MATHEMATICS STUDY GUIDE

Real Field

Archimedean property
If $x > 0$ and if y is an arbitrary real number, there exists a positive integer n such that $nx > y$. As a corollary, if three real numbers a, x, and y satisfy the inequalities $a \leq x \leq a + \frac{y}{n}$ for every integer $n \geq 1$, then $x = a$.

Properties of supremum and infimum
Let h be a given positive number and let S be a set of real numbers.
(a) If S has a supremum, then for some x in S we have $x > \sup S - h$.
(b) If S has an infimum, then for some x in S we have $x < \inf S + h$.

Well-ordering principle
Every nonempty set of positive integers contains a smallest member.

Triangle inequality
For arbitrary real numbers x and y, $|x + y| \leq |x| + |y|$. More generally, for arbitrary real numbers a_1, a_2, $\ldots$, a_n, we have $|\sum_{k=1}^{n} a_k| \leq \sum_{k=1}^{n} |a_k|$.

The Cauchy-Schwarz inequality
If $a_1, \ldots, a_n$ and $b_1, \ldots, b_n$ are arbitrary real numbers, we have $\left(\sum_{k=1}^{n} a_k b_k\right)^2 \leq \left(\sum_{k=1}^{n} a_k^2\right)\left(\sum_{k=1}^{n} b_k^2\right)$. The equality sign holds if and only if there is a real number x such that $a_k x + b_k = 0$ for each $k = 1, 2, \ldots, n$.

Complex Field

Field properties
$(a, b) = (c, d)$ means $a = c$ and $b = d$
$(a, b) + (c, d) = (a + c, b + d)$
$(a, b)(c, d) = (ac - bd, ad + bc)$ $x + y = y + x$
$x + (y + z) = (z + y) + z$
$x(y + z) = xy + xz$
$e^{z+2n\pi i} = e^z$

Polar coordinates
$x = r\cos\theta \quad y = r\sin\theta$
r is the modulus or absolute value of (x, y), equal to $\sqrt{x^2 + y^2}$.
θ is the angle between (x, y) and the x-axis, and is called the argument of (x, y), or the principal argument if $-\pi < \theta \leq \pi$.
Polar form of z: Every complex number $z \neq 0$ can be expressed as $z = re^{i\theta}$.

Complex exponential
If $z = (x, y)$, then $e^z = e^x(\cos y + i \sin y)$
$e^a e^b = e^{a+b}$

Derivatives and integrals
If $f = u + iv$, then $f'(x) = u'(x) + iv'(x)$
$\int_a^b f(x)\,dx = \int_a^b u(x)\,dx + i\int_a^b v(x)\,dx$
$(e^{tx})' = te^{tx}$
$\int e^{tx}\,dx = \frac{e^{tx}}{t}$

Integrals

Upper and lower integrals
Lower integral of f: $\underline{I} = \sup\left\{\int_a^b s(x)\,dx \mid s \leq f\right\}$ Upper integral of f: $\bar{I} = \inf\left\{\int_a^b t(x)\,dx \mid f \leq t\right\}$

Definition of the integral
Every function f which is boundend on $[a, b]$ has a lower integral $\underline{I}(f)$ and an upper integral $\bar{I}(f)$ satisfying the inequalities $\int_a^b s(x)\,dx \leq \underline{I}(f) \leq \bar{I}(f) \leq \int_a^b t(x)\,dx$ for all step functions s and t with $s \leq f \leq t$. The function f is integrable on $[a, b]$ if and only if its upper and lower integrals are equal, in which case we have $\int_a^b f(x)\,dx = \underline{I}(f) = \bar{I}(f)$.

Basic properties of the integral
Linearity: $\int_a^b [c_1 f(x) + c_2 g(x)]\,dx = c_1\int_a^b f(x)\,dx + c_2\int_a^b g(x)\,dx$.
Additivity: $\int_a^b f(x)\,dx + \int_b^c f(x)\,dx = \int_a^c f(x)\,dx$.
Invariance under translation: $\int_a^b f(x)\,dx = \int_{a+c}^{b+c} f(x-c)\,dx$.
Expansion or contraction of interval of integration: $\int_a^b f(x)\,dx = \frac{1}{k}\int_{ka}^{kb} f\left(\frac{x}{k}\right)\,dx$.

Definition of indefinite integral
$A(x) = \int_a^x f(t)\,dt, \qquad \text{if } a \leq x \leq b.$

Improper integrals
Improper integral of first kind: $\int_a^b$ where we let $b \to +\infty$.
Improper integral of second kind: $\int_a^b f$ where f is allowed to become unbounded at one or more points.

Average value of a function over an interval
$A(f) = \frac{1}{b-a}\int_a^b f(x)\,dx$.

Integral of rational powers
$\int x^n\,dx = \frac{x^{n+1}}{n+1}$

Integral of trigonometric functions
$\int \sin x\,dx = -\cos x$
$\int \cos x\,dx = \sin x$

Integrals involving logarithm
$\int \frac{du}{u} = \log|u| + C, \quad \text{where } u \neq 0$
$\int \frac{f'(x)}{f(x)}\,dx = \log|f(x)| + C, \quad \text{where } f(x) \neq 0$

Some useful integrals
$\int \frac{dx}{x+a} = \log|x+a|$
$\int \frac{dx}{(x+a)^n} = \frac{(x+a)^{1-n}}{(1-n)}, \quad$ if $n > 1$
For $\int \frac{x\,dx}{(x^2+bx+c)^m}$ and $\int \frac{dx}{(x^2+bx+c)^m}$, let $u = x + b/2$ and $\alpha = \frac{1}{2}\sqrt{4c-b^2}$ so that $x^2+bx+c = u^2+\alpha^2$.
Then, $\int \frac{u\,du}{(u^2+\alpha^2)} = \frac{1}{2}\log(u^2+\alpha^2)$, and $\int \frac{u\,du}{(u^2+\alpha^2)^m} = \frac{1}{2}\frac{(u^2+\alpha^2)^{1-m}}{(1-m)}$.
And $\int \frac{du}{(u^2+\alpha^2)} = \frac{1}{\alpha}\arctan\frac{u}{\alpha} + C$.
And $\int \frac{du}{(u^2+\alpha^2)^m} = \frac{1}{2\alpha^2(m-1)}\frac{u}{(u^2+\alpha^2)^{m-1}} + \frac{2m-3}{2\alpha^2(m-1)}\int \frac{du}{(u^2+\alpha^2)^{m-1}}$.
$\int e^{\alpha x}\cos\beta x\,dx = \frac{e^{\alpha x}(\alpha\cos\beta x+\beta\sin\beta x)}{\alpha^2+\beta^2}$
$\int e^{\alpha x}\sin\beta x\,dx = \frac{e^{\alpha x}(\alpha\sin\beta x-\beta\cos\beta x)}{\alpha^2+\beta^2}$

Techniques of integration

Substitution
$\int_c^x f[g(t)]g'(t)\,dt = \int_{g(c)}^{g(x)} f(u)\,du$

Parts
$\int_a^b f(x)g'(x)\,dx = f(b)g(b) - f(a)g(a) - \int_a^b f'(x)g(x)\,dx$
Or, $\int u\,dv = uv - \int v\,du + C$

Partial fractions
The trick is to find a way to decompose a rational function into partial fractions, which are easier to integrate.
CASE 1: Denominator product of distinct linear factors.
CASE 2: Denominator product of linear factors, some of which are repeated.
CASE 3: Denominator contains irreducible ($b^2 - 4c < 0$) quadratic factors, none of which are repeated.
CASE 4: Denominator contains irreducible quadratic factors, some of which are repeated.
If we have $\int R(\sin x, \cos x)\,dx$, where R is a RF of two variables, use substitution $u = \tan\frac{1}{2}x$ to give an RF.
If we have $\int R(x, \sqrt{a^2-x^2})\,dx$, use substitution $x = a\sin t, \quad dx = a\cos t\,dt$ to give an RF.

Derivatives

Definition
$f'(x) = \lim_{h\to 0}\frac{f(x+h)-f(x)}{h}$, if this limit exists.

Algebra
$(f+g)' = f' + g'$
$(f-g)' = f' - g'$
$(f\cdot g)' = f\cdot g' + g\cdot f'$
$(\frac{f}{g})' = \frac{g\cdot f' - f\cdot g'}{g^2}$, when $g(x) \neq 0$
$(c_1 f + c_2 g)' = c_1 f' + c_2 g'$
If $f = u\circ v$ and if $v'(x)$ and $u'(x)$ exist, then $f'(x) = u'(v(x))\cdot v'(x)$ (Chain rule).

Rolle's Theorem
If f is continuous on $[a,b]$ and differentiable on (a,b), and if $f(a) = f(b)$, then there is at least one point c in (a,b) such that $f'(c) = 0$.

Mean value theorem - derivatives
If f is continuous on $[a,b]$ and differentiable on (a,b), then there is at least one point c in (a,b) such that $f(b)-f(a) = f'(c)(b-a)$.

Derivatives of inverse functions
If f and g are inverses, and if f is strictly increasing and continuous, and if $f'(x)$ exists and is nonzero, and if $y = f(x)$, then $g'(y)$ also exists, and the derivatives are reciprocals, i.e., $g'(y) = \frac{1}{f'(x)}$, or, $\frac{dx}{dy} = \frac{1}{\left(\frac{dy}{dx}\right)}$.

Wronskian matrix
The Wronskian matrix of n functions $u_1, \ldots, u_n$:

$$W = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \\ u_1' & u_2' & \cdots & u_n' \\ \vdots & & & \vdots \\ u_1^{(n-1)} & u_2^{(n-1)} & \cdots & u_n^{(n-1)} \end{bmatrix}$$

Derivative of a scalar field with respect to a vector

$$f'(\boldsymbol{a};\boldsymbol{y}) = \lim_{h \to 0} \frac{f(\boldsymbol{a} + h\boldsymbol{y}) - f(\boldsymbol{a})}{h}$$

If f is a linear transformation, then $f'(\boldsymbol{a};\boldsymbol{y}) = f(\boldsymbol{y})$.
If $\|y\| = 1$, then $f'(\boldsymbol{a};\boldsymbol{y})$ is a *directional derivative.*
If $\boldsymbol{y} = \boldsymbol{e}_k$, then $f'(\boldsymbol{a};\boldsymbol{y})$ is a *partial derivative with respect to* $\boldsymbol{e}_k$.
$D_k f(\boldsymbol{a}) = f'(\boldsymbol{a};\boldsymbol{e}_k)$.

Mean value theorem for derivatives of scalar fields
If $f'(\boldsymbol{a} + t\boldsymbol{y});\boldsymbol{y})$ exists for $0 \leq t \leq 1$
...then for some real θ such that $0 < \theta < 1$,
...$f(\boldsymbol{a} + \boldsymbol{y}) - f(\boldsymbol{a}) = f'(\boldsymbol{a} + \theta\boldsymbol{y};\boldsymbol{y})$.

Total derivative
Even if directional derivatives exist for every direction at $\boldsymbol{a}$, f may not be differentiable at $\boldsymbol{a}$.
A scalar field f is "differentiable at $\boldsymbol{a}$" if:
(1) There is a linear transformation $T_a : \boldsymbol{R}^n \to \boldsymbol{R}$
(2) There is a scalar function $E(\boldsymbol{a}, \boldsymbol{v})$
(3) $f(\boldsymbol{a} + \boldsymbol{v}) = f(\boldsymbol{a}) + T_a(\boldsymbol{v}) + \|\boldsymbol{v}\|\, E(\boldsymbol{a}, \boldsymbol{v})$
(4) For all $\|\boldsymbol{v}\| < r$
(5) Where $E(\boldsymbol{a}, \boldsymbol{v}) \to 0$ as $\|\boldsymbol{v}\| \to 0$
$T_a(y) = f'(\boldsymbol{a};\boldsymbol{y}) = \nabla f(\boldsymbol{a}) \cdot \boldsymbol{y}$.
$f'(\boldsymbol{a};\boldsymbol{y}) = \sum_{k=1}^{n} D_k f(\boldsymbol{a}) y_k$.

Gradient
$\nabla f(\boldsymbol{a}) = (D_1 f(\boldsymbol{a}), \ldots, D_n f(\boldsymbol{a}))$.

Continuity and differentiability of scalar fields
If the partial derivatives exist and are continuous at $\boldsymbol{a}$, then f is differentiable at $\boldsymbol{a}$.

Scalar field chain rule
If $g(t) = f[\boldsymbol{r}(t)]$, then $g'(t) = \nabla f(\boldsymbol{r}(t)) \cdot \boldsymbol{r}'(t)$.

Level sets
A level set is the subset of a domain of a function for which a scalar field has a contant value.
The gradient is perpendicular to the level set (level curves or level surfaces) / normal to the tangent plane.

Derivatives of vector fields
Derivatives of vector fields are the same as for scalar fields, extended to more dimensions.
Jacobian matrix: $m \times n$ matrix whose kth row is $\nabla f_k(\mathbf{a})$. Jacobian matrix: Written $Df(\mathbf{a})$.
$\mathbf{T_a(y) = Df(a)y}$.

Vector field chain rule
Chain rule: $\mathbf{h}'(\mathbf{a}) = \mathbf{f}'(\mathbf{g}(\mathbf{a})) \circ \mathbf{g}'(\mathbf{a})$ (Composition of linear transformations).
Matrix form: $D\mathbf{h}(\mathbf{a}) = \mathbf{Df}(\mathbf{g}(\mathbf{a}))\,\mathbf{Dg}(\mathbf{a})$.
This can also be written as a system of scalar equations:
If $f : R^n \to R^m$, $g : R^n \to R^p$, then $h = f \circ g : R^m \to R^p$.
There are then mp scalar equations.
$D_j h_i(\boldsymbol{a}) = \sum_{k=1}^{n} D_k f_i(\boldsymbol{g}(\boldsymbol{a})) D_j g_k(\boldsymbol{a})$.
For $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, p$.

Sufficient condition for equality of mixed partial derivatives
If $D_1 f, D_2 f, and D_{2,1} f$ exist, and if $D_{2,1}$ is continuous in a neighborhood of a
Then $D_{1,2} f$ exists and the two mixed partials are equal.

Jacobian determinant

$$\frac{\partial(f_1, \ldots, f_n)}{\partial(x_1, \ldots, x_n)} = \det \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$$

Maxima, minima, and saddle points
For a scalar field, a *stationary point* or *critical point* occurs where $\nabla f = O$. At stationary points, the tangent plane is horizontal.
Three types of stationary points: maxima, minima, and saddle points.
Extremum: a relative (including absolute) minimum or maximum.
Saddle point: every n-ball around $\boldsymbol{a}$ contains $f(\boldsymbol{x}) > f(\boldsymbol{a})$ and $f(\boldsymbol{x}) < f(\boldsymbol{a})$.
If f has continuous second-order partial derivatives on $B(\boldsymbol{a})$
And if $H(\boldsymbol{a})$ is Hessian matrix
And if $\boldsymbol{a}$ is a stationary point, then:
(a) If all the eigenvalues of $H(\boldsymbol{a})$ are positive, $\boldsymbol{a}$ is a relative minimum.
(b) If all the eigenvalues of $H(\boldsymbol{a})$ are negative, $\boldsymbol{a}$ is a relative maximum.
(c) If $H(\boldsymbol{a})$ has both positive and negative eigenvalues, $\boldsymbol{a}$ is a saddle point.
(d) If all eigenvalues are zero, this gives us no information.
There are other tests for stationary points, using the determinant of the Hessian matrix, etc.

Hessian matrix of a scalar field
$H(\boldsymbol{x}) = [D_{ij} f(\boldsymbol{x})]_{i,j=1}^{n}$.

Lagrange multipliers
Method for dealing with the following problem:
Find extreme values of a scalar field $f(\boldsymbol{x})$ when $\boldsymbol{x}$ is restricted to a subset of the domain.
Method:
(1) Scalar field $f(x_1, \ldots, x_n)$ subject to $m < n$ constraints $g_1(x_1, \ldots, x_n) = 0$, etc.
(2) At each extremum there exist m scalars $\lambda_1, \ldots, \lambda_m$.
(3) Such that $\nabla f = \lambda_1 \nabla g_1 + \cdots + \lambda_m \nabla g_m$.
(4) Solve the $n + m$ equations for $x_1, \ldots, x_n$ and $\lambda_1, \ldots, \lambda_m$.
(5) The required extrema will be among the solutions.

General Calculus

First fundamental theorem of calculus
If $A(x) = \int_c^x f(t)\,dt$ for $a \le x \le b$, and $a \le c \le b$, then $A'(x) = f(x)$.

Second fundamental theorem of calculus
If f is continuous on (a, b), and P is any primitive of f on (a, b), then for every c, $x \in (a, b)$, $P(x) = P(c) + \int_c^x f(t)\, dt$.

Definition of primitive / antiderivative
A function P is called a primitive (or an antiderivative) of a function f on an open interval I if the derivative of P is f, that is, if $P'(x) = f(x)$ for all x in I.

Functions, Limits, Continuity

Definition of convex and concave function
For all x and y in $[a, b]$ and for $0 < \alpha < 1$:
Convex: $g[\alpha y + (1 - \alpha)x] \leq \alpha g(y) + (1 - \alpha)g(x)$.
Concave: $g[\alpha y + (1 - \alpha)x] \geq \alpha g(y) + (1 - \alpha)g(x)$.

Limit of a function
$\lim_{x \to p} f(x) = A$ means that for every neighborhood $N_1(A)$ there is some neighborhood $N_2(p)$ such that $f(x) \in N_1(A)$ whenever $x \in N_2(p)$ and $x \neq p$.

Continuity of a function at a point
f is defined at p, and $\lim_{x \to p} f(x) = f(p)$.

Limit algebra
Let f and g be functions such that $\lim_{x \to p} f(x) = A, \quad \lim_{x \to p} g(x) = B$.
$\lim_{x \to p}[f(x) + g(x)] = A + B$
$\lim_{x \to p}[f(x) - g(x)] = A - B$
$\lim_{x \to p} f(x) \cdot g(x) = A \cdot B$
$\lim_{x \to p} f(x)/g(x) = A/B, \quad$ if $B \neq 0$.

Continuity algebra
If f and g are continuous at a point p, then $f + g$, $f - g$, $f \cdot g$ are also continuous at p.
If $g(p) \neq 0$, then f/g is continuous at p.
If v is continuous at p, and u is continuous at $q = v(p)$, then $f = u \circ v$ is continuous at p.

Squeezing principle
If $f(x) \leq g(x) \leq h(x)$ for all $x \neq p$ in a neighborhood of p, and $\lim_{x \to p} f(x) = \lim_{x \to p} h(x) = a$, then $\lim_{x \to p} g(x) = a$.

Continuity of indefinite integrals
If f is integrable on $[a, b]$, then the indefinite integral from a to x is continuous at every point of $[a, b]$. At each endpoint, it has one-sided continuity.

Intermediate value theorem
If f is continuous on $[a, b]$, then for any two points $x_1, x_2 \in [a, b]$, $f(x_1) \neq f(x_2)$, then f takes on every value between $f(x_1)$ and $f(x_2)$ in the interval (x_1, x_2).

Inverse functions
Let there be two functions f and g, with f having domain A and range B, and with g having domain B and range A. If the value of g at each point $y \in B$ is that unique x in A such that $f(x) = y$, then f and g are inverse functions. Also, if $c = f(a)$ and $d = f(b)$, and f is strictly increasing on $[a, b]$, then g is strictly increasing on $[c, d]$ and is continuous on $[a, b]$.
Left inverse: Given sets V and W and a function $T : V \to W$. A function S is called a left inverse of T if $S : T(V) \to V$ and $S[T(x)] = x$ for all x in V. Equivalently, $ST = I_V$, where I_V is the idenity transformation on V.

Right inverse: $TR = I_{T(V)}$.
Left inverses may or may not exist. If they exist, they are unique and are also right inverses.
Right inverses may not be unique.
A function has a left inverse iff it is one-to-one.
We call a left inverse *the* inverse, and denote it by T^{-1}.

Boundedness of continuous functions on any interval
If f is continuous on $[a, b]$, then f is bounded on $[a, b]$.

Extreme value theorem
If f is continuous on $[a, b]$, then there exist points c, $d \in [a, b]$ such that $f(c) = \sup f$ and $f(d) = \inf f$.

Small span theorem
If f is continuous on $[a, b]$, then for every $\epsilon > 0$ there is a finite partition of $[a, b]$ such that the span of f in every subinterval is less than ϵ.

Integrability of continuous functions
If f is continuous on $[a, b]$, then f is integrable on $[a, b]$.

Mean value theorem - integrals
If f is continuous on $[a, b]$, then for some c in $[a, b]$, $\int_a^b f(x)\,dx = f(c)(b - a)$.

Mean value theorem - second for integrals
If g is continuous on $[a, b]$ and if f has a derivative which never changes sign on $[a, b]$, then for some c in $[a, b]$, $\int_a^b f(x)g(x) = f(a)\int_a^c g(x)\,dx + f(b)\int_c^b g(x)\,dx$.

Mean value theorem - weighted integrals
Let f, g be continuous on $[a, b]$, and let g never change sign in $[a, b]$. Then for some c in $[a, b]$, $\int_a^b f(x)g(x)\,dx = f(c)\int_a^b g(x)\,dx$.

Rational functions
Definition: Quotient of two polynomials.
Proper RF: The degree of the numerator is less than that of the denominator.
Every RF has an integral expressible in terms of polynomials, RFs, inverse tangents, and logarithms.
Every proper RF can be expressed as a finite sum of fractions of the forms:
$\frac{A}{(x+a)^k}$ and $\frac{Bx+C}{(x^2+bx+c)^m}$, where A, B, C, a, b, c are constants, k, m are positive integers, and $b^2 - 4c < 0$.
When an RF has been expressed in this form, it is decomposed into partial fractions.

Elementary functions
Can be obtained from polynomials, exponentials, logarithms, trigonometric, or inverse trigonometric functions in a finite number of steps by using addition, subtraction, multiplication, division, and composition.

Classic non-elementary functions
$\int_0^x e^{-t^2}\,dt \quad \int_0^x \frac{\sin t}{t}\,dt \quad \int_0^x \sin(t^2)\,dt \quad \int_0^x \sqrt{1 - k^2 \sin^2 t}\,dt$

Taylor / o-notation method for determining limits
Can replace indeterminate limit with Taylor + o-notation representation, which may then be easier to find.

Classic limits
$\lim_{x\to 0} \frac{a^x - b^x}{x} = \log \frac{a}{b} \quad \lim_{x\to 0} \frac{\log(1+ax)}{x} = a \quad \lim_{x\to 0}(1 + ax)^{1/x} = e^a$
$\lim_{x\to +\infty} \frac{(\log x)^b}{x^a} = 0$ for $a > 0, b > 0$
$\lim_{x\to +\infty} \frac{x^b}{e^{ax}} = 0$ for $a > 0, b > 0$.

L'Hopital's rule
If $\lim_{x\to a} f(x) = 0$ and $\lim_{x\to a} g(x) = 0$ and $g'(x) \neq 0$, then $lim_{x\to a} \frac{f(x)}{g(x)} = \lim_{x\to a} \frac{f'(x)}{g'(x)}$, if this second limit exists.

Limits of vector fields
For $\boldsymbol{f} : R^n \to R^m$,
$\lim_{\boldsymbol{x}\to\boldsymbol{a}} \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{b}$
means that $\lim_{\|\boldsymbol{x}-\boldsymbol{a}\|\to 0} \|\boldsymbol{f}(\boldsymbol{x}) - \boldsymbol{b}\| = 0$
or, $\lim_{\|\boldsymbol{h}\|\to 0} \|\boldsymbol{f}(\boldsymbol{a}+\boldsymbol{h}) - b\| = 0$.

Continuity of vector fields

$$\lim_{\boldsymbol{x}\to\boldsymbol{a}} \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{a})$$

o-notation

Definition
Near a, $f(x)$ is small compared with $g(x)$. More formally:
$f(x) = o(g(x))$ as $x \to a$ means that $\lim_{x\to a} \frac{f(x)}{g(x)} = 0$ (assuming $g(x) \neq 0$ near a).

Algebra
$o(g(x)) \pm o(g(x)) = o(g(x))$
$o(cg(x)) = o(g(x))$ if $c \neq 0$
$f(x) \cdot o(g(x)) = o(f(x)g(x))$
$o(o(g(x))) = o(g(x))$

Trigonometric functions

Fundamental properties of sine and cosine
Defined on entire real line.
$\cos 0 = \sin \frac{1}{2}\pi = 1, \cos \pi = -1$.
$\cos(y - x) = \cos y \cos x + \sin y \sin x$.
For $0 < x < \frac{1}{2}\pi$, $0 < \cos x < \frac{\sin x}{x} < \frac{1}{\cos x}$.

Additional properties of sine and cosine
$\sin^2 x + \cos^2 x = 1$.
$\sin 0 = \cos \frac{1}{2}\pi = \sin \pi = 0$.
$\cos(-x) = \cos(x)$, $\sin(-x) = -\sin(x)$.
$\sin(\frac{1}{2}\pi + x) = \cos x$, $\cos(\frac{1}{2}\pi + x) = -\sin(x)$.
$\sin(x + 2\pi) = \sin x$, $\cos(x + 2\pi) = \cos x$.
$\cos(x + y) = \cos x \cos y - \sin x \sin y$, $\sin(x + y) = \sin x \cos y + \cos x \sin y$.
$\sin a - \sin b = 2 \sin \frac{a-b}{2} \cos \frac{a+b}{2}$, $\cos a - \cos b = -2 \sin \frac{a-b}{2} \sin \frac{a+b}{2}$.
In the interval $[0, \frac{1}{2}\pi]$, the sine is strictly increasing, and the cosine is strictly decreasing.

Additional sine and cosine relationships
If $0 < a \leq \frac{1}{2}\pi$ and $n \geq 1$, $\frac{a}{n} \sum_{k=1}^{n} \cos \frac{ka}{n} < \sin a < \frac{a}{n} \sum_{k=0}^{n-1} \cos \frac{ka}{n}$.

Integrals
$\int_0^a \cos x \, dx = \sin a$
$\int_0^a \sin x \, dx = 1 - \cos a$.

Other trigonometric functions
$\tan x = \frac{\sin x}{\cos x}$ $\quad \cot x = \frac{\cos x}{\sin x}$ $\quad \sec x = \frac{1}{\cos x}$ $\quad \csc x = \frac{1}{\sin x}$

Inverse trigonometric functions
$u = \arcsin v$ means $v = \sin u$, $\quad u = \arccos v$ means $v = \cos u$, $\quad u = \arctan v$ means $v = \tan u$
$\operatorname{arccot} x = \frac{\pi}{2} - \arctan x$ for all real x, $\quad \operatorname{arcsec} x = \arccos \frac{1}{x}$ for $|x| \geq 1$, $\quad \operatorname{arccsc} x = arcsin \frac{1}{x}$ for $|x| \geq 1$

Inverse trigonometric derivatives and integrals
$D \arcsin x = \frac{1}{\sqrt{1-x^2}}$ if $-1 < x < 1$
$\int_0^x \frac{1}{\sqrt{1-t^2}}\, dt = \arcsin x$
$\int \arcsin x\, dx = x \arcsin x + \sqrt{1-x^2} + C$
$D \arccos x = \frac{-1}{\sqrt{1-x^2}}$
$D \arctan x = \frac{1}{1+x^2}$
$\int \frac{dx}{\sqrt{1-x^2}} = -\arccos x + C$
$\int \frac{dx}{1+x^2} = \arctan x + C$
$\int \arccos x\, dx = x \arccos x - \sqrt{1-x^2} + C$
$\int \arctan x\, dx = x \arctan x - \frac{1}{2} \log(1+x^2) + C$

Exponential, logarithm, hyperbolic functions

Logarithm definition
$L(x) = \int_1^x \frac{1}{t}\, dt$

Exponential definition
$y = E(x)$ means $L(y) = x$

Logarithm properties
$L(1) = 0$
$L'(x) = \frac{1}{x}$ for $x > 0$
$L(ab) = L(a) + L(b)$ for $a > 0$, $b > 0$

Exponential properties
$E(0) = 1$
$E(1) = e$
$E'(x) = E(x)$
$E(a+b) = E(a)E(b)$
$a^x = e^{x \log a}$, for $a > 0$, x real

Derivatives and integrals
$(e^x)' = e^x$
$(e^{x \log a})' = a^x \log a$
$\int e^x\, dx = e^x + C$
$\int a^x\, dx = \frac{a^x}{\log a} + C$
$\int e^{f(x)} f'(x)\, dx = e^{f(x)} + C$
$\int a^{f(x)} f'(x)\, dx = \frac{a^{f(x)}}{\log a} + C$

Logarithmic differentiation
If $g(x) = \log|f(x)|$, then $g'(x) = \frac{f'(x)}{f(x)}$. In cases where $f'(x)$ is difficult to compute, $(\log|f(x)|)'$ may be easier to compute. Then we know $f(x)$ and $g'(x)$, which we can multiply to give $f'(x)$.

Hyperbolic definitions
$\sinh x = \frac{e^x - e^{-x}}{2}$, $\cosh x = \frac{e^x + e^{-x}}{2}$, $\tanh x = \frac{\sinh x}{\cosh x} = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
$\operatorname{csch} x = \frac{1}{\sinh x}$, $\operatorname{sech} x = \frac{1}{\cosh x}$, $\coth x = \frac{1}{\tanh x}$

Hyperbolic properties
$\cosh^2 x - \sinh^2 x = 1$
$\sinh(-x) == \sinh x$
$\cosh(-x) = \cosh x$
$\tanh(-x) = -\tanh x$
$\sinh(x+y) = \sinh x \cosh y + \sinh x \sinh y$
$\cosh(x+y) = \cosh x \cosh y + \sinh x \sinh y$
$\sinh 2x = 2 \sinh x \cosh x$
$\cosh 2x = \cosh^2 x + \sinh^2 x$
$\cosh x + \sinh x = e^x$
$\cosh x - \sinh x = e^{-x}$
$(\cosh x + \sinh x)^n = \cosh nx + \sinh nx$, where n is an integer
$2 \sinh^2 \frac{1}{2} x = \cosh x - 1$
$2 \cosh^2 \frac{1}{2} x = \cosh x + 1$
$\tanh^2 x + \operatorname{sech}^2 x = 1$
$\coth^2 x - \operatorname{csch}^2 x = 1$

Taylor polynomials

Definition
$T_n f(x) = \sum_{k=0}^{n} \frac{f^{(k)}(a)}{k!} (x-a)^k$

Remainder / error
$E_n(x) = \frac{1}{n!} \int_a^x (x-t)^n f^{(n+1)}(t)\, dt$

Error o-notation
$f(x) = T_n f(x) + o((x-a)^n) \quad \text{as} \quad x \to a$

Error estimate
If $m \le f^{(n+1)}(t) \le M$, then:
If $x > a$: $m \frac{(x-a)^{n+1}}{(n+1)!} \le E_n(x) \le M \frac{(x-a)^{n+1}}{(n+1)!}$
If $x < a$: $m \frac{(a-x)^{n+1}}{(n+1)!} \le (-a)^{n+1} E_n(x) \le M \frac{(a-x)^{n+1}}{(n+1)!}$

Algebra
Linearity: $T_n(c_1 f + c_2 g) = c_1 T_n(f) + c_2 T_n(g)$
Differentiation: $(T_n f)' = T_{n-1}(f')$
Substitution: If $g(x) = f(cx)$, then $T_n g(x; a) = T_n f(cx; ca)$
Integration: An indefinite integral of a Taylor polynomial of f is a Taylor polynomial of an indefinite integral of f.
Integration, precisely: If $g(x) = \int_a^x f(t)\, dt$, then $T_{n+1} g(x) = \int_a^x T_n f(t)\, dt$.

Linear Differential Equations

First order
$\{f'(x) = f(x),\ f(0) = C\} \quad \to \quad f(x) = Ce^x$
$\{y' + P(x)y = 0,\ f(a) = b\} \quad \to \quad f(x) = be^{-A(x)},\ A(x) = \int_a^x P(t)\, dt$
$\{y' + P(x)y = Q(x),\ f(a) = b\} \quad \to \quad f(x) = be^{-A(x)} + e^{-A(x)} \int_a^x Q(t) e^{A(t)}\, dt$

Second order - homogeneous, constant coefficients
Form: $y'' + ay' + by = 0$
Let $d = a^2 - 4b$. Every solution has the form:
$y = e^{-ax/2}[c_1 u_1(x) + c_2 u_2(x)]$ where c_1 and c_2 are constants, and u_1 and u_2:
If $d = 0$, then $u_1(x) = 1$ and $u_2(x) = x$
If $d > 0$, then $u_1(x) = e^{kx}$ and $u_2(x) = e^{-kx}$ where $k = \frac{1}{2}\sqrt{d}$
If $d < 0$, then $u_1(x) = \cos kx$ and $u_2(x) = \sin kx$ where $k = \frac{1}{2}\sqrt{-d}$

Second order - nonhomogeneous, constant coeffcients
Form: $y'' + ay' + by = R$
Let $v_1(x)$ and $v_2(x)$ be a basis of solutions to $L(y) = 0$ given by $v_1(x) = e^{-ax/2}u_1(x)$ and $v_2(x) = e^{-ax/2}u_2(x)$.
Let $W(x) = v_1(x)v_2'(x) - v_2(x)v_1'(x)$. (Wronskian of v_1 and v_2)
Then $L(y) = R$ has a particular solution given by $y_1(x) = t_1(x)v_1(x) + t_2(x)v_2(x)$
$t_1(x) = -\int v_2(x)\frac{R(x)}{W(x)}\,dx$, $t_2(x) = \int v_1(x)\frac{R(x)}{W(x)}\,dx$
The general solution is given by $y = c_1 v_1 + c_2 v_2 + y_1$.

Second order - nonhomogeneous, constant coefficients - special methods
R is a polynomial (of degree n) and $b \neq 0$:
(1) Take $Ax^n + Bx^{n-1} + \cdots + D$ and differentiate twice
(2) Plug into the differential equation, and equate coeffs of like powers to find y_1.
$R(x) = p(x)e^{mx}$, $p(x)$ is polynomial of degree n, and m is constant.
(1) Use the substitution $y = u(x)e^{mx}$
(2) This transforms the diff eq into $u'' + (2m + a)u' + (m^2 + am + b)u = p$, where p is a polynomial
(3) This can be solved as in the previous case

Order n
Form: $P_0(x)y^{(n)} + P_1(x)y^{(n-1)} + \cdots + P_n(x)y = R(x)$
Singular point: where $P_0(x) = 0$ (often introduce complications)
Form ($P_0(x) \neq 0$): $y^{(n)} + P_1(x)y^{(n-1)} + \cdots + P_n(x)y = R(x)$
Operator notation: $L(f) = f^{(n)} + P_1 f^{(n-1)} + \cdots + P_n f$
Operator notation (alternate): $L = D^n + P_1 D^{n-1} + \cdots + P_n$
L is a linear differential operator of order n
Homogeneous: $L(y) = 0$
Nonhomogeneous: $L(y) = R$
Solution of the homogeneous equation is the null space of L, and $\dim N(L) = n$
Homogeneous existence-uniqueness: there is one and only one solution satisfying an order n eqn and initial conditions
For n independent solutions, all linear combinations of these solutions is the general solution
General solution of nonhomogeneous is one particular solution of nonhomogeneous, plus linear combinations of homogeneous

Order n - constant coefficients
Constant-coefficient operator: $A = a_0 D^n + a_1 D^{n-1} + \cdots + a_{n-1}D + a_n$
The set of all constant-coefficient operators is a linear space
Constant-coefficient operators commute, since $D^r D^s = D^s D^r$
Characteristic polynomial of A: $p_A(r) = a_0 r^n + a_1 r^{n-1} + \cdots + a_n$
Let A and B be constant-coefficient operators and λ real:
(a) $A = B$ iff $p_A = p_B$
(b) $p_{A+B} = p_A + p_B$
(c) $p_{AB} = p_A \cdot p_B$
(d) $p_{\lambda A} = \lambda \cdot p_A$
Factorizations of a characteristic polynomial are also factorizations of the operators
If you factor an operator, the solution space of the operator contains the solution space of each factor
Annihilation: If an operator maps an element to O, it is said to *annihilate* that element
If a factor annihilates u, then L annihilates u as well
Solutions:
(a) $p_L(r) = 0$ has real distinct roots: $y = \sum_{k=1}^{n} c_k e^{r_k x}$

(b) Real roots, some repeated: $u_1(x) = e^{rx}, u_2(x) = xe^{rx}, u_3 = x^2e^{rx}$ etc.
(c) Complex roots $\alpha + i\beta$: $u_1(x) = e^{\alpha x}\cos\beta x$, $u_2(x) = e^{\alpha x}\sin\beta x$
(d) Repeated complex roots: Add x^{m-1} to each solution

Determining a particular solution of a nonhomogeneous equation - Variation of parameters
Want to determine one solution of $L(y) = R$
Let $u_1, \ldots, u_n$ be n independent solutions of the homogeneous equation $L(y) = 0$
Define an $n \times 1$ column matrix v as:

$$v(x) = \int_c^x R(t)W(t)^{-1} \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix} dt$$

where W is the Wronskian matrix of $u_1, \ldots, u_n$ and c is any point in J.
Then $y_1(x) = \sum_{k=1}^n u_k(x)v_k(x)$
The Wronskian matrix of n independent solutions is always nonsingular.

Determining a particular solution of a nonhomogeneous equation - Reduction to first-order system
Works if: constant coefficients
Make a substitution, say $u = (D-2)y$
This reduces the order of the equation by one.
When you find the solution to one equation, substitute it back into other equations.

Determining a particular solution of a nonhomogeneous equation - Annihilator method
Works if:
Constant coefficients and R is annihilated by a constant-coefficient differential operator
Steps:
(1) Apply the annihilator operator to both sides, producing a homogeneous constant-coefficient equation.
(2) Find the general solution to this equation. The particular solution is one of these.
(3) Plug the general solution with undetermined coefficients back into $L(y) = R$.
(4) Equate coefficients of like terms to determine the coefficients of this particular solution.
Functions and corresponding annihilators:

$y = x^{m-1}$	D^m
$y = e^{\alpha x}$	$D - \alpha$
$y = x^{m-1}e^{\alpha x}$	$(D-\alpha)^m$
$y = \cos\beta x$ or $y = \sin\beta x$	$D^2 + \beta^2$
$y = x^{m-1}\cos\beta x$ or $y = x^{m-1}\sin\beta x$	$(D^2+\beta^2)^m$
$y = e^{\alpha x}\cos\beta x$ or $y = e^{\alpha x}\sin\beta x$	$D^2 - 2\alpha D + (\alpha^2+\beta^2)$
$y = x^{m-1}e^{\alpha x}\cos\beta x$ or $y = x^{m-1}e^{\alpha x}\sin\beta x$	$[D^2 - 2\alpha D + (\alpha^2+\beta^2)]^m$

Second order - analytic coefficients
A function is *analytic* if it has a power series expansion in an interval.
For $y'' + P_1(x)y' + P_2(x)y$, if P_1 and P_2 are analytic, then the equation has two independent analytic solutions.
More generally, a homogeneous linear equation of order n with analytic coefficients has n independent analytic solutions.
Finding power series solutions:
(1) Write a general possible solution, $y = \sum_{n=0}^\infty a_n(x-x_0)^n$
(2) Differentiate this possible solution twice to find y' and y''
(3) Plug these into the differential equation
(4) Determine what the coefficients a_n would have to look like (recursion formula) to satisfy the equation
(5) Verify the determined power series converges in the given interval
(6) Determine specific coefficients to satisfy initial conditions or to find independent solutions

Legendre equation and polynomials
Legendre equation: $(1-x^2)y'' - 2xy' + \alpha(\alpha+1)y = 0$, where α is any real constant.

Legendre equation (Sturm-Liouville operator): $[(x^2-1)y']' = \alpha(\alpha+1)y$.
$\rightarrow$ So $T(y) = \lambda y$, where $T(f) = (pf')'$ and $p(x) = x^2 - 1$ and $\lambda = \alpha(\alpha+1)$.
$\rightarrow$ So the nonzero solutions are eigenfunctions belonging to $\alpha(\alpha+1)$.
General solution on $(-1, 1)$:

$$y = a_0 u_1(x) + a_1 u_2(x)$$

$$u_1(x) = 1 + \sum_{n=1}^{\infty} (-1)^n \frac{\alpha(\alpha-2)\cdots(\alpha-2n+2)\cdot(\alpha+1)(\alpha+3)\cdot(\alpha+2n-1)}{(2n)!} x^{2n}$$

$$u_2(x) = x + \sum_{n=1}^{\infty} (-1)^n \frac{(\alpha-1)(\alpha-3)\cdots(\alpha-2n+1)\cdot(\alpha+2)(\alpha+4)\cdots(\alpha+2n)}{(2n+1)!} x^{2n+1}$$

When α is zero or a positive integer, the equation has polynomial solutions called Legendre polynomials:

$$P_n(x) = \frac{1}{2^n} \sum_{r=0}^{[n/2]} \frac{(-1)^r (2n-2r)!}{r!(n-r)!(n-2r)!} x^{n-2r}$$

where $[n/2]$ is the greatest integer $\leq n/2$.
These are the same (except for scalar factors) as the result of the Gram-Schmidt process on $\{1, x, x^2, x^3, \dots\}$.
$\rightarrow$ So, they are special in that they are orthogonal in Euclidean space with inner product $\int_{-1}^{1} fg\,dx$.
Rodrigues' formula for the Legendre polynomials: $P_n(x) = \frac{1}{2^n n!} \frac{d^n}{dx^n}(x^2-1)^n$.
$P_n(-x) = (-1)^n P_n(x)$, so P_n is even when n is even and odd when n is odd.

Method of Frobenius
Works if:
(1) Differential equation has the form: $(x-x_0)^2 y'' + (x-x_0)P(x)y' + Q(x)y = 0$.
(2) P and Q are analytic in some open interval about x_0.
(Note) We say x_0 is a *regular* singular point of the equation.
Preliminaries:
Indicial equation of the differential equation: $t(t-1) + P(x_0)t + Q(x_0) = 0$.
Denote the two roots of the indicial equation by α_1 and α_2.
First case: $\alpha_1 - \alpha_2$ is not an integer
There are two independent solutions, u_1 and u_2.
$u_1(x) = |x-x_0|^{\alpha_1} \sum_{n=0}^{\infty} a_n (x-x_0)^n$, with $a_0 = 1$.
$u_2(x) = |x-x_0|^{\alpha_2} \sum_{n=0}^{\infty} b_n (x-x_0)^n$, with $b_0 = 1$.
Second case: $\alpha_1 - \alpha_2 = N$, a nonnegative integer
u_1 is the same as in the first case.
$u_2 = |x-x_0|^{\alpha_2} \sum_{n=0}^{\infty} b_n (x-x_0)^n + C u_1(x) \log|x-x_0|$, with $b_0 = 1$.
$\rightarrow$ C is nonzero if $N = 0$; otherwise it may or may not be zero.

Gamma function
For complex z with real part > 0:
$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t}\,dt$
It can also be extended to be defined for the rest of the complex plane except for negative integers.

Bessel equation and functions
Bessel equation: $x^2 y'' + xy' + (x^2 - \alpha^2)y = 0$, where $\alpha \geq 0$
Bessel function of the first kind of order α:
$J_a(x) = \left(\frac{x}{2}\right)^{\alpha} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\Gamma(n+1+\alpha)} \left(\frac{x}{2}\right)^{2n}$
Valid for $x > 0$
Bessel function of the first kind when $\alpha = p$ is a nonnegative integer:
$J_p(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!(n+p)!} \left(\frac{x}{2}\right)^{2n+p}$
Valid for $x \neq 0$
Bessel function valid if α is not zero or a positive integer:
$J_{-\alpha}(x) = \left(\frac{x}{2}\right)^{-\alpha} \sum_{n=0}^{\infty} \frac{(-1)^n}{n!\Gamma(n+1-\alpha)} \left(\frac{x}{2}\right)^{2n}$

Valid for $x > 0$
If α is not an integer, $J_\alpha(x)$ and $J_{-\alpha}(x)$ are linearly independent and general solution is:
$y = c_1 J_\alpha(x) + c_2 J_{-\alpha}(x)$
Valid for $x > 0$
Bessel function of the second kind of order p:
$Y_p(x) = J_p(x)\log x + x^{-p}\sum_{n=0}^{\infty} C_n x^n$
With last term determined: $Y_p(x) = J_p(x)\log x - \frac{1}{2}\left(\frac{x}{2}\right)^{-p}\sum_{n=0}^{p-1}\frac{(p-n-1)!}{n!}\left(\frac{x}{2}\right)^{2n} - \frac{1}{2}\left(\frac{x}{2}\right)^{p}\sum_{n=0}^{\infty}(-1)^n\frac{h_n+h_{n+p}}{n!(n+p)!}\left(\frac{x}{2}\right)^{2n}$
Where $h_0 = 0$ and $h_n = 1 + \frac{1}{2} + \cdots + 1/n$ for $n \geq 1$
Valid when $\alpha = p$ is a nonnegative integer and $x > 0$
If α is a nonnegative integer, general solution is:
$y = c_1 J_p(x) + c_2 Y_p(x)$
Valid for $x > 0$

Systems of differential equations

Concepts
Any order differential equation can be rewritten as a system of first-order equations.
A general linear system has the form $Y' = P(t)Y + Q(t)$
...where Y' is a column matrix, $P(t)$ is a coefficient matrix, and $Q(t)$ is a column matrix.

Basic exponential differential equation
If $F'(t) = F(t)A$ or $F'(t) = FE(t)$, then $F(t) = e^{tA}$.
$F'(t) = AF(t)$ and $F(0) = B \quad\rightarrow\quad F(t) = e^{tA}B$.
$F'(t) = F(t)A$ and $F(0) = B \quad\rightarrow\quad F(t) = Be^{tA}$.

Exponential matrices
$e^O = I$

$$e^A = \sum_{k=0}^{\infty}\frac{A^k}{k!}$$

A series of matrices is convergent if the series for each element is convergent.

Matrix norm: $\|A\| = \sum_{i=1}^{m}\sum_{j=1}^{n}|a_{ij}|$

If $\sum_{k=1}^{\infty}\|C_k\|$ converges, then $\sum_{k=1}^{\infty} C_k$ also converges.
In general two matrices commute under multiplication if they are both diagonalizable
...but there are other circumstances where they commute as well, for example λI commutes with any matrix.
e^{tA} is always nonsingular, and its inverse is e^{-tA}.
If A and B commute, then $e^{A+B} = e^A + e^B$.

Homogeneous linear systems with constant coefficients
For $Y'(t) = AY(t)$, $Y(a) = B \quad\rightarrow\quad Y(t) = e^{(t-a)A}B$.

Calculating e^{tA}
Diagonal matrix: If $A = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, then $e^{tA} = \mathrm{diag}(e^{t\lambda_1}, \ldots, e^{t\lambda_n})$.
If A can be diagonalized ($A = CDC^{-1}$): $e^{tA} = Ce^{tD}C^{-1}$.
Putzer's method: can be used for any $n \times n$ matrix.
There are other special methods for matrices where eigenvalues are identical, distinct, or of multiplicity 1 and $n-1$.

Nonhomogeneous systems with constant coefficients
Form: $Y'(t) = AY(t) + Q(t)$, $Y(a) = B$.
Solution: $Y(x) = e^{(x-a)A}B + e^{xa}\int_a^x e^{-tA}Q(t)\,dt$.

Cayley-Hamilton theorem
An $n \times n$ matrix satisfies its own characteristic equation.
For an $n \times n$ matrix A,
...characteristic polynomial is $f(\lambda) = \det(\lambda I - A) = \lambda^n + c_{n-1}\lambda^{n-1} + \cdots + c_1\lambda + c_0$.
...Then $f(A) = O$
...Or, $A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 I = O$.

Homogeneous linear systems
If $A(t)$ is analytic, you can attempt to find a power series solution.

General linear system
Form:
$Y'(x) = P(x)Y(x) + Q(x)$, $Y(a) = B$
There is one and only one solution to this initial value problem.
Solution:
$Y(x) = F(x)^{-1}Y(a) + F(x)^{-1}\int_a^x F(t)Q(t)\,dt$
$F(x)$ is the transpose of the matrix whose kth column is the solution of the initial-value problem:
$Y'(x) = -P(x)^t Y(x)$, $Y(a) = I_k$, where I_k is the kth column of I.
Note: this is not always a useful formula since determining $F(x)$ can be difficult.

Method of successive approximations
Basically, start with something, apply an operation, plug that back into the original, apply the operation again, etc.
Can be used in an existence theorem for homogeneous linear systems $Y'(t) = A(t)Y(t)$, $Y(a) = B$.
Process:
(1) Start with $Y(x) = B$.
(2) Plug into equation, $Y'(t) = A(t)B$.
(3) Solve to get $Y_1(x) = B + \int_a^x A(t)B\,dt$.
(4) Plug $Y_1(x)$ back into original equation, repeat.
(5) Come up with a recursive formula describing all terms, determining a series.
(6) Prove that the series converges on the required interval.
Note: In practice, it is often very difficult to compute the successive approximations,
...so this method is often more useful for existence proofs than solving problems.

Nonlinear Differential Equations

First order - separable
If $y' = Q(x)R(y)$, this can be written as $A(y)y' = Q(x)$. Then, if $y = Y(x)$ is a solution, and if G is any primitive of A, then we have an implicit formula for y: $G(y) = \int Q(x)\,dx + C$, for some C.

First order - homogeneous
For $y' = f(x, y)$, *homogeneous* means that $f(tx, ty) = f(x, y)$ for all $x, y, t \neq 0$. In this case, we can use the substitution $v = y/x$, which gives us $x\frac{dv}{dx} = f(1, v) - v$, which is a first-order separable equation for v.

Using power series to solve
Technique: Assume there is a power series expansion of the solution about a point (say 0). Then, differentiate the power series sum function as needed, and plug into the original equation, equate coefficients of like terms, etc. Can also make use of the uniqueness theorem for power series which says that two power series with the same sum function about the same point must be equal, term by term. Then, I think you verify convergence of the resulting series. Normally, will have undetermined coefficients, which gives a general class of solutions to the diff eq. Can also use the fact the $a_n = \frac{f^{(n)}(0)}{n!}$, if the expansion is about 0.

Nonlinear systems
Form:
$Y'(x) = F(x, Y(x))$, $Y(a) = B$
Lipschitz condition:
For a vector mapping $F(x, Y)$,
$\|F(x, Y) - F(x, Z)\| \leq A\|Y - Z\|$
where A is a positive constant.
If $\|F(x, Y)\| \leq M$ where M is a positive constant,
...and if $G(x) = F(x, Y(x))$ is continuous on $(a - h, a + h)$,
...and if F satisfies the Lipschitz condition
...and let S be a k-cell $\{(x, Y)| \quad |x - a| \leq h, \|Y - B\| \leq k\}$
...where h and k are positive constants
...then there is one and only one function defined on $(a - c, a + c)$ with $c = \min\{h, k/M\}$
...which satisfies the nonlinear system given above.

Partial Differential Equations

General concepts
It often happens that in solving a PDE, an equation of one variable is introduced
that can be anything - and so the solution space is often infinite-dimensional.

Fixed points and contraction operators

Fixed point of an operator
If $T(Y) = Y$, then Y is a *fixed point* of T.

Norms
A function in a linear space such that:
(1) $\|x\| \geq 0$
(2) $\|cx\| = |c|\|x\|$
(3) $\|x + y\| \leq \|x\| + \|y\|$
(4) $\|x\| = 0$ implies $x = O$ A linear space with a norm defined is called a *normed linear space.*
The max norm:
For a function $\varphi \in C(J)$, the max norm is $\|\varphi\| = \max_{x \in J} |\varphi(x)|$.

Contraction operators
An operator is a *contraction operator* if:
$T : C(J) \to C(J)$
$\|T(\varphi) - T(\psi)\| \leq \alpha\|\varphi - \psi\|$
with contraction constant α such that $0 \leq \alpha < 1$.
Every contraction operator has one and only one fixed point such that $T(\varphi) = \varphi$.

Sequences and series

Definitions
Infinite sequence: a function whose domain is the set of all positive integers.
Sequence convergence: can be for real or complex sequences. In the complex case, each component must converge separately.
Series or infinite series: a sequence $\{s_n\}$ of partial sums of an infinite sequence such that $s_n = \sum_{k=1}^{n} a_k$.
If $\lim_{n \to \infty} s_n = S$, for some finite S, then the series converges and S is its sum.

Properties
A monotonic sequence converges iff it is bounded.
Linearity: $\sum_{n=1}^{\infty}(\alpha a_n + \beta b_n) = \alpha \sum_{n=1}^{\infty} a_n + \sum_{n=1}^{\infty} b_n$.
If $\sum a_n$ conveges and $\sum b_n$ diverges, then $\sum(a_n + b_n)$ diverges.
If $\{a_n\}$ and $\{b_n\}$ are two sequences of complex numbers st $a_n = b_n - b_{n+1}$, then $\sum a_n$ converges iff $\{b_n\}$ converges, in which case $\sum_{n=1}^{\infty} a_n = b_1 - \lim_{n\to\infty} b_n$.
For complex x, if $|x| < 1$, $\sum_{n=0}^{\infty} x^n$ converges and equals $\frac{1}{1-x}$. Otherwise, it diverges.
Finite sums can be rearranged without affecting the sum. This is not necessarily true for series.
Absolutely convergent series can be rearranged without affecting the sum.
If $\sum a_n$ is a real conditionally convergent series, then for any real number S, there is a rearrangement $\sum b_n$ of $\sum a_n$ which converges to S.

Convergence and divergence
If $\sum a_n$ converges, then $\lim_{n\to\infty} a_n = 0$.
If all a_n are nonnegative, then $\sum a_n$ converges iff $\{a_n\}$ is bounded above.
If all a_n and b_n are nonnegative, and if there is some c st $a_n \leq cb_n$ for all n, then convergence of $\sum b_n$ implies convergence of $\sum a_n$.
If $\{a_n\}$ is a monotonic decreasing sequence with limit 0, then $\sum_{n=1}^{\infty}(-1)^{n-1}a_n$ converges. Also, $0 < (-1)^n(S - s_n) < a_{n+1}$, for $n \geq 1$.
If $\sum |a_n|$ converges, then $\sum a_n$ also converges, and $|\sum_{n=1}^{\infty} a_n| \leq \sum_{n=1}^{\infty} |a_n|$.

Series Tests
Comparison test:
  If a_n and b_n nonnegative, and $a_n \leq cb_n$ for some positive constant c
  Then convergence of $\sum b_n$ implies convergence of $\sum a_n$.
  When the inequality is satisfied, $\sum b_n$ *dominates* $\sum a_n$.
Limit comparison test:
  If all a_n and b_n are positive
  And if $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$ (asymptotically equal)
  Then $\sum a_n$ and $\sum b_n$ converge or diverge together.
Integral test:
  If $f(x)$ is a positive decreasing fn defined on $x \geq 1$
  And if $s_n = \sum_{k=1}^{n} f(k)$ and $t_n = \int_1^n f(x)\,dx$
  Then $\{s_n\}$ and $\{t_n\}$ converge or diverge together.
Root test:
  a_n nonnegative and $\lim_{n\to\infty} \sqrt[n]{a_n} = R$
  $R < 1$: Converges
  $R > 1$: Diverges
  $R = 1 or limit does n' texist$: Inconclusive
Ratio test:
  a_n positive and $\frac{a_{n+1}}{a_n} \to L$ as $n \to \infty$
  $L < 1$: Converges
  $L > 1$: Diverges
  $L = 1 or limit does n' texist$: Inconclusive
Dirichlet's test:
  If the partial sums of $\sum a_n$ form a bounded sequence
  And if $\{b_n\}$ is a decreasing sequence which converges to 0
  Then $\sum a_n b_n$ converges.
Abel's test:
  If $\sum a_n$ is a convergent complex series
  And if $\{b_n\}$ is a monotonic convergent real sequence
  Then $\sum a_n b_n$ converges.

Converging series
$\sum \frac{1}{n^2+n)}$

$\sum \frac{1}{n^2}$
$\sum \frac{1}{n^s}$, for $s > 1$

Diverging series
$\sum \frac{1}{n}$

Miscellaneous
Harmonic series: $\sum_{k=1}^{n} \frac{1}{k}$
Geometric series: $\sum_{k=0}^{n} x^k$
Power series: $\sum_{n=0}^{\infty} a_n x^n$ or $\sum_{n=0}^{\infty} a_n (z-a)^n$
Riemann zeta-function: $\zeta(s) = \sum_{n=1}^{\infty} \frac{1}{n^s}$, if $s > 1$

Abel's partial summation formula
Let $\{a_n\}$ and $\{b_n\}$ be sequences of complex numbers, and let $A_n = \sum_{k=1}^{n} a_k$. Then, $\sum_{k=1}^{n} a_k b_k = A_n b_{n+1} + \sum_{k=1}^{n} A_k (b_k - b_{k+1})$.

Sequences and series of functions

General
In general, the limit of integrals is not equal to the integral of the limit.
For a *uniformly convergent* sequence, limit and integration can be interchanged.

Pointwise vs. Uniform convergence
Pointwise vs. uniform convergence
If each function f_n is continuous at a point p, and if $f_n \to f$ uniformly, then f is also continuous at p.
If a series of functions converges uniformly, and if each term is continuous at a point, then the sum is also continuous at that point.
If $f \to f_n$ uniformly on an interval, then $\lim_{n\to\infty} \int_a^x f_n(t)\,dt = \int_a^x \lim_{n\to\infty} f_n(t)\,dt$.
If $\sum u_k$ converges uniformly, then $\sum_{k=1}^{\infty} \int_a^x u_k(t)\,dt = \int_a^x \sum_{k=1}^{\infty} u_k(t)\,dt$.

Tests for uniform convergence
Weierstrass M-Test: If $\sum u_n$ converges pointwise to f on S, and if there is a convergent series of positive constants $\sum M_n$ such that $0 \le |u_n(x)| \le M_n$ for every $n \ge 1$ and $x \in S$, then $\sum u_k$ converges uniformly on S.

Circle of convergence for complex power series
For every complex power series $\sum_{n=0}^{\infty} a_n (z-a)^n$, there is a circle of convergence. Inside the circle of convergence, the series converges absolutely for any fixed z, and converges uniformly for variable z with domain inside the circle. The series diverges outside the circle. On the boundary, the series may or may not converge.

Power series expansion / representation
Power series expansion of f about a: $f(x) = \sum_{n=0}^{\infty} a_n (x-a)^n$

Properties of power-series represented functions
• Assume a function f is represented by the power series $f(x) = \sum_{n=0}^{\infty} a_n (x-a)^n$ in an open interval $(a-r, a+r)$. Then f is continuous on this interval, and its integral over any closed subinterval may be computed by integrating the series term by term. In particular, for every x in $(a-r, a+r)$, we have $\int_a^x f(t)\,dt = \sum_{n=0}^{\infty} a_n \int_a^x (t-a)^n\,dt = \sum_{n=0}^{\infty} \frac{a_n}{n+1} (x-a)^{n+1}$
• If f is represented by a power series as above in the interval of convergence $(a-r, a+r)$, then the derivative $f'(x)$ exists for each x in the interval of convergence, and is given by the differentiated series $f'(x) = \sum_{n=1}^{\infty} n a_n (x-a)^{n-1}$, which also has radius of convergence r.
• The sum function of a power series has derivatives of *every* order, which may be obtained by repeated term-by-term differentiation of the power series.

Uniqueness theorem for power series expansions: If a function has a power series expansion, it has the form $f(x) = \sum_{k=0}^{\infty} \frac{f^{(k)}(a)}{k!}(x-a)^k$. In other words, $a_n = \frac{f^{(n)}(a)}{n!}$.
• A function has a power series expansion about a if and only if it is infinitely differentiable in the interval, and the error term of its Taylor polynomial about a tends to 0 as $n \to \infty$.
• Sufficient condition for Taylor's series converging to $f(x)$ in an interval: If f is infinitely differentiable in an open interval, and if $|f^{(n)}(x)| \leq A^n$ for $n \geq 1$ and $x \in I$, for some positive constant A.

Classic power series expansions
$\sin x = x - \frac{x^3}{3!} + \frac{x^t}{5!} - \frac{x^7}{7!} + \cdots + (-1)^{n-1}\frac{x^{2n-1}}{(2n-1)!} + \cdots$
$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + \cdots$
$e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + \cdots$

General mathematics

Binomial theorem
$(1+x)^n = \sum_{k=0}^{n} \binom{n}{k} x^k$
Also, for $|x| < 1$, $(1+x)^\alpha = \sum_{n=0}^{\infty} \binom{\alpha}{n} x^n$, with α an arbitrary real number.

Euler's constant
$\gamma = \lim_{n\to\infty} \left(1 + \frac{1}{2} + \cdots + \frac{1}{n} - \log n\right)$
γ is approximately equal to 0.5772156649

Vector algebra

Basic Properties
Cauchy-Schwarz Inequality: For A and B (vectors in V_n), we have $(A \cdot B)^2 \leq (A \cdot A)(B \cdot B)$. The equality sign holds iff one of the vectors is a scalar multiple of the other.
Length or norm: $\|A\| = (A \cdot A)^{1/2}$.
Triangle inequality: $\|A + B\|^2 \leq \|A\| + \|B\|$. The equality sign holds iff $A = O$, $B = O$, or if $B = cA$ for some $c > 0$.
Projection of A along B: tB, where $t = \frac{A \cdot B}{B \cdot B}$.
Angle between two vectors: $\theta = \arccos \frac{A \cdot B}{\|A\|\|B\|}$, $0 \leq \theta \leq \pi$.
Dot product for $V_n(C)$: $A \cdot B = \sum_{k=1}^{n} a_k \overline{b}_k$.
Two vectors A and B in V_n are linearly dependent iff they lie on the same line through the origin.
Eqn for a line: $L(P; A) = \{P + tA \ |treal\}$.
Cartesian eqn for line: $y - q = \frac{b}{a}(x - p)$, or $(X - P) \cdot N = 0$, where $N = (b, -a)$, and is normal to the line. Or, $X \cdot N = P \cdot N$.
Eqn of a plane in V_n, with A and B independent: $M = \{P + sA + tB\}$.
Three vectors A, B, C are linearly dependent iff they lie on the same plane through the origin.
Alternate description of plane in V_3: $(X - P) \cdot N = 0$, where $N = A \times B$.

Concepts
Linear span: All vectors which can be represented by a linear combination of the set of vectors.
Linear independence: Spans the O vector in one way only (the trivial representation).
Basis: A set S of vectors is called a *basis* of V_n if S spans every vector in V_n uniquely.

Cross product
$A \times B = (a_2b_3 = a_3b_2, a_3b_1 - a_1b_3, a_1b_2 - a_2b_1)$
$A \times B = -(B \times A)$
$A \times (B + C) = (A \times B) + (A \times C)$
$c(A \times B) = (CA) \times B$
$a \cdot (A \times B) = 0$
$B \cdot (A \times B) = 0$
$\|A \times B\|^2 = \|A\|^2\|B\|^2 - (A \cdot B)^2$
$A \times B = O$ iff A and B are linearly dependent
Cross product is *not* associative
$\|A \times B\| = \|A\|\|B\| \sin\theta$
The length of $A \times B$ is equal to the area of the parallelogram determined by A and B.
Alternate defintion of cross product:

$$A \times B = \begin{vmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{vmatrix}$$

Scalar triple product
$A \cdot B \times C = A \cdot (B \times C)$.
A, B, C are linearly dependent iff $A \cdot B \times C = 0$.
Scalar triple product is equal to the volume of the parallellipiped defined by A, B, C.
Alternate notation for scalar triple product: $A \cdot B \times C = [ABC]$.

Conic sections
Definition 1: Slicing right circular cones with a plane.
Definition 2: Ellipse: $d_1 + d_2$ constant, Hyperbola: $|d_1 - d_2|$ constant, Parabola $d_1 = d_2$ (focus, directrix).
Definition 3: Eccentricity (e): constant ratio of distance from a fixed point and fixed line. Ellipse ($0 < e < 1$), Parabola ($e = 1$), Hyperbola ($e > 1$)
Vector eqn for conic section: $\|X - F\| = e\, d(X, L)$, where N is a normal to L and $d(X, L) = \frac{|(X-P)\cdot N|}{\|N\|}$.
If N is a unit normal, eqn for conic section becomes $\|X - F\| = e|(X - F) \cdot N - d|$.
Cartesian eqn for ellipse and hyperbola: $\frac{x^2}{a^2} + \frac{y^2}{a^2(1-e^2)} = 1$, where $a = \frac{ed}{(1-e^2)}$, where d is distance between F and L.

Vector calculus

Algebra + differentiation
$(F + G)' = F' + G'$
$(uF)' = u'F = uF'$
$(F \cdot G)' = F' \cdot G + F \cdot G'$
$(F \times G)' = F' \times G + F \times G'$
$(F \circ u)' = F'[u(t)]u'(t)$

Properties
If F has constant length on I, then $F \cdot F' = 0$ on I.
$C \cdot \int_a^b F(t)\, dt = \int_a^b C \cdot F(t)\, dt$
If F and $\|F\|$ are integrable, we have $\|\int_a^b F(t)\, dt\| \leq \int_a^b \|F(t)\|\, dt$.
Unit tangent: $T(t) = \frac{X'(t)}{\|X'(t)\|}$
Principal normal: $N(t) = \frac{T'(t)}{\|T'(t)\|}$
Relationship b/w acceleration, $T(t)$, $N(t)$: $\boldsymbol{a}(t) = v'(t)T(t) + v(t)T'(t) = v'(t)T(t) + v(t)\|T'(t)\|N(t)$
Acceleration vector is linear combination of $T(t)$ and $N(t)$
Osculating plane: plane determined by $T(t)$ and $N(t)$
Rectifiable curve: $|\pi(P)| \leq M$ for a positive M and for all partitions P.

Arc length: $\Lambda(a,b) = \sup|\pi(P)| = \int_a^b v(t)\,dt = \int_a^b \|X'(t)\|\,dt$.
Arc length function: $s(t) = \Lambda(a,t)$, for $t > a, s(a) = 0$.
Curvature vector: rate of change of the unit tangent with respect to arc length.
$\frac{dT}{ds} = \frac{dt}{ds}\frac{dT}{dt} = \frac{1}{s'(t)}T'(t) = \frac{1}{v(t)}T'(t) = \frac{\|T'(t)\|}{v(t)}N(t)$.
Curvature of a curve at t: $\kappa(t)$ = length of curvature vector $= \frac{\|T'(t)\|}{v(t)}$.
Acceleration in terms of radial and transverse components: $\boldsymbol{a} = \left(\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2\right)\boldsymbol{u_r} + \left(r\frac{d^2\theta}{dt^2} + 2\frac{dr}{dt}\frac{d\theta}{dt}\right)\boldsymbol{u_\theta}$

Linear spaces

Concepts
Linear space: a set which has operations "multiplication" and "addition" defined, which obey certain axioms, such as closure under addtion and closure under multiplication by scalars.
Subspace: subset of linear space which is itself a linear space, i.e. follows closure axioms.
Dependence: A set S is dependent if it spans the zero element in more than one way.
Finite basis for V: independent and spans V.
Dimension: the number of elements in a basis.
Euclidean space: linear space with an inner product defined.
Orthogonal set: every pair of distinct elements has $\langle x, y\rangle = 0$.
Orthonormal set: orthogonal set with each element having norm 1.
Every finite dimensional Euclidean space has an orthogonal / orthonormal basis.
Projection of x along y: $\frac{\langle x,y\rangle}{\langle y,y\rangle}y$, if $y \neq O$.
$S^\perp$: the set of all elements orthogonal to S.
$S^\perp$ is always a subspace. If S is also a subspace, then $S^\perp$ is called the orthogonal complement of S.
Projection of x on the subspace S: $s = \sum_{i=1}^n \langle x, e_i\rangle e_i$, where $\{e_1, \ldots, e_n\}$ is an orthonormal basis for S.
The projection of x on S is nearer to X than any other element of S.

Important applications
Trigonometric polynomials:
  Elements of a subspace S of a linear space $V = C(0, 2\pi)$,
  with inner product $\int_0^{2\pi} f(x)\,g(x)\,dx$,
  spanned by the orthonormal basis $\varphi_0(x) = \frac{1}{\sqrt{2\pi}}$, $\varphi_{2k-1}(x) = \frac{\cos kx}{\sqrt{\pi}}$, $\varphi_{2k}(x) = \frac{\sin kx}{\sqrt{\pi}}$.
Projection of $f \in C(0, 2\pi)$ on S (trig polys):
  $f_n = \sum_{k=0}^{2n}\langle f, \varphi_k\rangle\varphi_k$,
  where $\langle f, \varphi_k\rangle = \int_0^{2\pi} f(x)\,\varphi_k(x)\,dx$.
Fourier coefficients of f:
  The number $\langle f, \varphi_k\rangle$ so defined.
Alternate notation for projection of f:
  $f_n(x) = \frac{1}{2}a_0 + \sum_{k=1}^n (a_k \cos kx + b_k \sin kx)$,
  where $a_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\,\cos kx\,dx$
  and $b_k = \frac{1}{\pi}\int_0^{2\pi} f(x)\,\sin kx\,dx$.
Legendre polynomials for approximating functions:
  If $f \in C(-1, 1)$, with inner product $\int_{-1}^1 f(x)\,g(x)\,dx$,
  then the polynomial of degree n which "best approximates" f
  is the projection of f on the subspace of polynomials of degree n,
  which can be obtained by the projection formula using the normalized Legendre polynomials.

Inner product
(1) $\langle x, y\rangle = \overline{\langle y, x\rangle}$
(2) $\langle x, y + z\rangle = \langle x, y\rangle + \langle x, z\rangle$
(3) $c\langle x, y\rangle = \langle cx, y\rangle$
(4) $\langle x, x\rangle > 0$ if $x \neq O$

Norms
$\|x\| = \langle x, x\rangle^{1/2}$
$\|cx\| = |c|\ \|x\|$
Triangle inequality: $\|x + y\| \leq \|x\| + \|y\|$

Cauchy-Schwarz Inequality
In a Euclidean space: $|\langle x, y\rangle|^2 \leq \langle x, x\rangle\langle y, y\rangle$.
Or, equivalently: $|\langle x, y\rangle| \leq \|x\|\ \|y\|$.
The equality holds iff x and y are dependent.

Examples of linear spaces
Real numbers, complex numbers, V_n, $V_n(C)$, all functions defined on an interval, all polynomials, all functions continuous or differentiable or integrable on an interval.

General properties
If an element x is described as a linear combination of elements of some orthogonal basis $x = \sum_{i=1}^{n} c_i e_i$, then its components relative to that basis are $c_j = \frac{\langle x, e_j\rangle}{\langle e_j, e_j\rangle}$.
If the basis is an orthonormal basis, then $c_j = \langle x, e_j\rangle$.
Parseval's formula: If V has a finite orthonormal basis, then $\langle x, y\rangle = \sum_{i=1}^{n} \langle x, e_i\rangle\overline{\langle y, e_i\rangle}$.
Parseval's for $y = x$: $\|x\|^2 = \sum_{i=1}^{n} |\langle x, e_i\rangle|^2$.
Gram-Schmidt process: A process for taking a finite or infinite sequence of elements, and constructing a sequence of elements which is orthogonal and which has the same span as the sequence.

Legendre polynomials
Take the linear space of all polynomials, with inner product $\langle x, y\rangle = \int_{-1}^{1} x(t)\, y(t)\, dt$, and take the infinite sequence $x_0, x_1, x_2, \ldots,$ where $x_n(t) = t^n$. If the orthogonalization theorem / Gram-Schmidt process is applied to this sequence, a new sequence of orthogonal polynomials is obtained. These are the Legendre polynomials. If each polynomial is divided by its norm, we have the normalized Legendre polynomials.

Linear transformations and matrices

Concepts
Linear transformation: A function from one linear space to another, with the properties of additivity and homogeneity.
Basically, a linear transformation is equivalent to a series of linear equations with no constant terms.
If there are constant terms, you have instead an *affine* transformation.
Range of a linear transformation is always a subspace of the target space.
Linear transforms always map zero element to zero element.
Null space of T: All elements of source space that map to zero element in target space (also called kernel of T).
The null space of T is denoted by $N(T)$.
Null space is always a subspace of source space.
Nullity: the dimension of the null space.
Rank: the dimension of the range.
Nullity + Rank = dimension of the source space.
$\mathscr{L}(V, W)$: set of all linear transformations from V to W, where V and W have the same scalars. $\mathscr{L}(V, W)$ is itself a linear space, with addition and multiplication by scalars as defined below.
The composition of two linear transformations is also linear (for same scalar spaces).
There is one and only one linear transformation that maps each of the n basis elements of V onto any n arbitrary elements of W.
If $T(e_k) = u(k)$, and $x = \sum_{k=1}^{n} x_k e_k$, then $T(x) = \sum_{k=1}^{n} x_k u_k$.
If $T(e_k) = \sum_{i=1}^{m} t_{ik} w_i$, then the matrix representation of T is $m \times n$ and each column represents the components of the target space basis elements T maps a given source space basis element to.
$T(x) = \sum_{i=1}^{m} y_i w_i$, where $y_i = \sum_{k=1}^{n} t_{ik} x_k$, for $i = 1, 2, \ldots, m$.

Diagonal matrix: A matrix with all non-diagonal elements = 0.
For any finite linear transformation, you can create a diagonal matrix representation with 1's along the diagonal equal to the rank of the transformation, and 0's on the diagonal after that. This is done by choosing the bases appropriately.
Linear space of matrices: $M_{m,n}$, with appropriately defined addtion and scalar multiplication.
Linear transformations and matrices are isomorphic.
Matrix multiplication: If A is $m \times p$, and B is $p \times n$, then AB is such that $c_{ij} = \sum_{k=1}^{p} a_{ik} b_{kj}$.
Matrix multiplication corresponds to linear transformation composition.
Computing a transformation corresponds to matrix multiplication between the transformation matrix and a column matrix of x.
If we have a system of linear equations such that $T(x) = c$, then if we can find $k = \text{nullity}(T)$ independent solutions of the homogeneous system $T(x) = O$, which we denote $v_1, \ldots, v_k$, and if we can find one particular solution of the nonhomogeneous system, which we denote by b, then the general solution of the system is $x = b + t_1 v_1 + \cdots + t_k v_k$, where $t_1, \ldots, t_n$ are arbitrary scalars. In other words, the solution is any particular solution, plus any element of the image of the null space of T.
A matrix is nonsingular iff T is invertible.
You can obtain a matrix inverse by applying the Gauss-Jordan method to a matrix augmented with an identity matrix.
For the linear system $AX = C$ (square), if A is nonsingular there is a unique solution of the system: $X = A^{-1}C$.
Transpose of a matrix A^t has its i, j entry as a_{ji}.
Invariance: a subspace U of S is called invariant under T if T maps every element of U onto an element of U.

Gauss-Jordan elimination method
You can perform the following operations and have an equivalent system: 1) Interchanging two rows, 2) multiplying all terms of a row by a nonzero scalar, and 3) adding a multiple of one row to another row.
Process: 1) obtain a 1 in the top left position, 2) make all other elements of column 1 to be 0, 3) Continue to get an upper triangular matrix, 4) Continue to make a diagonal matrix.

One-to-one linear transformations
T is one-to-one $\iff$ T is invertible and T^{-1} is linear $\iff$ $N(T)$ contains only O.
If V is finite-dimensional, then: T is one-to-one $\iff$ Independent elements in V map to independent elements in $T(V) \iff \dim T(V) = \dim V$

Algebra of linear transformations
$(S + T)(x) = S(x) + T(x)$
$(cT)(x) = cT(x)$
Composition: $(ST)(x) = S[T(x)]$, with domains and ranges appropriately defined.
$R(ST) = (RS)T$
$T^0 = I$, and $T^n = TT^{n-1}$ for $n \geq 1$.
$R(S + T) = RS + RT \quad \text{and} \quad R(cS) = c(RS)$

Algebra of matrices
$A(BC) = (AB)C$
$(A + B)C = AC + BC$
$C(A + B) = CA + CB$
For a square matrix, $A^0 = I$, and $A^n = AA^{n-1}$, for $n \geq 1$.
$(AB)_i = A_i B$, where A_i denotes the ith row of A.

Examples of linear transformations
Identity transformation, zero transformation, multiplication by a scalar, linear eqns with no constant terms, inner product with a fixed element, projection on a subspace, differentiation, integration.

Determinants

Concepts
In 3-space, the scalar triple product is the determinant of the matrix whose rows are the vectors: $A_1 \times A_2 \cdot A_3$.
In 3-space, this is the area of the parallelipiped determined by the three vectors.
Only defined for square matrices.
The determinant is unique (in satisfying the axioms).
A'_{kj} denotes the matrix obtained from A by replacing row A_k with unit coordinate vector I_j.
Cofactor of $a_{kj} = \det A'_{kj}$.
Minor: The kj minor of A is the matrix obtained from A by deleting the kth row and the jth column from A.
The minor is denoted by A_{kj}.
Cofactor matrix: $\operatorname{cof} A = (\operatorname{cof} a_{ij})$

Axioms
(1) Homogeneity in each row: $d(\dots, tA_k, \dots) = t\, d(\dots, A_k, \dots)$.
(2) Additivity in each row: $d(\dots, A_k + C, \dots) = d(\dots, A_k, \dots) + d(\dots, C, \dots)$.
(3) Determinant is zero if any two rows are equal.
(4) Determinant of identity matrix is 1.

Properties
(1) If any row is O, the determinant is 0.
(2) If any two rows are interchanged, the determinant changes sign.
(3) If the rows are dependent, the determinant is 0.
(4) A set of n vectors in n-space is independent iff their determinant is 0.

Calculating determinants
Determinant of 1×1 matrix is just the element.
Determinant of 2×2 matrix is $\left[\begin{smallmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{smallmatrix}\right] = a_{11}a22 - a_{12}a_{21}$.
Determinant of a diagonal matrix is the product of its diagonal elements.
Determinant of an upper triangular matrix is the product of its diagonal elements.

Gauss-Jordan method for finding determinant
Use Gauss-Jordan to obtain an upper triangular matrix, then take the product of the diagonal elements.
(1) Interchange two rows (changes sign).
(2) Multiply a row by a nonzero scalar (is multiplied by the scalar).
(3) Add a scalar multiple of one row to another row (no change).

Algebra
$\det(AB) = (\det A)(\det B)$
For A nonsingular, $\det A^{-1} = \frac{1}{\det A}$
$\det\left[\begin{smallmatrix} A & O \\ O & B \end{smallmatrix}\right] = (\det A)(\det B)$, where A and B form part of a block diagonal matrix.
Expansion by cofactors: $\det A = \sum_{j=1}^{n} a_{kj} \operatorname{cof} a_{kj}$ for any row k.
Relationship between cofactor and determinant of the minor: $\operatorname{cof} a_{kj} = (-1)^{k+j} \det A_{kj}$.
$\det A = \det A^t$
$A(\operatorname{cof} A)^t = (\det A)I$, for any square matrix with $n \geq 2$.
$A^{-1} = \frac{1}{\det A}(cofA)^t$. This is a good way to compute the inverse of A.

Cramer's rule
For n linear equations in n unknowns, we have $AX = B$.
Solution: $x_j = \frac{1}{\det A} \sum_{k=1}^{n} b_k \operatorname{cof} a_{kj}$.

Eigenvalues and eigenvectors

Concepts
Eigenvalue / eigenvector: When a linear tranformation has $T(x) = \lambda x$ for some scalar λ, λ is an eigenvalue and x is an eigenvector, as long as $x \neq 0$.
For $T : V \to V$, $T \in \mathscr{L}(V, V)$, $\dim V = n$, if there are n independent elements u_k such that $T(u_k) = \lambda_k u_k$, then $A = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$ is a respresentation of T relative to the u_k. The converse is also true.
Eigenspace: the linear subspace consisting of all elements such that $T(x) = \lambda x$ for a given λ (including zero). This is a linear subspace.
Zero *can* be an eigenvalue.
In a real linear space, T preserves the "direction" of an eigenvector. This is not generally true for complex linear spaces.
Eigenvectors corresponding to distinct eigenvalues are independent.
$\lambda_1 \cdots \lambda_n = \det A$. Trace of A: Sum of the roots of $f(\lambda)$, which is equal to the sum of the diagonal elements of A.

Characteristic polynomials
Eigenvalues must satisfy $T(x) = \lambda x \quad \Longrightarrow \quad (\lambda I - T)(x) = O$ for $x \neq 0$.
If this is satisfied, it means that some x other than 0 is mapped to 0, which means that $(\lambda I - T)$ is not one-to-one.
This means that $(\lambda I - T)$ has no inverse, and thus that the determinant of its matrix A is 0.
So $\det(\lambda I - A) = 0$.
We thus need to find the roots of the polynomial $f(\lambda) = \det(\lambda I - A)$.
This is a polynomial in λ of degree n, the term or highest degree is λ^n, and the constant term is $(-1)^n \det A$.
The eigenvalues are the roots of the characteristic polynomial which lie in F (real or complex scalars).
Once eigenvalues are found, you can find corresponding eigenvectors by solving $AX = \lambda X$.

Similar matrices
If A and B are $n \times n$ matrices representing the same linear transformation, there is some matrix C such that $B = C^{-1}AC$.
If A is relative to basis E, and B is relative to basis U, the bases are related by $U = EC$.
Similar matrices: square matrices related by $B = C^{-1}AC$.
Similar matrices have the same determinant, same characteristic polynomial, and the same eigenvalues.

Eigenvalues and eigenvectors in Euclidean spaces

General concepts
$\lambda = \frac{\langle T(x), x\rangle}{\langle x, x\rangle}$ and $\bar{\lambda} = \frac{\langle x, T(x)\rangle}{\langle x, x\rangle}$
An eigenvalue is real or 0 iff $\langle T(x), x\rangle = \langle x, T(x)\rangle$.
An eigenvalue is pure imaginary or 0 iff $\langle T(x), x\rangle = -\langle x, T(x)\rangle$.
Hermitian/symmetric transformation: $\langle T(x), y\rangle = \langle y, T(x)\rangle$
skew-Hermitian/skew-symmetric transformation: $\langle T(x), y\rangle = -\langle y, T(x)\rangle$
We use "Hermitian" terminology if the Euclidean space is complex, and "symmetric" if real.
Hermitian transformations have real eigenvalues.
Skew-Hermitian transformations have pure imaginary eigenvalues.
Diagonal elements for: Hermitian(real), skew-Hermitian(pure imaginary), skew-symmetric(zero).
Distinct eigenvalues of Hermitian or skew-Hermitian transformations have orthogonal eigenvectors.
If T operates on a finite-dimensional complex space and is Hermitian or skew-Hermitian, then there exist n orthonormal eigenvectors (which form a basis for V).
T can be determined to be (skew-)Hermitian based only on its action on any basis.
Adjoint of a matrix: The transpose of the conjugate, $\bar{A}^t$, denoted by A^*. Hermitian matrix: $a_{ij} = \bar{a}_{ji}$, or $A = A^*$.
Skew-Hermitian matrix: $a_{ij} = -\bar{a}_{ji}$, or $A = -A^*$.
For (skew-)Hermitian matrices, the diagonalizing matrix C (where $\Lambda = C^{-1}AC$) has $C^{-1} = C*$, or $C\,C^* = I$.
For real matrices, you always have $A^t = A^*$.
Unitary matrix: $A\,A^* = I$.

Orthogonal matrix: $A\,A^t = I$.
Every real unitary matrix is also orthogonal, since $A^* = A^t$.
Symmetric matrix: $A = A^t$.
Skew-symmetric matrix: $A = -A^t$.
Every square matrix A can be expressed as a sum of a Hermitian matrix ($\frac{1}{2}(A + A^*)$) and a skew-Hermitian matrix ($\frac{1}{2}(A - A^*)$).
Every square matrix A can be expressed as a sum of a symmetric matrix ($\frac{1}{2}(A + A^t)$) and a skew-symmetric matrix ($\frac{1}{2}(A - A^t)$).
Orthogonal matrices have determinant $= \pm 1$.

How to find a diagonalizing matrix for (skew-)Hermitian matrices
(1) Find an orthonormal set of eigenvectors $u_1, \ldots, u_n$
(2) Use the components of u_j (relative to the basis of unit coordinate vectors) as entries of the jth column.

Sturm-Liouville operators
Let V be subspace of $C(a, b)$ with continuous second derivative on $[a, b]$.
Let p and q be fixed functions in $C(a, b)$, with p having a continuous derivative on $[a, b]$.
Let $f \in V$.
Let p and q satisfy the boundary conditions $p(a)f(a) = 0$ and $p(b)f(b) = 0$.
Let $T : V \to C(a, b)$ be the following operator: $T(f) = (pf')' + qf$.
T is then called a *Sturm-Liouville* operator.
Sturm-Liouville operators are symmetric.

Quadratic forms
For a symmetric operator (on a real space), quadratic form associated with T is: $Q(x) = \langle T(x), x\rangle$.
If A is a representation of T relative to an orthonomal basis, then $Q(x) = \sum_{i=1}^n \sum_{j=1}^n a_{ij} x_i x_j$.
Where x_k is the component of x relative to e_k.
You can actually associate a quadratic form with *any* square matrix (not just symmetric).
Diagonal form: If A is a diagonal matrix, $Q(x) = \sum_{i=1}^n a_{ii} x_i^2$.
The quadratic form can also be written as $Q(x) = X\,A\,X^t$, where X is a row matrix.
$X\,A\,X^t = X\,[\frac{1}{2}(A + A^t)]\,X^t$.

Converting a real quadratic form to diagonal form
If A is a real symmetric matrix, and C is a diagonalizing matrix (so $\Lambda = C^t AC$), then $XAX^t = \sum_{i=1}^n \lambda_i y_i^2$, where $Y = XC$, and $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A. This is because $XAX^t = Y\Lambda Y^t$.

Relation between eigenvalues of a symmetric transformation and its quadratic form
For eigenvectors x with norm 1, $Q(x) = \lambda$.
So, any eigenvalues of T are found on the values Q takes on the unit sphere.
If Q has a maximum or minumum on the unit sphere at x, then x is an eigenvector, and $Q(x)$ is its eigenvalue.
If you find the minimum value of Q on the unit sphere, this is the smallest eigenvalue.
You can then look at the subset of the unit sphere in $S^\perp$ - all elements not spanned by the found eigenvector.
The minimum of Q on this unit sphere subset will be the next smallest eigenvalue, etc.

Unitary transformations
Unitary transformation: $\langle T(x), T(y)\rangle = \langle x, y\rangle$.
When E is a real space, a unitary transformation is also called an *orthogonal* transformation.
Unitary transformations preserve inner products, orthogonality, norms, and distances.
Unitary transformations are invertible, and their inverses are unitary on $T(V)$.

Unitary transformations and eigenvectors
Any eigenvalues of T have $|\lambda| = 1$.
Distinct eigenvalues have corresponding eigenvectors which are orthogonal.
For $T : V \to V$ and finite dimensional V, an orthonormal T-eigenvector basis of V exists,
and the matrix of T relative to this basis is the diagonal matrix Λ.

Probability

Concepts
Finitely additive set function: $f(A \cup B) = f(A) + f(B)$, when A and B are disjoint.
Countably additive set function: Also holds for unions of countably infinite subsets.
Boolean algebra of sets: closed under union and complement (also closed under intersection and difference).
Boolean σ-algebra: also closed under countably infinite union.
Finitely additive measure: a finitely additive set function that is nonnegative.
Probability measure: set function which is finitely additive, nonnegative, and $P(S) = 1$.
Probability space: S (sample space), $\mathscr{B}$ (boolean algebra of subsets of S), P (probability measure).
Experiment: Various definitions.
Descriptive vs. Inferential statistics.
Number of distinct subsets of size k from a set of size n: $\binom{n}{k} = \frac{n!}{k!(n-k)!}$.
Sampling with replacement: n^k.
Sampling without replacement: $n(n-1)(n-2)\cdots(n-k+1) = \frac{n!}{(n-k)!}$.
Total number of subsets: $\sum_{k=0}^{n} \binom{n}{k} = 2^n$.
Conditional probability: $P(A|B) = \frac{P(A \cap B)}{P(B)}$.
Independence: $P(A \cap B) = P(A)P(B)$.
Random variable: a real vector-valued function on S.
Distribution function: $F(t) = P(X \leq t)$.
Probability mass function: $p(t) = P(X = t)$.
If S is uncountable, there can be at most countably many points at which $P > 0$.
Probability density function: $f(t)$, where $P(X \leq t) = \int_{-\infty}^{t} f(u)\,du$.
Mean squared error: the mean squared difference between an estimator and the true value.
Standard error: $\sqrt{\text{MSE}}$.
Sampling distribution: distribution of an estimator of a parameter.

Estimation
Maximum likelihood fn: A joint probability distribution, viewed as fixed observed values and variable parameter(s).
Maximum likelihood estimator: A parameter estimator which maximizes the maximum likelihood function.
Invariance: $g(\hat{\theta})$ is the MLE of $g(\theta)$.
Normal distribution: $\hat{\mu} = \bar{X}_n$, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$.

Sampling distributions
Normal distribution, $\bar{X}_n$: mean $= \mu$, var $= \sigma^2/n$.
Normal distribution, $(1/n)\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$: χ^2 distribution with $n-1$ df.
For normal distributions, sample mean and sample variance are independent RV's.

Chi-Square distribution
Gamma distribution with $\alpha = n/2$ and $\beta = 1/2$.

$$f(x) = \frac{1}{2^{n/2}\Gamma(n/2)} x^{(n/2)-1} e^{-x/2}$$

n is number of "degrees of freedom".
$E(X) = n$, $\text{Var}(X) = 2n$.
The sum of k independent χ^2 distributed RV's has a χ^2 distribution with $df = n_1 + \cdots + n_k$.
If X has a normal distribution, then $Y = X^2$ has a χ^2 distribution with one degree of freedom.
If k RV's are independent and normally distributed, then $X_1^2 + \cdots + X_k^2$ are χ^2 with k df.

t distribution
If Y is standard normal and Z is χ^2 with n df, then

$$X = \frac{Y}{\left(\frac{Z}{n}\right)^{1/2}}$$

has a t distribution with n df.

$$f(x) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{(n\pi)^{1/2}\Gamma\left(\frac{n}{2}\right)}$$

When $n = 1$, the t distribution is a Cauchy distribution.
When $n \to \infty$, the t distribution approaches the normal distribution.

Bernoulli trials
Definition: Sequence of repeated, independent trials, p = probability of success, q = failure.
$P(k$ successes in n trials): $\binom{n}{k}p^k q^{n-k}$.
$P($at least r successes in n trials): $\sum_{k=r}^{n}\binom{n}{k}p^k q^{n-k}$.

Cauchy distribution
$F(t) = \frac{1}{2} + \frac{1}{\pi}\arctan t$.
$f(t) = \frac{1}{\pi(1+t^2)}$.

Exponential distributions

$$F(t) = \begin{cases} 1 - e^{-\lambda t} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}$$

$$f(x) = \begin{cases} \lambda e^{-\lambda t} & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}$$

Normal distributions

$$F(t) = \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{t} e^{-[(u-m)/\sigma]^2/2}\,du$$

$$f(t) = \frac{1}{\sigma\sqrt{2\pi}}e^{-[(t-m)/\sigma]^2/2}$$

Approximating binomial distribution with normal

$$\sum_{k=a}^{b}\binom{n}{k}p^k q^{n-k} \sim \Phi\left(\frac{b - np + \frac{1}{2}}{\sqrt{npq}}\right) - \Phi\left(\frac{a - np - \frac{1}{2}}{\sqrt{npq}}\right)$$

Expectation

$$E(X) = \int_{-\infty}^{+\infty} t f_X(t)\,dt$$

$$E(X) = \sum_{k=1}^{\infty} x_k p_k$$

Variance

$$\mathrm{Var}(X) = \int_{-\infty}^{+\infty} [t = E(X)]^2 f_X(t)\,dt$$

$$\mathrm{Var}(X) = \sum_{k=1}^{\infty}[x_k - E(X)]^2 p_k$$

Expectation of $Y = \varphi(X)$

$$E(Y) = \int_{-\infty}^{+\infty} \varphi(t) f_X(t)\, dt$$

Confidence intervals
To determine confidence interval:
(1) Find a way that say 0.95 of the area of a relevant distribution falls within boundaries.
(2) Make an inequality.
(3) Manipulate the inequality to isolate the parameter of interest.
For $X_1, \ldots, X_n$ from a normal distribution,

$$\Pr(\bar{X}_n - \frac{c\sigma'}{\sqrt{n}} < \mu < \bar{X}_n + \frac{c\sigma'}{\sqrt{n}}) = 0.95$$

where c is a constant such that

$$\int_{-c}^{c} g_{n-1}(x)\, dx = 0.95$$

and $g_{n-1}(x)$ is a t distribution with $n-1$ degrees of freedom.

Unbiased estimators
An estimator is *unbiased* if its expected value equals the parameter for all parameter values.
Unbiased estimator of σ^2: $\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X}_n)^2$.

Hypothesis testing
Type I Error: Rejecting H_0 when it is true (α).
Type II Error: Accepting H_0 when it is false (β).
Can test two alternate values of a parameter, or one-sided alternatives.
Unbiased vs. biased tests (they are almost always biased).

Neyman-Pearson Lemma
A test procedure of the form: "Accept H_0 if $f_a(x) > kf_1(x)$"
...where H_0 and H_a are two point hypotheses
...and where $f_0(x)$ is the joint pdf of the sample observations when H_0 is true
...and with specified α
...will minimize β.
Likelihood ratio: $f_a(x)/f_0(x)$.

Uniformly Most Powerful tests
For a test procedure $\delta*$
...and a fixed level of significance α_0
...and for any other procedure δ
...δ is UMP if $\pi(\theta|\delta) \leq \pi(\theta|\delta*)$
...for every value of $\theta \in \Omega_a$.

Monotone likelihood ratio
$f_n(\boldsymbol{x} \,|\, \theta)$ has a MLR in the statistic $T = r(\boldsymbol{x})$
...if for any two values $\theta_1 \in \Omega$ and $\theta_2 \in \Omega$
...with $\theta_1 < \theta_2$
...the ratio $f_n(\boldsymbol{x} \,|\, \theta_2)/f_n(\boldsymbol{x} \,|\, \theta_1)$
...depends on $\boldsymbol{x}$ only through $r(\boldsymbol{x})$
...and this ratio is an increasing function of $r(\boldsymbol{x})$.

t test, one-sided
H_0: $\mu \leq \mu_0$
H_a: $\mu > \mu_0$
Normal distribution, μ and σ^2 unknown.

Likelihood ratio: $$\frac{\sup_{(\mu,\sigma^2)\in\Omega_a} f_n(\boldsymbol{x} \mid \mu, \sigma^2)}{\sup_{(\mu,\sigma^2)\in\Omega_0} f_n(\boldsymbol{x} \mid \mu, \sigma^2)}$$

H_0 should be rejected if $r(\boldsymbol{x}) \geq k$.
There is a version of this ratio that differs by a constant factor (and translated) and which has a t distribution:

$$U = \frac{n^{1/2}(\overline{X}_n - \mu_0)}{\left[\frac{S_n^2}{n-1}\right]^{1/2}}$$

This has $n-1$ degrees of freedom.
Thus, H_0 should be rejected if $U \geq c$, where c is chosen to have desired α value.
Comparing two means, equal sample size, equal variance:

$$U = \frac{\overline{X}_1 - \overline{X}_2}{S_{X_1X_2} \cdot \sqrt{\frac{2}{n}}} \text{ where } S_{X_1X_2} = \sqrt{\frac{S_{X1}^2 + S_{X2}^2}{2}} \text{ and with } 2n-2 \text{ degrees of freedom}$$

Comparing two means, unequal sample size, equal variance:

$$U = \frac{\overline{X}_1 - \overline{X}_2}{S_{X_1X_2} \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \text{ where } S_{X_1X_2} = \sqrt{\frac{(n_1-1)S_{X_1}^2 + (n_2-1)S_{X_2}^2}{n_1+n_2-2}} \text{ and with } n_1+n_2-2 \text{ degrees of freedom}$$

Comparing two means, dependent:

$$U = \frac{\overline{X}_D - \mu_0}{s_D/\sqrt{N}} \text{ with } N-1 \text{ degrees of freedom}$$

Distributions of functions of random variables
If you know the distribution of a random variable (or the joint distribution of multiple random variables), then you can determine the distribution of functions of those random variables [described in Apostol].

Chebyshev's inequality
For every positive number c:

$$P[|X - E(X)| > c] \leq \frac{\mathrm{Var}(X)}{c^2}$$

Weak law of large numbers
For n random variables with the same expectation (m) and variance, $\lim_{n\to\infty} P(|\overline{X} - m| > \epsilon) = 0$.

Central limit theorem
Central limit property:
Take a sequence of random variables $X_1, X_2, \ldots, X_n$
where each X_k has a finite expectation m_k and a finite variance σ_k^2.
Take $S_n = \sum_{k=1}^n (X_k - m_k)$ and $T_n = \frac{S_n}{\sqrt{\mathrm{Var}(S_n)}}$.
This sequence is said to satisfy the central limit property if, for all $a \leq b$,

$$\lim_{n\to\infty} P(a \leq T_n \leq b) = \frac{1}{\sqrt{2\pi}} \int_a^b e^{-u^2/2} du$$

If the variables are independent with a common distribution, this condition is satisfied.
More generally, another condition, the Lindeberg condition, is necessary and sufficient for independent variables to satisfy the CLP.
The CLP can also be satisfied for dependent variables, but the theory on this is still incomplete.

Line Integrals

General concepts

Continuous path in n-space: $\boldsymbol{\alpha} : [a, b] \in R^1 \to R^n$, continuous on $[a, b]$.
Smooth path: $\boldsymbol{\alpha}'$ exists and is continuous on (a, b).
If $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ trace out the same path in the same direction, the line integrals are equal.
If $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ trace out same path in opposite directions, the line integrals have opposite sign.
Independence of path: If a line integral depends only on the endpoints.
Potential function: If a scalar field φ has a gradient $\boldsymbol{f}$, φ is a potential function for $\boldsymbol{f}$.

Definitions

Line integral of a scalar field with respect to arc length:
(a) φ: $R^n \to R^1$.
(b) $\boldsymbol{\alpha} : piecewisesmoothpathin$n$-spacedefinedon$[a,b].
$(c) Definedas : \int_C \varphi \, ds = \int_a^b \varphi[\boldsymbol{\alpha}(t)]s'(t) \, dt$.
(d) Where $s(t) = \int_a^t \|\boldsymbol{\alpha}'(u)\| \, du$.
(e) Where $s'(t) = \|\boldsymbol{\alpha}'(t)\|$. Line integral of a scalar field with respect to a domain variable:
(a) $\int_C f(x, y) \, dx = \int_a^b f(x(t), y(t))x'(t) \, dt$.
Line integral of a vector field:
(a) $\boldsymbol{f} : R^n \to R^n$.
(b) $\boldsymbol{\alpha}$: piecewise smooth path in n-space defined on $[a, b]$.
(c) Defined as: $\int f_1 d\alpha_1 + \cdots + f_n d\alpha_n$.
(d) Written: $\int \boldsymbol{f} \cdot d\boldsymbol{\alpha}$.
(e) Equivalent to: $\int_a^b \boldsymbol{f}[\boldsymbol{\alpha}(t)] \cdot \boldsymbol{\alpha}'(t) \, dt$.
(f) $\int_C \boldsymbol{f} \cdot d\boldsymbol{\alpha} = \int_C \boldsymbol{f} \cdot \boldsymbol{T} \, ds$, $\boldsymbol{T}(t) = \frac{\boldsymbol{\alpha}'(t)}{\|\boldsymbol{\alpha}'(t)\|}$.

Properties

(a) $\int (a\boldsymbol{f} + b\boldsymbol{g}) \cdot d\boldsymbol{\alpha} = a \int \boldsymbol{f} \cdot d\boldsymbol{\alpha} + b \int \boldsymbol{g} \cdot d\boldsymbol{\alpha}$.
(b) $\int_C \boldsymbol{f} \cdot d\boldsymbol{\alpha} = \int_{C_1} \boldsymbol{f} \cdot d\boldsymbol{alpha} + \int_{C_2} \boldsymbol{f} \cdot d\boldsymbol{\alpha}$.

First fundamental theorem for line integrals

If the line integral of $\boldsymbol{f}$ is path independent,
And if $\varphi(\boldsymbol{x}) = \int_{\boldsymbol{a}}^{\boldsymbol{x}} \boldsymbol{f} \cdot d\boldsymbol{\alpha}$,
Then $\nabla\varphi(\boldsymbol{x}) = \boldsymbol{f}(\boldsymbol{x})$.

Second fundamental theorem for line integrals

For a scalar field φ with a continuous gradient $\nabla\varphi$ on an open connected set in R^n:
$\int_{\boldsymbol{a}}^{\boldsymbol{b}} \nabla\varphi \cdot d\boldsymbol{\alpha} = \varphi(\boldsymbol{a}) - \varphi(\boldsymbol{b})$.
This means the line integral of a gradient is path independent.

Conditions for a vector field to be a gradient (conservative)

A vector field is *conservative* if it is a gradient of some scalar field.
The following are equivalent:
(a) $\boldsymbol{f}$ is the gradient of some potential function.
(b) The line integral of $\boldsymbol{f}$ is independent of the path.
(c) The line integral of $\boldsymbol{f}$ is zero around every piecewise smooth closed path.
Necessary for a vector field to be a gradient:
(a) $D_i f_j(\boldsymbol{x}) = D_j f_i(\boldsymbol{x})$.
(b) If the domain of $\boldsymbol{f}$ is convex, then this is also sufficient.
(c) If the domain of $\boldsymbol{f}$ is open and simply connected, then this is sufficient.

Constructing potential functions

If $\boldsymbol{f}$ is a gradient, a potential function can be constructed.
Take the line integral of $\boldsymbol{f}$ from any point $\boldsymbol{a}$ to an arbitrary point $\boldsymbol{x}$, using any path.

Multiple Integrals

Repeated or iterated integration
If the given integrals exist and $Q = [a, b] \times [c, d]$,

$$\iint\limits_Q f(x, y)\, dx\, dy = \int_c^d \left[\int_a^b f(x, y)\, dx \right] dy$$

Integrability of continuous functions
If f is continuous on Q, it is integrable, and the integral can be obtained by iterated integration.
If f is defined and bounded on Q, and if its set of discontinuities is of content zero, then it is integrable.

More general regions
Type I:
$\{(x, y) \quad | \quad a \leq x \leq b \qquad \text{and} \qquad \varphi_1(x) \leq y \leq \varphi_2(x)\}$.
Can be evaluated by iterated integration, first on y, then on x.
Type II:
$\{(x, y) \quad | \quad c \leq y \leq d \qquad \text{and} \qquad \psi_1(y) \leq x \leq \psi_2(x)\}$.
Can be evaluated by iterated integraion, first on x, then on y.

Green's Theorem
If P and Q are scalar fields that are continuously differentiable
...and if C is a piecewise smooth Jordan curve [but more generally rectifiable Jordan I think],
...and if the below line integrals are traversed in a counterclockwise direction, then:

$$\oint_C P\, dx + Q\, dy = \iint\limits_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx\, dy$$

$$\oint_C Q\, dy = \iint\limits_R \frac{\partial Q}{\partial x}\, dx\, dy$$

$$\oint_C P\, dx = -\iint\limits_R \frac{\partial P}{\partial y}\, dx\, dy$$

[These last two don't seem necessarily to follow from Green's Theorem.]

Green's Theorem for multiply connected regions
If appropriate conditions are met:

$$\oint_{C_1} (P\, dx + Q\, dy) - \sum_{k=2}^{n} \oint_{C_k} (P\, dx + Q\, dy) = \iint\limits_R \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx\, dy$$

Winding number
If $\boldsymbol{\alpha}(t) = X(t)\boldsymbol{i} + Y(t)\boldsymbol{j}$ for $a \leq t \leq b$, and for a point P_0, And if C is a piecewise smooth closed curve in the plane, then

$$W(\boldsymbol{\alpha}; P_0) = \frac{1}{2\pi} \int_a^b \frac{[X(t) - x_0]Y'(t) - [Y(t) - y_0]X'(t)}{[X(t) - x_0]^2 - [Y(t) - y_0]^2}\, dt$$

If C is a Jordan curve (simple), then W is:
(a) Zero for every P_0 outside of C
(b) +1 for every P_0 inside C if $\boldsymbol{\alpha}$ traces C in a positive / counterclockwise diretion
(c) -1 for every P_0 inside C if $\boldsymbol{\alpha}$ traces C in a negative / clockwise diretion

Change of variables
For single integrals:

$$\int_{g(c)}^{g(d)} f(x)\,dx = \int_c^d f[g(t)]g'(t)\,dt$$

For double integrals:

$$\iint\limits_S f(x,y)dx\,dy = \iint\limits_T f[X(u,v),Y(u,v)]\left|\frac{\partial(X,Y)}{\partial(u,v)}\right| du\,dv$$

Where $x = X(u,v)$ and $y = Y(u,v)$ and $\boldsymbol{r}(u,v) = X(u,v)\boldsymbol{i} + Y(u,v)\boldsymbol{j}$ and $\boldsymbol{r}(T) = S$.

Higher dimension integrals
Pretty much everything follows analagously from two dimensions.

Surface Integrals

Concepts
Simple parametric surface: If $\boldsymbol{r}$ is one-to-one on T.
Regular point of $\boldsymbol{r}$: $\partial\boldsymbol{r}/\partial u$ and $\partial\boldsymbol{r}/\partial v$ are continuous, and fundamental vector product is nonzero.
Smooth surface: A surface $\boldsymbol{r}(T)$ is smooth if all of its points are regular points.

Parametric representations of surfaces
Sphere with radius a: $x = a\cos u\cos v \qquad y = a\sin u\cos v \qquad z = a\sin v$

Fundamental vector product
Take $\boldsymbol{r}(u,v) = X(u,v)\boldsymbol{i} + Y(u,v)\boldsymbol{j} + Z(u,v)\boldsymbol{k}$, where $(u,v) \in T$.

$$\frac{\partial\boldsymbol{r}}{\partial u} = \frac{\partial X}{\partial u}\boldsymbol{i} + \frac{\partial Y}{\partial u}\boldsymbol{j} + \frac{\partial Z}{\partial u}\boldsymbol{k}$$

$$\frac{\partial\boldsymbol{r}}{\partial v} = \frac{\partial X}{\partial v}\boldsymbol{i} + \frac{\partial Y}{\partial v}\boldsymbol{j} + \frac{\partial Z}{\partial v}\boldsymbol{k}$$

$$\text{Fundamental vector product} = \frac{\partial\boldsymbol{r}}{\partial u}\times\frac{\partial\boldsymbol{r}}{\partial v} = \begin{pmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ \frac{\partial X}{\partial u} & \frac{\partial Y}{\partial u} & \frac{\partial Z}{\partial u} \\ \frac{\partial X}{\partial v} & \frac{\partial Y}{\partial v} & \frac{\partial Z}{\partial v} \end{pmatrix} = \frac{\partial(Y,Z)}{\partial(u,v)}\boldsymbol{i} + \frac{\partial(Z,X)}{\partial(u,v)}\boldsymbol{j} + \frac{\partial(X,Y)}{\partial(u,v)}\boldsymbol{k}$$

The length of the fundamental vector product can be thought of as a sort of local magnification factor for areas.
The fundamental vector product is normal to the surface of $\boldsymbol{r}(T)$.

Surface area

$$a(S) = a(\boldsymbol{r}(T)) = \iint\limits_T \left\|\frac{\partial\boldsymbol{r}}{\partial u}\times\frac{\partial\boldsymbol{r}}{\partial v}\right\| du\,dv$$

Definition of surface integral
For a scalar field f:

$$\iint\limits_{\boldsymbol{r}(T)} f\,dS = \iint\limits_T f[\boldsymbol{r}(u,v)]\left\|\frac{\partial\boldsymbol{r}}{\partial u}\times\frac{\partial\boldsymbol{r}}{\partial v}\right\| du\,dv$$

Change of parametric representation
If $\boldsymbol{r}(u,v) = A$ and $\boldsymbol{R}(s,t) = B$, and if $G(s,t) = U(s,t,)\boldsymbol{i} + V(s,t)\boldsymbol{j}$ so $\boldsymbol{r}$ and $\boldsymbol{R}$ are smoothly equivalent, then

$$\frac{\partial \boldsymbol{R}}{\partial s} \times \frac{\partial \boldsymbol{R}}{\partial t} = \left(\frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v}\right) \frac{\partial(U,V)}{\partial(s,t)}$$

In other words, fundamental vector products of smoothly equialent functions have the same direction and only differ in magnitude.
Which means that surface integrals are invariant under change of smoothly equivalent parametric representations.

Alternate notation
Sometimes we want to take a surface integral of a vector field, in the direction normal to the f.v.p.:

$$\iint\limits_S \boldsymbol{F} \cdot \boldsymbol{n}\, dS = \iint\limits_T \boldsymbol{F}[\boldsymbol{r}(u,v)] \cdot \boldsymbol{n}(u,v) \| \frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v} \| du\, dv = \iint\limits_T \boldsymbol{F}[\boldsymbol{r}(u,v)] \cdot \frac{\partial \boldsymbol{r}}{\partial u} \times \frac{\partial \boldsymbol{r}}{\partial v} du\, dv$$

where $\boldsymbol{n} = \frac{\boldsymbol{N}}{\|\boldsymbol{N}\|}$.
This effectively turns this into a scalar field, where the value is the magnitude of the vector in the direction normal to the surface.
For $\boldsymbol{f}(x,y,z) = P(x,y,z)\boldsymbol{i} + Q(x,y,z)\boldsymbol{j} + R(x,y,z)\boldsymbol{k}$ and $\boldsymbol{r}(u,v) = X(u,v)\boldsymbol{i} + Y(u,v)\boldsymbol{j} + Z(u,v)\boldsymbol{k}$:

$$\begin{aligned}\iint\limits_S \boldsymbol{F} \cdot \boldsymbol{n}\, dS &= \iint\limits_T P[\boldsymbol{r}(u,v)] \frac{\partial(Y,Z)}{\partial(u,v)} + \iint\limits_T Q[\boldsymbol{r}(u,v)] \frac{\partial(Z,X)}{\partial(u,v)} + \iint\limits_T R[\boldsymbol{r}(u,v)] \frac{\partial(X,Y)}{\partial(u,v)} \\ &= \iint\limits_S P(x,y,z) dy \wedge dz + \iint\limits_S Q(x,y,z) dz \wedge dx + \iint\limits_S = R(x,y,z) dx \wedge dy\end{aligned}$$

So basically,

$$\iint\limits_S P\, dy \wedge dz = \iint\limits_T P[\boldsymbol{r}(u,v)] \frac{\partial(Y,Z)}{\partial(u,v)} du\, dv$$

Stokes' theorem

$$\oint_C P\, dx + Q\, dy + R\, dz = \iint\limits_S \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right) dy \wedge dz + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right) dz \wedge dx + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right) dx \wedge dy$$

Or,

$$\oint_C \boldsymbol{F} \cdot d\boldsymbol{\alpha} = \iint\limits_S (\operatorname{curl} \boldsymbol{F}) \cdot \boldsymbol{n}\, dS$$

where $\boldsymbol{\alpha}(t) = \boldsymbol{r}(\boldsymbol{\gamma}(t)]$, where $\boldsymbol{\gamma}(t)$ parameterizes Γ.

Stokes' theorem extended
For non-simply connected regions:

$$\oint_C \boldsymbol{F} \cdot d\boldsymbol{\rho} - \sum_{k=1}^{n} \oint_{C_k} \boldsymbol{F} \cdot d\boldsymbol{\rho_k} = \iint\limits_S (\operatorname{curl} \boldsymbol{F}) \cdot \boldsymbol{n}\, dS$$

where $\boldsymbol{\rho_k}(t) = \boldsymbol{r}[\boldsymbol{\gamma_k}(t)]$ and where $\boldsymbol{\gamma_k}(t)$ is the parameterization of Γ_k.

Curl
Take $\boldsymbol{F}(x,y,z) = P(x,y,z)\boldsymbol{i} + Q(x,y,z)\boldsymbol{j} + R(x,y,z)\boldsymbol{k}$.

$$\operatorname{curl} \boldsymbol{F} = \left(\frac{\partial R}{\partial y} - \frac{\partial Q}{\partial z}\right)\boldsymbol{i} + \left(\frac{\partial P}{\partial z} - \frac{\partial R}{\partial x}\right)\boldsymbol{j} + \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y}\right)\boldsymbol{k}$$

$$\operatorname{curl} \boldsymbol{F} = \begin{pmatrix} \boldsymbol{i} & \boldsymbol{j} & \boldsymbol{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{pmatrix}$$

Divergence
$\mathrm{div}\boldsymbol{F} = \mathrm{tr}\, D\boldsymbol{F}(x, y, z)$ (The trace of the Jacobian matrix.)

Laplacian
$\nabla^2\varphi = \mathrm{div}(\mathrm{grad}\,\varphi) = \frac{\partial^2\varphi}{\partial x^2} + \frac{\partial^2\varphi}{\partial y^2} + \frac{\partial^2\varphi}{\partial z^2}$.
Harmonic function: $\nabla^2\varphi = 0$.

Symbolic formulas
$\nabla = \frac{\partial}{\partial x}\boldsymbol{i} + \frac{\partial}{\partial y}\boldsymbol{j} + \frac{\partial}{\partial z}\boldsymbol{k}$
$\mathrm{curl}\,\boldsymbol{F} = \nabla \times \boldsymbol{F}$
$\mathrm{div}\boldsymbol{F} = \nabla \cdot \boldsymbol{F}$
$\mathrm{grad}\,\varphi = \nabla\varphi$
$\boldsymbol{F}$ is a gradient on an open convex set iff $\mathrm{curl}\,\boldsymbol{F} = O$.

Other relationships
If φ is a scalar field with continuous second-order mixed partial derivatives, $\mathrm{curl}(\mathrm{grad}\,\varphi) = O$.
Irrotational: A vector field with $\mathrm{curl}\,\boldsymbol{F} = 0$.
If $\boldsymbol{F}$ is a vector field with continous mixed partial dervatives, $\mathrm{div}(\mathrm{curl}\,\boldsymbol{F}) = 0$ and $\mathrm{curl}(\mathrm{curl}\,\boldsymbol{F}) = \mathrm{grad}(\mathrm{div}\boldsymbol{F}) - \nabla^2\boldsymbol{F}$.
Curl and divergence are linear operators.
Let $\boldsymbol{F}$ be continously differentiable on an open interval S in 3-space.
Then there exists a vector field $\boldsymbol{G}$ such that $\mathrm{curl}\,\boldsymbol{G} = \boldsymbol{F}$ iff $\mathrm{div}\boldsymbol{F} = 0$ everywhere in S.
Solenoidal: A vector field with $\mathrm{div}\boldsymbol{F} = 0$.

Divergence theorem
Let V be a solid in 3-space bounded by an orientable closed surface S.
Let $\boldsymbol{n}$ be the unit outer normal to S.
Let $\boldsymbol{F}$ be a continously differentiable vector field defined on V.

$$\iint\limits_S \boldsymbol{F} \cdot \boldsymbol{n}\, dS = \iiint\limits_V (\mathrm{div}\boldsymbol{F}) dx\, dy\, dz$$