}{x-c}$. If $f'(c)$ exists then f is said to be differentiable at c.

Definition (Linear Rate of Convergence)**.** Let $[x_n]$ be a sequence of real numbers tending to a limit $x^\star$. We say that the rate of convergence is at least linear if there is a constant $c < 1$ and an $N \in \mathbb{Z}$ such that $\mid x_{n+1} - x^\star \mid\leq c \mid x_n - x^\star \mid$ for $n \geq N$.

Definition (Superlinear Rate of Convergence)**.** Let $[x_n]$ be a sequence of real numbers tending to a limit $x^\star$. We say that the rate of convergence is at least superlinear if there exists a sequence ϵ_n tending to 0 and an $N \in \mathbb{Z}$ such that $\mid x_{n+1} - x^\star \mid\leq \epsilon_n \mid x_n - x^\star \mid$ for $n \geq N$.

Note a sequence that has superlinear rate of convergence also has linear rate of convergence.

Definition (Quadratic Rate of Convergence)**.** Let $[x_n]$ be a sequence of real numbers tending to a limit $x^\star$. We say that the rate of convergence is at least quadratic if there is a constant C and an $N \in \mathbb{Z}$ such that $\mid x_{n+1} - x^\star \mid\leq C \mid x_n - x^\star \mid^2$ for $n \geq N$.

In general, if there are positive constants C and α and $N \in \mathbb{Z}$ such that $\mid x_{n+1} - x^\star \mid\leq C \mid x_n - x^\star \mid^\alpha$ for $n \geq N$, then we say that the rate of convergence is of *order* α at least.

Remark. Quadratic convergences $\Rightarrow$ Superlinear convergence $\Rightarrow$ Linear convergence

Definition (Big $\mathcal{O}$ Notation for Sequences)**.** Let $[x_n]$ and $[\alpha_n]$ be two different sequences. We write $x_n = \mathcal{O}(\alpha_n)$ if there are constants C and n_0 such that $|x_n| \leq C\,|\alpha_n|$ when $n \geq n_0$.

Definition (Little o Notation for Sequences)**.** Let $[x_n]$ and $[\alpha_n]$ be two different sequences. We write $x_n = o(\alpha_n)$ if for some $\epsilon_n \geq 0$ we have $\epsilon_n \to 0$ and $|x_n| \leq \epsilon_n\,|\alpha_n|$.

Definition (Big $\mathcal{O}$ Notation for Functions)**.** We write $f(x) = \mathcal{O}(g(x))$ as $x \to x^\star$ when there is a constant C and a $\delta > 0$ such that $|f(x)| \leq C\,|g(x)|$ for $|x - x^\star| < \delta$.

$f(x) = \mathcal{O}(g(x))$ as $x \to \infty$ means there exists constants C and r such that $|f(x)| \leq C\,|g(x)|$ whenever $x \geq r$.

Fact (Least Upper Bound Axiom)**.** *Any nonempty set of real numbers that possesses an upper bound has a least upper bound.*

Definition (Supremum (Least Upper Bound))**.** v is said to be the supremum of S, written $v = \sup S = \text{lub } S$, if and only if

1. v is an upper bound for S and
2. no real number smaller than v is an upper bound for S.

Definition (Infimum (Greatest Lower Bound))**.** u is said to be the infimum of S, written $u = \inf S = \text{glb } S$, if and only if

1. u is a lower bound for S and
2. no real number greater than u is a lower bound for S.

Definition (Shift/Displacement Operator)**.** Let $x = [x_1, x_2, \ldots]$ be a vector (or sequence). Then a shift/displacement operator denoted by E is defined as $(Ex)_n = x_{n+1}$. Example, $Ex = [x_2, x_3, \ldots]$ where $x = [x_1, x_2, \ldots]$.

Repeated application of E gives $(E^k x)_n = x_{n+k}$. Example, $(EEx)_n = x_{n+2}$.

Definition (Linear Difference Operator)**.** A linear difference operator L is defined as $L = \sum_{i=0}^{m} c_i E^i$. Here E^0 is the identity operator, i.e. $(E^0 x)_n = (Ix)_n = x_n$. Since L is a polynomial in E, it can be written as $L = p(E)$ where the polynomial p is called the *characteristic polynomial* of L and is defined by $p(\lambda) = \sum_{i=0}^{m} c_i \lambda^i$. The equation $Lx = 0$ is called a *linear difference equation.*

Definition (Stable Difference Equation)**.** A difference equation of the form $p(E)x = 0$ is said to be stable if all of its solutions are bounded.

Chapter 2

Computer Arithmetic

Definition (Normalized Scientific Notation)**.** Any real number x is written in a decimal form by shifting all of its digits to the right of the decimal point (so that only 0 remains before the decimal point) and using appropriate powers of 10 (in base 10). In base 10, the first digit displayed is not 0. Such a representation of a real number is called normalized scientific notation. Examples: $732.5051 = 0.7325051 \times 10^3$ and $-0.005612 = -0.5612 \times 10^{-2}$.

In base 2, scientific notation is of the form $x = \pm q \times 2^m$. The number q is called the *mantissa* and the integer m is called the *exponent*. Both q and m are base 2 numbers.

Definition (Left-shifted normalized binary number)**.** is such that the first nonzero bit in the mantissa is just before the binary point, i.e. $q = (1.f)_2$. This bit can be assumed to be 1 and does not require storage.

Definition (Normalized Floating Point Form)**.** Any real number x expressed as $(-1)^s \times q^m$ where $q = (1.f)_2$, $m = e - 127$ and s is the bit representing the sign of x (positive: bit 0, negative: bit 1) is called a normalized floating point form.

When a real number x is approximated by another number $x^\star$, the

1. *error* is $x - x^\star$.
2. *absolute error* is $|x - x^\star|$.
3. *relative error* is $\left|\dfrac{x - x^\star}{x}\right|$.

Definition (Unit Roundoff Error)**.** The bound of the relative error is called the unit roundoff error.

Definition (Conditioning)**.** A problem is *ill conditioned* if small changes in the data can produce large changes in the answers. Otherwise the problem is called *well conditioned.*

Definition (Condition Number for a Function)**.** Consider a function f evaluated at a point x. If x is perturbed slightly then the effect on $f(x)$ is given by $\dfrac{xf'(x)}{f(x)}$. This factor is called the condition number.

Definition (Condition Number for Linear Systems)**.** Condition number of a matrix A, denoted $\kappa(A)$, is $\kappa(A) = \|A\| \cdot \left\|A^{-1}\right\|$ where $\|\cdot\|$ is the norm of the matrix. In solving linear systems $Ax = b$, this condition number studies how small perturbations in b affects the computed solution x. It is always true that $\kappa(A) \geq 1$.

Chapter 3

Solution of Nonlinear Equations

Definition (Bisection (Interval Halving) Method)**.** Let f be a continuous function on $[a, b]$. A consequence of the Intermediate-Value Theorem is that if $f(a)f(b) < 0$ then f must have a zero in (a, b).

The bisection method exploits this idea in the following way. If $f(a)f(b) < 0$ then compute $c = \frac{a+b}{2}$ and test whether $f(a)f(c) < 0$. If this is true, then f has at least one zero in $[a, c]$. So rename c as b and start again with the new interval $[a, b]$, which is half as large as the original interval. If $f(a)f(c) > 0$ then $f(c)f(b) < 0$, and in this case rename c as a. In either case, a new interval containing a zero of f has been produced, and the process can be repeated. If $f(a)f(c) = 0$ then $f(c) = 0$ and a zero has been found.

Look at Figure 3.2(a) and (b) on page 75 to understand why the bisection method only finds one zero but not all of them in $[a, b]$.

Definition (Newton's Method)**.** Let f be the function whose zeros are to be determined numerically. Let r be a zero of f and let x be an approximation to r. If f'' exists and is continuous, then by Taylor's Theorem,

$$0 = f(r) = f(x + h) = f(x) + hf'(x) + \mathcal{O}(h^2)$$

where $h = r - x$. If h is small (that is, x is near r), then it is reasonable to ignore the $\mathcal{O}(h^2)$ term and solve the remaining equation for h. When this is done, $h = -\frac{f(x)}{f'(x)}$. If x is an approximation to r, then $x - \frac{f(x)}{f'(x)}$ should be a better approximation to r.

Newton's method begins with an estimate x_0 of r and then defines inductively $x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$, $n \geq 0$.

Definition. Newton's method is faster than the bisection and the secant methods since its convergence is quadratic rather than linear or superlinear. Unfortunately, this method is not guaranteed always to converge. Look at Figure 3.4 on page 83 to understand the graphical interpretation of Newton's method.

Definition (Secant Method)**.** One of the drawbacks of Newton's method is that it involves the derivative of the funnction whose zero is sought. So, replace $f'(x_n)$ by $\frac{f(x_n)-f(x_{n-1})}{x_n-x_{n-1}}$ which comes from the limit definition of a derivative. This resulting algorithm is called the secant method and its formula is $x_{n+1} = x_n - f(x_n)\left(\frac{x_n-x_{n-1}}{f(x_n)-f(x_{n-1})}\right)$, $n \geq 1$.

Definition (Quadratic Convergence)**.** Let $e_n = x_n - r$ be the error where r is simple zero of f and x_n is the sequence. Then $e_{n+1} = Ce_n^2$ for some constant C is called quadratic convergence.

Definition (Functional Iteration)**.** Functional iteration is given by $x_{n+1} = F(x_n)$, $n \geq 0$. For example, in Newton's method, the function F is given by $F(x) = x - \frac{f(x)}{f'(x)}$.

Definition (Fixed Point)**.** Assume $\lim_{n\to\infty} x_n$ exists, and is equal to s. If the functional iteration F is continuous then $F(s) = F(\lim_{n\to\infty} x_n) = \lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} x_{n+1} = s$. Thus, $F(s) = s$, and s is called a fixed point of F.

Definition (Contractive Mapping)**.** A mapping (or function) F is said to be contractive if there exists a number $\lambda < 1$ such that $|F(x) - F(y)| \leq \lambda\,|x - y|$ for all points x and y in the domain of F. Look at Figure 3.7, page 101, for a graphical interpretation.

Definition (Cauchy Criterion)**.** Let $[x_n]$ be a sequence. Given any ϵ, there exists an integer N such that $|x_n - x_m| < \epsilon$ whenever $n, m \geq N$.

Definition (Horner's Algorithm)**.** If $p(z) = a_n z^n + a_{n-1} z^{n-1} + \cdots + a_1 z + a_0$ and a complex number z_0 are given, then Horner's algorithm produces the number $p(z_0)$ and the polynomial $q(z) = \frac{p(z) - p(z_0)}{z - z_0}$ whose degree is one less than the degree of p. So, write $q(z) = b_0 + b_1 z + \cdots + b_{n-1} z^{n-1}$. Substituting $p(z)$ and the new form of $q(z)$ into the equation $p(z) = (z - z_0)q(z) + p(z_0)$ (which is obtained from the original form of $q(z)$), and then setting the coefficients of like powers equal to each other from both sides of the equation gives,

$$\begin{aligned} b_{n-1} &= a_n \\ b_{n-2} &= a_{n-1} + z_0 b_{n-1} \\ &\vdots \\ b_0 &= a_1 + z_0 b_1 \\ p(z_0) &= a_0 + z_0 b_0. \end{aligned}$$

In compact form, Horner's algorithm can be written as

$$\begin{array}{c|cccc} & a_n & a_{n-1} & \cdots & a_0 \\ z_0 & & z_0 b_{n-1} & & z_0 b_0 \\ \hline & b_{n-1} & b_{n-2} & \cdots & \boxed{a_0 + z_0 b_0} \end{array}$$

The boxed number is the value of $p(z_0)$.

Definition (Laguerre Iteration)**.** Let p be a polynomial of degree n. Then this algorithm proceeds iteratively from one approximate root z to a new one by calculating

$$\begin{aligned} A &= -\frac{p'(z)}{p(z)} \\ B &= \frac{A^2 - p''(z)}{p(z)} \\ C &= \frac{A \pm \sqrt{(n-1)(nB - A^2)}}{n} \\ z_{new} &= z + \frac{1}{C} \end{aligned}$$

In the definition of C, the sign of C is chosen such that $|C|$ as large as possible.

Definition (Basin of Attraction)**.** If Newton's method is started at a point z in the complane plane, it produces a sequence defined by the equations

$$\begin{cases} z_0 &= z \\ z_{n+1} &= z_n - \frac{p(z_n)}{p'(z_n)} \end{cases}$$

where $n \geq 0$. If $\lim_{n \to \infty} z_n = \xi$, we say that z (the starting point) is *attracted* to ξ. The set of all points z that are attracted to ξ is called the *basin of attraction* corresponding to ξ.

Every root of p has a basin of attraction, and these are mutually disjoint sets because a sequence that converges to one root of p cannot converge to another root.

Definition (Homotopy)**.** A homotopy between two functions $f, g : X \to Y$ is a continuous map $h : [0,1] \times X \to Y$ such that $h(0, x) = g(x)$ and $h(1, x) = f(x)$. If such a map exists, then f is said to be homotopic to g.

Homotopy forms an equivalence relation among the continuous maps from X to Y, where X and Y are any two topological spaces. Read page 135 to see how homotopy relates to Newton's method.

Chapter 4

Solving Systems of Linear Equations

Definition (Symmetric Matrix)**.** means $A^T = A$.

Definition (Diagonally Dominant Matrix)**.** satisfies the inequality $|a_{ii}| > \sum_{j=1, j\neq i}^{n} |a_{ij}|$ for $1 \leq i \leq n$. Example, $\begin{bmatrix} 4 & -1 & 0 & -1 \\ -1 & 4 & 0 & -1 \\ -1 & 0 & 4 & -1 \\ 0 & -1 & -1 & 4 \end{bmatrix}$.

Definition (Tridiagonal Matrix)**.** $a_{ij} = 0$ for all pairs (i, j) that satisfy $|i - j| > 1$. Thus, in the ith row, only $a_{i,i-1}$, a_{ii}, $a_{i,i+1}$ can be different from 0.

Definition (Vector Norm)**.** A norm is a function $\|\cdot\|$ from a vector space V to the set of nonnegative reals that obeys the following three properties:

1. $\|x\| > 0$ if $x \neq 0$, $x \in V$
2. $\|\lambda x\| = |\lambda| \, \|x\|$ if $\lambda \in \mathbb{R}$, $x \in V$
3. $\|x + y\| \leq \|x\| + \|y\|$ if x,$y \in V$

Definition (Euclidean l_2-Norm)**.** is defined as $\|x\|_2 = \left(\sum_{i=1}^{n} x_i^2 \right)^{\frac{1}{2}}$ where $x = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$.

Definition (l_∞-Norm)**.** $\|x\|_\infty = \max_{1\leq i\leq n} |x_i|$

Definition (l_1- Norm)**.** $\|x\|_1 = \sum_{i=1}^{n} |x_i|$

Definition (Matrix Norm Subordinate to a Vector Norm)**.** If a vector norm $\|\cdot\|$ has been specified, the matrix norm *subordinate* to it is defined by $\|A\| = \sup \{\|Au\| : u \in \mathbb{R}^n, \|u\| = 1\}$. This is also called the matrix norm *associated* with the given vector norm. Here, A is an $n \times n$ matrix.

Definition (l_2-matrix norm/Spectral Norm and Spectral Radius)**.** There are two definitions:

1. l_2-matrix norm/Spectral Norm is defined as $\|A\|_2 = \sup_{\|x\|_2=1} \|Ax\|_2$. This is the matrix norm subordinate to the Euclidean vector norm. In fact, $\|A\|_2 = \max_{1\leq i\leq n} |\sigma_i|$ where σ_i are the singular values of A. If $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n \geq 0$, then $Av_1 = \sigma_1 u_1$ and $A^T u_1 = \sigma_1 v_1$. Hence, $A^T A v_1 = \sigma_1^2 v_1$. So σ_1^2 is the largest eigenvalue of $A^T A$. Thus, the 2-matrix norm is defined as $\|A\|_2 = \sqrt{\rho(A^T A)}$ where $\rho(A^T A)$ is the *spectral radius* of $A^T A$ and is the largest eigenvalue of $A^T A$.

2. The *eigenvalues* of a matrix A are the complex numbers λ for which the matrix $A-\lambda I$ is not invertible. These numbers are then the roots of the *characteristic equation* of A: $\det(A-\lambda I)=0$. The *spectral radius* of A is defined by $\rho(A)=\max\{|\lambda| : \det(A-\lambda I)=0\}$.

Definition (Normed Linear Space)**.** If a vector space V is assigned a norm $\|\cdot\|$, then the pair $(V,\|\cdot\|)$ is a normed linear space.

Definition (Vector Convergence)**.** A given sequence of vectors $v^{(1)}, v^{(2)}, \ldots$ is said to *converge* to a vector v if $\lim_{k\to\infty} \left\|v^{(k)}-v\right\|=0$.

Definition (Iterative Refinement)**.** Suppose that $Ax=b$ has been solved by Gaussian elimination and $x^{(0)}$ is an approximate solution. Compute the residual vector $r^{(0)}=b-Ax^{(0)}$ in double precision, and $Ae^{(0)}=r^{(0)}$ where $e^{(0)}=A^{-1}(b-Ax^{(0)})$ is the error vector, and the next iteration $x^{(1)}=x^{(0)}+e^{(0)}$. To obtain better solutions $x^{(2)}, x^{(3)}, \ldots$, this process can be repeated. The sequence of vectors $x^{(m)}$ converge to x.

Definition (Row and Column Equilibrium)**.** is the process of dividing each row of the coefficient matrix by the maximum element in absolute value in that row; that is, mulitplying row i by $r_i=\frac{1}{\max\limits_{1\le j\le n}|a_{i,j}|}$ for $1\le i\le n$. Column equilibrium is similar except division is done columnwise, that is, mulitply column j by $c_j=\frac{1}{\max\limits_{1\le i\le n}|a_{i,j}|}$ for $1\le j\le n$.

Definition (Similar Matrix)**.** A matrix A is said to be similar to a matrix B if there is a nonsingular matrix S such that $S^{-1}AS=B$.

Definition (Positive Definite)**.** There are two definitions:

1. The matrix A is positive definite if $x^T Ax>0$ for all nonzero vector x.

2. A matrix A is said to be positive definite if $\langle Ax, x\rangle>0$ for all $x\neq 0$.

Definition (Orthogonal Vectors)**.** Vectors u and v are orthogonal if $\langle u, v\rangle=0$.

Definition (Orthonormal Vectors)**.** Vectors u and v are orthogonal if $\langle v_i, v_j\rangle=\delta_{ij}$.

Definition (A-orthonormality)**.** Assume A is an $n\times n$ symmetric and positive definite matrix. Suppose that a set of vectors $\left\{u^{(1)}, u^{(2)}, \ldots, u^{(n)}\right\}$ is provided and has the property $\langle u^{(i)}, Au^{(j)}\rangle=\delta_{ij}$ $(1\le i,j\le n)$. This property is called A-orthonormality.

Chapter 5

Selected Topics in Numerical Linear Algebra

Definition (Unitary/Hermitian Matrix)**.** Consider a matrix U such that $UU^\star = I$. If $\mathbb{F} = \mathbb{C}$ then $U^\star = \left(\overline{U}\right)^T$ and U is called a unitary matrix. If $\mathbb{F} = \mathbb{R}$ then $U^\star = U^T$ and U is called a Hermitian matrix. In both cases, $UU^\star = I$ implies that $U^\star = U^{-1}$.

Definition (Unitarily Similar)**.** Matrices A and B are unitarily similar if $B = UAU^\star$ for some unitary matrix U.

Chapter 6

Approximating Functions

Definition (Interpolating Polynomial)**.** Given a table of $n+1$ data points $(x_0, y_0), \ldots, (x_n, y_n)$, we seek a polynomial p of lowest possible degree for which $p(x_i) = y_i$ $(0 \leq i \leq n)$. Such a polynomial is said to interpolate the data.

Definition (Interpolating Polynomial in Newton's Form)**.** is $p_k(x) = \sum_{i=0}^{k} c_i \prod_{j=0}^{i-1} (x - x_j)$. If $i - 1 < 0$, then $\prod_{j=0}^{i-1} (x - x_j) = 1$. The coefficients $c_i = f[x_0, x_1, \ldots, x_i]$ is found by the method of Divided Differences. The polynomials in the product can be found by Horner's algorithm/nested multiplication. A pseudocode for this is on page 310.

Definition (Lagrange Form of the Interpolating Polynomial)**.** The (Lagrange form) of the interpolating polynomial is expressed as $p(x) = \sum_{k=0}^{n} y_k l_k(x)$ where $l_0, l_1, \ldots, l_n$ are polynomials that depend on the (distinct) nodes $x_0, x_1, \ldots, x_n$ but not on the *ordinates* $y_0, y_1, \ldots, y_n$. The ordinates are set to be $l_i(x_j) = \delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$. Then, a general form of the polynomials l_i is $l_i(x) = \prod_{j=0, j\neq i}^{n} \frac{x - x_j}{x_i - x_j}$ $(0 \leq i \leq n)$. (For example, setting $x = x_0$ gives that $l_0(x_0) = 1$.) The set of nodes $x_0, x_1, \ldots, x_n$ is called the *cardinal functions.*

Definition (Vandermonde Matrix)**.** Given a vector $x = (x_0, x_1, \ldots, x_n)$, the Vandermonde matrix is

$$\begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{bmatrix}$$

According to Problem 6.1.34 (p.327), the determinant of this matrix is $\prod_{0 \leq j < k \leq n} (x_k - x_j)$. This determinant is nonzero, and therefore, this matrix is nonsingular. Thus, the system

$$\begin{aligned} p(x) &= a_0 + a_1 x + \cdots + a_n x^n \\ &= \begin{bmatrix} 1 & x_0 & x_0^2 & \cdots & x_0^n \\ 1 & x_1 & x_1^2 & \cdots & x_1^n \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & \cdots & x_n^n \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_n \end{bmatrix} \end{aligned}$$

has a unique solution for any choice of $y_0, y_1, \ldots, y_n$. However, the Vandermonde matrix is often ill conditioned, and the coefficients a_i may be inaccurately determined.

Definition (Chebyshev Polynomials)**.** (of the first kind) are defined recursively as follows:

$$\begin{cases} T_0(x) & = 1 \\ T_1(x) & = x \\ T_{n+1}(x) & = 2xT_n(x) - T_{n-1}(x), (n \geq 1) \end{cases}$$

Definition (Hermite Interpolation)**.** refers to the interpolation of a function and some of its derivatives at a set of nodes.

Definition (Lagrange Interpolation)**.** refers to the interpolation of a function where no derivatives are interpolated at a set of nodes.

Definition (Spline Function of Degree k)**.** Suppose that $n+1$ points $t_0, t_1, \ldots, t_n$ (called *knots*) have been specified and satisfy $t_0 < t_1 < \cdots < t_n$. Suppose also that an integer $k \geq 0$ has been prescribed. A spline function of degree k having knots $t_0, t_1, \ldots, t_n$ is a function S such that:

1. On each interval $[t_{i-1}, t_i)$, S is a polynomial of degree $\leq k$.
2. S has a continuous $(k-1)$st derivative on $[t_0, t_n]$.

Definition (Tension Spline)**.** is a function f having the following properties:

1. $f \in C^2[t_0, t_n]$
2. The interpolation conditions $f(t_i) = y_i$ hold for $0 \leq i \leq n$.
3. On each open interval (t_{i-1}, t_i), f satisfies $f^{(4)} - \tau^2 f'' = 0$.

Definition (Truncated Power Function)**.** is a function of continuity class C^{n-1} defined as $\begin{cases} x^n & x \geq 0 \\ 0 & x < 0 \end{cases}$.

Definition (Modulus of Continuity)**.** for a function f (continuous or not) is defined to be $\omega(f;\delta) = \max\limits_{|s-t|\leq\delta} |f(s) - f(t)|$.

Definition (B-Splines of Degree 0)**.** is defined to be $B_i^0(x) = \begin{cases} 1 & t_i \leq x < t_{i+1} \\ 0 & \text{otherwise} \end{cases}$ where $i \in \mathbb{Z}$. Its graph is on page 366. Its properties are:

1. The *support* of B_i^0, defined as the set of x where $B_i^0(x) \neq 0$, is the interval $[t_i, t_{i+1})$.
2. $B_i^0(x) \geq 0$ for all i and all x.
3. B_i^0 is continuous from the right on the entire real line.
4. $\sum\limits_{i=-\infty}^{\infty} B_i^0(x) = 1$ for all x.

Definition (B-Splines of Degree 1)**.** is defined to be

$$\begin{aligned} B_i^1(x) &= \left(\frac{x - t_i}{t_{i+1} - t_i}\right) B_i^0(x) + \left(\frac{t_{i+2} - x}{t_{i+2} - t_{i+1}}\right) B_{i+1}^0(x) \\ &= \begin{cases} 0 & x < t_i \text{ or } x \geq t_{i+2} \\ \frac{x - t_i}{t_{i+1} - t_1} & t_i \leq x \leq t_{i+1} \\ \frac{t_{i+2} - x}{t_{i+2} - t_{i+1}} & t_{i+1} \leq x < t_{i+2} \end{cases} \end{aligned}$$

Its graph is on page 368. Its properties are:

1. The *support* of B_i^1 is (t_i, t_{i+2}).

2. $B_i^1(x) \geq 0$ for all i and all x.

3. B_i^1 is continuous and is differentiable at every point except t_i, t_{i+1}, and t_{i+2}.

4. $\sum\limits_{i=-\infty}^{\infty} B_i^1(x) = 1$ for all x.

Definition (Radius of Convergence)**.** For every power series $\sum\limits_{k=0}^{\infty} a_k(x-c)^k$, there is a number r in the range $[0,\infty]$ such that the series diverges for $|x-c| > r$ and converges for $|x-c| < r$. This number r is called the radius of convergence.

Definition (Critical Set of a Function)**.** The critical set of a function f in the space of all continuous real-valued functions $C(X)$ is $crit(f) = \{x \in X : |f(x)| = \|f\|\}$ where $\|f\| = dist(f, G)$ and G is a finite-dimensional subspace in $C(X)$.

Definition (Convex Combinations)**.** Linear combinations of vectors of the form $\sum\limits_{i=1}^{k} \theta_i u_i$ are called convex combinations when $\sum\limits_{i=1}^{k} \theta_i = 1$ and $\theta_i \geq 0$.

Definition (Convex Hull)**.** The set of all convex combinations of points selected from a given set S is called the convex hull of S. Thus, $co(S) = \left\{ \sum\limits_{i=1}^{k} \theta_i u_i : k \in \mathbb{N}, u_i \in S, \theta_i \geq 0, \sum\limits_{i=1}^{k} \theta_i = 1 \right\}$.

Definition (Convex Set)**.** A set K in a linear space is said to be convex if it contains every line segment connecting two points of K. Thus, for $u, v \in K$ and $0 \leq \theta \leq 1$, $\theta u + (1-\theta)v \in K$.

Definition (Haar Subspace)**.** An n-dimensional subspace G in $C(X)$ is called a Haar subspace if no element of G (except 0) has n or more zeros in X.