Lecture 12: Revision

Robert Garrard

Binary Relations

A **binary relation**, (X, Y, G) is an ordered triple where the set X is called the domain, the set Y is called the codomain, and the set G, called the graph, is a subset of the cartesian product of these sets, $G \subseteq X \times Y$.

The graph is a set of ordered pairs, $G = \{(x, y) \mid x \in X, y \in Y\}$. If the pair $(x, y) \in G$, we often write xRy and say "x relates to y".

When the domain and codomain are the same set, the relation *may* have some of the following interesting properties:

Binary Relations

Properties of Relations

Reflexive if $\forall x \in X \quad xRx$

Irreflexive if $\forall x \in X \quad \neg xRx$

Symmetric if $xRy \Rightarrow yRx \quad \forall x, y \in X$

Anti-symmetric if $xRy \wedge yRx \Rightarrow x = y \quad \forall x, y \in X$

Asymmetric if $xRy \Rightarrow \neg yRx \quad \forall x, y \in X$

Transitive if $xRy \wedge yRz \Rightarrow xRz \quad \forall x, y, z \in X$

Weakly Connected if $xRy \vee yRx \quad \forall x \neq y \in X$

Complete if reflexive and weakly connected.

Exercise 12.1 Consider the game Rock-Paper-Scissors. Let $X = \{R, P, S\}$ and xRy be the relation representing x beats y. Which of the above properties does this relation have?

Functions

A **function**, f, is a binary relation, (X, Y, G), such that:

1. $\forall x \in X \; \exists y \in Y \;\; s.t \; (x, y) \in f$
2. $(x, y) \in G \wedge (x, z) \in G \Rightarrow y = z$

The **image** of a set A under a function f is the set

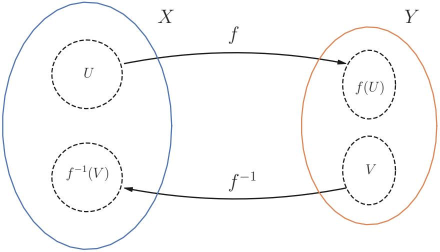
$$f(A) = \{y \in Y \mid \exists x \in A \text{ s.t. } f(x) = y\}$$

The **pre-image** of a set B under a function f is the set

$$f^{-1}(B) = \{x \in X \mid f(x) \in B\}$$

The **range** of a function, $im(f)$, is the image of its domain.

Functions



A function $f : X \to Y$ is said to be an

Injection (one-to-one) if $f(a) = f(b) \Rightarrow a = b \;\; \forall a, b \in X$

Surjection (onto) if $\forall y \in Y \; \exists x \in X \; s.t \; f(x) = y$

Bijection if it is both injective and surjective

Functions that are bijective have an **inverse function**.

Functions

Exercise 12.2

a) Show that the composition of two bijetive functions is bijective.

b) Over what subset of $\mathbb{R}$ does $f(\theta) = \sin\theta$ have an inverse?

Metric Spaces

A **metric space** is a pair (X, d) where X is a set and $d : X \times X \to \mathbb{R}$ is a function which satisfies the following conditions

1. $d(x, y) \geq 0 \ \ \forall x, y \in X$ and $d(x, y) = 0 \Leftrightarrow x = y$
2. $d(x, y) = d(y, x) \ \ \forall x, y \in X$
3. $d(x, z) \leq d(x, y) + d(y, z) \quad \forall x, y, z \in X \ \ (\textit{Triangle inequality})$

One advantage of being able to talk about distance is that we could define **continuity** for functions:

$\varepsilon - \delta$ **Continuity**

$\forall \varepsilon > 0 \ \ \exists \delta > 0 \ \ s.t \ \ d(x, c) < \delta \Rightarrow d\left(f(x), f(c)\right) < \varepsilon$

Open Sets

It was sometimes useful to talk about all the points *within* some distance of another point.

An **open ball** of radius $\varepsilon > 0$ centered at a is the set:

$$B(a, \varepsilon) = \{x \in X \mid d(x, a) < \varepsilon\}$$

We defined a set to be **open** if every point had a little neighborhood it could move around in and stay inside the set:

$$\forall x \in X \quad \exists \varepsilon > 0 \;\; s.t \;\; B(x, \varepsilon) \subseteq X$$

A set X is **closed** if its complement is open.

Open Sets

One strategy for showing that a particular set is open (closed) is to show that you can place an open ball around any point in the set (set's complement) that's contained in the set (complement).

Another was to exploit some useful properties:

- The union of any collection of open sets is open.
- The intersection of any collection of closed sets is closed.
- The intersection of a finite collection of open sets is open.
- The union of a finite collection of closed sets is closed.

Exercise 12.3 Show that the closed ring centered at the origin, $\{x \in \mathbb{R}^2 \mid r \leq d(0, x) \leq R\}$, is a closed set.

Sequences

A **sequence** is a function whose domain is the natural numbers.

A sequence, $\{x_n\}_{n\in\mathbb{N}}$, is said to **converge** to a point a if

$$\forall\varepsilon > 0 \quad \exists N \in \mathbb{N} \;\; s.t \;\; \forall n > N \;\; d(x_n, a) < \varepsilon$$

or in other words:

Every ε-ball centered at a must contain *all but finitely many* elements of the sequnce.

Convergence of Sequences

Hypothesizing that a sequence converges to a particular point a and showing that the definition is satisfied is one strategy for proving convergence. But we had some shortcuts:

Theorem (Monotone Convergence Theorem)

A monotone sequence of real numbers converges iff it is bounded. Further, if it is increasing (resp. decreasing), it converges to its supremum (infimum).

Convergence of Sequences

A sequence $\{x_n\}$ is called a **Cauchy sequence** if:

$$\forall \varepsilon > 0 \; \exists N \in \mathbb{N} \text{ s.t. } \forall m, n > N \; d(x_m, x_n) < \varepsilon$$

Theorem (Cauchy Convergence Criterion)

A sequence of real numbers is convergent iff it is Cauchy

Exercise 12.4 Show that the sequence $x_0 = 100$, $x_{k+1} = \sqrt{x_k}$ is convergent.

Completeness

It would be very convenient if every Cauchy sequence were convergent. Such a metric space is said to be **complete**.

What conditions are sufficient to ensure this?

Maybe closedness? We know the following property of closed sets

Proposition: Closed sets contain their limit points

Let (X, d) be a metric space. If $A \subseteq X$ is a closed subset and $a_n \in A$ is a sequence, then

$$a_n \to a \Rightarrow a \in A$$

But we saw that that wasn't enough (e.g., sequnce of rationals converging to $\sqrt{2}$).

Compactness

A set is **compact** if *every* open cover has a finite sub-cover.

That does the trick, and a few others:

- A compact metric space is complete.
- A closed subset of a compact set is compact.
- A compact subset of R has a maximum and minimum value.

In $\mathbb{R}^n$, the Heine-Borel theorem tells us that the compact sets are ones that are **closed** and **bounded**.

Compactness

With these concepts, we saw how they applied to micro theory, particularly representing preferences and showing that utility maximization problems had solutions.

We also got a powerful fixed point theorem that would later show up in dynamic programming.

Theorem (Banach Fixed-point Theorem)

Let (X, d) be a non-empty complete metric space with a contraction mapping $f : X \to X$. Then f admits a unique fixed-point x^*, $f(x^*) = x^*$.

Furthermore, x^* can be found by iterating f on an arbitrary initial value.

Linear Algebra

In linear algebra, we first started out being interested in finding the solutions to a system of linear equation:

$$\begin{array}{ccccccccc} a_{11}x_{11} & + & a_{12}x_{12} & + & \cdots & + & a_{1n}x_{1n} & = & b_1 \\ a_{21}x_{21} & + & a_{22}x_{22} & + & \cdots & + & a_{2n}x_{2n} & = & b_2 \\ a_{31}x_{31} & + & a_{32}x_{32} & + & \cdots & + & a_{3n}x_{3n} & = & b_3 \\ \vdots & + & \vdots & + & \cdots & + & \vdots & = & \vdots \\ a_{m1}x_{m1} & + & a_{m2}x_{m2} & + & \cdots & + & a_{mn}x_{mn} & = & b_m \end{array}$$

which we would represent more conveniently using matrices: $Ax = b$.

Linear Algebra

We learned a method that works every time for any system:

Recipe for Solving a System of Linear Equations

1. Use row operations to put the augmented system in reduced row echelon form.
2. Is the last column of the augmented system a pivot column? If yes, there is no solution. If no, proceed.
3. Are there any free variables? If no, read the unique solution off of each row. If yes, proceed.
4. Solve for each basic variable in terms of the free variables and a constant.

The 'put the system in RREF' part, we can make a computer do.

Vector Spaces

We saw that we could get some powerful tools if we thought about matrices as functions. But functions of what?

A **vector space** is a structure formed by a set V (whose elements are called vectors) together with two operations: addition and scalar multiplication.

Vector space axioms

1. *Commutativity of vector addition* $\mathbf{x}+\mathbf{y}=\mathbf{y}+\mathbf{x}$
2. *Associativity of vector addition* $(\mathbf{x}+\mathbf{y})+\mathbf{z}=\mathbf{x}+(\mathbf{y}+\mathbf{z})$
3. *Additive identity:* There is a $\mathbf{0}$ s.t $\mathbf{0}+\mathbf{x}=\mathbf{x}$
4. *Additive inverse:* $\forall\mathbf{x}$ there exists a $-\mathbf{x}$ s.t $\mathbf{x}+-\mathbf{x}=\mathbf{0}$
5. *Associativity of scalar multiplication* $r(s\mathbf{x})=(rs)\mathbf{x}$
6. *Distributivity of scalar sums* $(r+s)\mathbf{x}=r\mathbf{x}+s\mathbf{x}$
7. *Distributivity of vector sums* $r(\mathbf{x}+\mathbf{y})=r\mathbf{x}+r\mathbf{y}$
8. *Scalar identity* $1\cdot\mathbf{x}=\mathbf{x}$

Linear Independence and Span

Now that we can add and scalar multiply the elements of our set, there are new properties we can talk about:
The vectors $v_1, \ldots, v_n$ are **linearly independent** if, for scalars $c_1, \ldots, c_n \in \mathbb{R}$, the only solution to the equation

$$c_1 v_1 + \cdots + c_n v_n = 0$$

The vectors $v_1, \ldots, v_n$ are **spanning** if for every vector $v \in V$, we can find $c_1, \ldots, c_n \in \mathbb{R}$ such that

$$v = c_1 v_1 + \cdots + c_n v_n$$

In which case we may write $V = \text{span}\{v_1, \ldots, v_n\}$.

Linear Independence and Span

How do you check if a set of vectors are linearly independent and spanning?

Place the vectors as columns of a matrix: $[v_1, \ldots, v_n]$.

Put the matrix in RREF.

To check linear independence, check if there are n basic columns.

To check spanning, check that every row has a pivot.

If the matrix is square, you can do both at once by checking the determinant.

The vectors $v_1, \ldots, v_n$ form a **basis** for V if they are linearly independent and spanning.

A vector space is finite dimensional with dimension n if we can find a vectors $v_1, \ldots, v_n \in V$ which form a basis for V.

Subspaces

A subset $W \subset V$ is called a **subspace** if and only if

1. $\mathbf{0} \in W$
2. If $\mathbf{u}, \mathbf{v} \in W$ then $\mathbf{u} + \mathbf{v} \in W$
3. If $\mathbf{u} \in W$ then $\alpha\mathbf{u} \in W$ for any scalar $\alpha \in \mathbb{R}$

$$\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{x} = \mathbf{c} + \mathbf{v} \text{ for some } \mathbf{v} \in V\}$$

is called an **affine subspace** of $\mathbb{R}^n$ (but it is not a subspace!).

Matrices as Functions

A function $f : \mathbb{R}^m \to \mathbb{R}^n$ is a **linear map** if:

- $f(x+y) = f(x) + f(y) \;\; \forall x, y \in V.$
- $f(\alpha x) = \alpha f(x) \;\; \forall \alpha \in \mathbb{R},\; x \in V.$

Proposition

Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be a linear map. Further, let the vectors of $\mathbb{R}^m$ and $\mathbb{R}^n$ be column vectors. Then there exists a matrix A such that

$$f(\mathbf{v}) \; = A\mathbf{v}$$

Exercise 12.5 Consider the function which takes a vector in $\mathbb{R}^2$ and returns its mirror image about the y-axis. Is this a linear mapping? What matrix corresponds to it?

Matrices as Functions

If matrices are linear functions, do we have analogues of the image, zeros, and fixed points?

Yes, and these are subspaces!

The image of a matrix is its **column space**,
$\text{Col}(A) = \{b \in \mathbb{R}^n \mid Ax = b \text{ for some } x\}$.

The set of zeros of a matrix is its **null space**,
$\text{Null}(A) = \{x \in \mathbb{R}^n \mid Ax = 0\}$.

The rank-nullity theorem connects these and the total number of columns, n, by:

$$\dim \text{Col}(A) + \dim \text{Null}(A) = n$$

Solutions to Systems Again

Theorem

If $\mathcal{S}$ is non-empty, such that there is at least one particular solution, $\mathbf{x}^*$, such that $A\mathbf{x}^* = \mathbf{b}$. Then $\mathcal{S}$ is the affine subspace

$$\mathcal{S} = \{\mathbf{x}^* + \mathbf{v} \mid \mathbf{v} \in Null(A)\}$$

Exercise 12.6 Find the solutions to the augmented system

$$\begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} x = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

Show that the solution set is an affine subspace.

Eigenvalues and Eigenvectors

For the analogue of fixed points we have:

Let $A : \mathbb{R}^n \to \mathbb{R}^n$ be a linear map. An **eigenvector** $\mathbf{v}$ of A and its corresponding **eigenvalue**, λ, are a vector and scalar satisfying

$$A\mathbf{v} = \lambda\mathbf{v}$$

We can find the eigenvalues by solving:

$$\det(\lambda\mathbf{I} - A) = 0$$

We can solve for the eigenvectors by solving the augmented system;

$$(\lambda I - A \mid 0)$$

for each λ. For small systems we can use the trick $tr(A) = \sum \lambda_i$ and $det(A) = \Pi\lambda_i$.

Diagonalization

Theorem

Let A be an $n \times n$ matrix. Let $\lambda_1, \ldots, \lambda_n$ be eigenvalues of A with corresponding eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n$. Construct the matrix

$$P = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$$

whose columns are A's eigenvectors. If P in invertible, then

$$P^{-1}AP = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ldots & \vdots \\ 0 & 0 & 0 & \lambda_n \end{pmatrix}$$

Diagonalization

Theorem (The Spectral Theorem)

Let A be an $n \times n$ matrix of real entries that is symmetric (i.e., $A' = A$). Then A can be diagonalized

$$A = QDQ'$$

where Q is an orthogonal matrix ($Q^{-1} = Q'$).

Diagonalization

We saw how we can use this property to solve homogeneous first-order difference equations, $\mathbf{x_{t+1}} = A\mathbf{x}_t$.

$$\mathbf{x}_t = c_1\lambda_1^t\mathbf{v}_1 + \cdots + c_n\lambda_n^t\mathbf{v}_n$$

Quadratic Forms

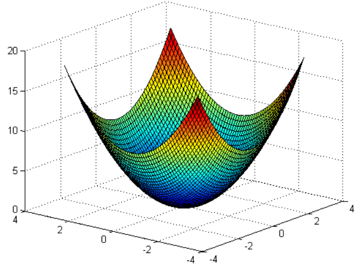
$$\mathcal{Q}(x_1, x_2, \ldots, x_n) = \sum_{i \leq j} a_{ij} x_i x_j$$

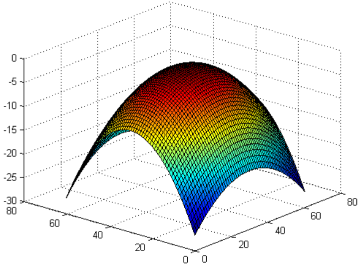
Any quadratic form can be written more compactly as

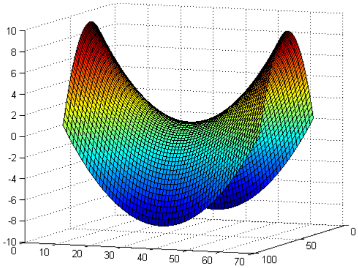
$$\mathcal{Q}(x_1, \ldots, x_n) = \mathbf{x}' A \mathbf{x}$$

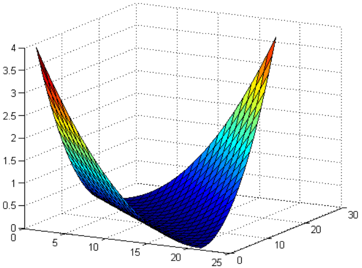
where A is a unique real-valued symmetric matrix.

Quadratic Forms









Definiteness

Theorem

Let A be a symmetric matrix. Then

a) *A is positive definite iff all eigenvalues are > 0.*

b) *A is negative definite iff all eigenvalues are < 0.*

c) *A is positive semidefinite iff all eigenvalues are ≥ 0.*

d) *A is negative semidefinite iff all eigenvalues are ≤ 0*

e) *A is indefinite iff it has at least one positive and one negative eigenvalue.*

Axioms of Probability

A **probability** is a function, $\mathbb{P} : 2^\Omega \to [0, 1]$ such that

1. $0 \leq \mathbb{P}(E) \leq 1$
2. $\mathbb{P}(\Omega) = 1$
3. For any countable sequence of *disjoint* events E_i,

$$\mathbb{P}\left(\bigcup_{i=1}^{\infty} E_i\right) = \sum_{i=1}^{\infty} \mathbb{P}(E_i)$$

Probability *measures* the size of sets in a particular way.

NB: This is not entirely correct. The domain of the function is not the whole power set but a subset that is a “measurable space” and $\mathbb{P}$ must be a “measurable function”, but this is beyond our scope.

Properties of Probability

Let A and B be events. The probability of A **conditional** on having observed B is

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Two events A and B are said to be **independent** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$$

The **law of total probability**:

$$\mathbb{P}(A) = \sum_{i=1}^{n} \mathbb{P}(A|B_i)\mathbb{P}(B_i)$$

and **Bayes' theorem**

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A)\mathbb{P}(A)}{\mathbb{P}(B)}$$

Expectation

Rather than dealing with outcomes, we want to deal with **random variables**, $f : \Omega \to \mathbb{R}$.

$$\mathbb{E}[X] = \sum_{x} x \cdot \mathbb{P}(x)$$

$$\mathbb{E}[X] = \int_{-\infty}^{\infty} x f(x) \mathrm{d}x$$

For this to be well defined, we require that $\mathbb{E}[|X|] < \infty$.

Moments

The k-th **central moment** is defined as

$$\mu_k = \mathbb{E}[(X - \mathbb{E}[X])^k]$$

The k-th **raw moment** is defined as

$$\mu'_k = \mathbb{E}[X^k]$$

The **variance** is the second central moment.

Moment Generating Functions

If all of the moments exist, we can talk about the **moment generating function**.

$$M_X(t) = \mathbb{E}[e^{tX}]$$

The MGF has the following useful properties:

1. If random variables X and Y have the same MGF, they have the same distribution.
2. If X and Y are independent random variables, then their sum $X + Y$ has MGF: $M_{X+Y}(t) = M_X(t) \cdot M_Y(t)$.
3. $M_{cX}(t) = M_X(ct)$.
4. $\mu'_k = \frac{\mathrm{d}^n M_X}{\mathrm{d}t^n}\Big|_{t=0}$

Jointly Distributed Random Variables

$$F(a,b) = \mathbb{P}(X \le a, Y \le b) \qquad -\infty < a, b < \infty$$

$$\mathrm{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

We get **correlation** by scaling covariance by the appropriate variances.

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$$

Properties of Multivariate Normal

Proposition

Let $\mathbf{x} \sim N(\mu, \mathbf{\Sigma})$. Then for some non-zero matrix A (or vector) we have

$$\mathbf{Ax} \sim N(\mathbf{A}\mu, \mathbf{A\Sigma A}')$$

Theorem

Let $\mathbf{x} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$ *and let* M *be a symmetric idempotent matrix of rank* m*. Then*

$$\frac{\mathbf{x}'M\mathbf{x}}{\sigma^2} \sim \chi^2(m)$$

Inequalities

Theorem (Markov's Inequality)

If X is a random variable that takes only nonnegative values, then for any value $a > 0$

$$\mathbb{P}(X \geq a) \leq \frac{\mathbb{E}[X]}{a}$$

Theorem (Chebyshev's Inequality)

If X is a random variable with mean μ and variance σ^2, then for any $\varepsilon > 0$

$$\mathbb{P}(|X - \mu| > \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

Theorem (Jensen's Inequality)

Let X be a random variable and $f(\cdot)$ be a convex function. Then

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)]$$

Estimation

Let F be a distribution with parameter θ. An **estimator** is a function of a sample from F.

$$\hat{\theta} = T(X_1, \ldots, X_n; F)$$

An estimator is **unbiased** if $\mathbb{E}[\hat{\theta}] = \theta$.

An estimator, $\hat{\theta}$, is **consistent** for a parameter θ if $\hat{\theta} \xrightarrow{p} \theta$.

Convergence of Random Variables

We had some different modes of convergence for random variables:
$\{X_n\}$ **converges in probability** to X, if
$\forall \varepsilon > 0 \ \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$.

$\{X_n\}$ **converges almost surely** to X, if $\mathbb{P}(\lim_{n\to\infty} X_n = X) = 1$.

$\{X_n\}$ **converges in distribution** to X, if $F_n(x) \to F(x) \quad \forall x$.

Convergence almost surely $\Rightarrow$ convergence in probability $\Rightarrow$ convergence in distribution

Convergence of Random Variables

Theorem (Continuous Mapping Theorem)

Let $g(\cdot)$ be a continuous function. Then for $i \in \{d, p, a.s.\}$

$$X_n \xrightarrow{i} X \Rightarrow g(X_n) \xrightarrow{i} g(X)$$

Theorem (Slutsky's Theorem)

Let $X_n \xrightarrow{d} X$ and $Y_n \to c$.

$$X_n + Y_n \xrightarrow{d} X + c$$

$$X_n Y_n \xrightarrow{d} Xc$$

$$X_n / Y_n \xrightarrow{d} X/c, \textit{provided } c \not\to 0\ .$$

Limit Theorems

We frequently appeal to one of the following (family of) laws:

Theorem (law of large numbers)

Let $X_1, \ldots, X_n \sim F$ be an iid sample from a distribution with mean μ.

$$\bar{X}_n = \frac{1}{n} \sum X_i \xrightarrow{p} \mu$$

Theorem (Central Limit Theorem)

Let $X_1, \ldots X_n \overset{iid}{\sim} F$, with $\mathbb{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$, both finite. Let $\bar{X}_n = \frac{1}{n} \sum X_i$.

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

Limit Theorems for Random Vectors

Theorem (Multivariate CLT)

Suppose $\mathbf{X}$ *is a random vector with finite mean* μ *and variance* $\mathbf{\Sigma}$. *Then*

$$\sqrt{n}\left(\bar{\mathbf{X}}_n - \mu\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{\Sigma})$$

Using Slutsky's theorem, if $\mathbf{A}_n \xrightarrow{p} A$ then $\sqrt{n}\mathbf{A}_n\left(\bar{\mathbf{X}}_n - \mu\right) \xrightarrow{d} N(\mathbf{0}, \mathbf{A}\mathbf{\Sigma}\mathbf{A}')$.

Taylor Series

Smooth functions admit a Taylor series.

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

We can use a few terms of the series and be confident our error is small if we're close to the point we expanded the series around.

Theorem

Let $f : \mathbb{R} \to \mathbb{R}$ be k times differentiable at the point $a \in \mathbb{R}$. Then there exists a function $R_k : \mathbb{R} \to \mathbb{R}$ such that

$$f(x) = f(a) + f'(a)(x-a) + \frac{f''(a)}{2}(x-a)^2 + \cdots + \frac{f^{(k)}}{k!}(x-a)^k + R_k(x)$$

such that

$$lim_{x \to a} R_k(x)/x^k = 0$$

Multivariate Functions

A **level set** of a function $f : \mathbb{R}^n \to \mathbb{R}$ is the set

$$\mathcal{L} = \{\mathbf{x} \in \mathbb{R}^n \mid f(\mathbf{x}) = c\}$$

for some $c \in \mathbb{R}$.

The **gradient** is the vector of partial derivatives:

$$\nabla f(\mathbf{x}) = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n}\right)$$

At any point, the tangent to the level set and the gradient vector are orthogonal.

Log-linearization

In log-linearizing an equation, we want to transform it from being in levels to being in percentage deviation from the steady state

$$\hat{x}_t = \frac{x_t - \bar{x}}{\bar{x}}$$

For most equations, we won't be able to write it like this exactly. We'll need to write it approximately:

$$\begin{aligned} f(x_t) &\approx f(\bar{x}) + f'(\bar{x})(x_t - \bar{x}) \\ &= f(\bar{x}) + f'(\bar{x})\bar{x}\hat{x}_t \end{aligned}$$

Unconstrained optimization

To optimize a function on the interior, set the gradient to zero and solve. $\nabla f(x^*) = 0$.

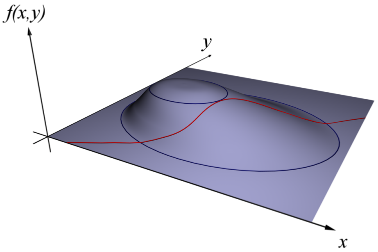
Construct the Hessian and check it's negative (positive) definite for a local maximum (minimum).

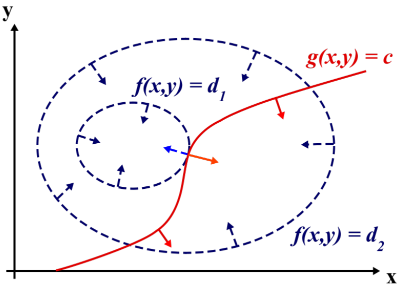
Equality Constrained Optimization

Consider the optimisation problem

$$\max_{\{x_1,x_2\}} \quad f(x_1,x_2)$$
$$\text{s.t} \quad g(x_1,x_2)=h$$

To determine what conditions might be necessary for an optimum, consider these functions and their level sets.





Equality Constrained Optimization

At an optimum, the level set of the function and the constraint set are tangent.

$$\nabla f(\mathbf{x}^*) = \lambda \nabla g(\mathbf{x}^*)$$

However, for λ to be uniquely defined, we need to require that $\nabla g(\mathbf{x}^*) \neq 0$.
Construct the Lagrangian

$$\mathcal{L} = f(x) + \lambda[h - g(x)]$$

Solve this unconstrained problem, $\nabla \mathcal{L} = 0$.

Inequality Constrained Optimization

Consider the problem with one inequality constraint

$$\max_{\{x_1,x_2\}} \quad f(x_1,x_2)$$
$$\text{s.t} \quad g(x_1,x_2) \leq h$$

As before, it'll be necessary that at a local max/min, $\mathbf{x}^*$

$$\nabla f(\mathbf{x}^*) = \mu \nabla g(\mathbf{x}^*)$$

But this time we can't take partials with respect to Lagrange multipiliers (that gives us binding constraints). Instead we use the conditions:

$$g((x) \leq h \qquad \mu \geq 0 \qquad \mu[h - g(\mathbf{x})^*] = 0$$

The Principle of Optimality

Principle of Optimality (Bellman, 1957)

An optimal policy has the property that whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.

Tells us to look for a recursive formulation of the problem.

$$V(k_t) = \max_{c_t} \quad \{U(c_t, k_t) + \beta V(g(c_t, k_t))\}$$

This is not a sequence problem but a **functional equation** problem. The solution is a pair of functions: $\sigma(k_t)$ and $V(k_t)$.

Properties (Stokey, Lucas, Prescott)

If

A1. The constraint set $\{x \mid x = g(c_t, k_t)\}$ is non-empty for all k_t.

A2. The constraint set is compact and $f(\cdot)$ and $g(\cdot)$ are continuous.

A3. $f(\cdot)$ is strictly concave. The constraint set is convex.

A4. $f(\cdot)$ is strictly increasing.

A5. $f(\cdot)$ is $\mathcal{C}^1$ on the interior of its domain.

then there are unique value and policy functions that solve the problem, and the solution is identical to the sequence problem's solution. The value function is strictly increasing, strictly concave, continuous, and differentiable.

The Bellman Operator

Define the operator $T : X \to X$ (where X is a function space) to be:

$$T(V) = \max_{c} \{f(c,k) + \beta V(g(c,k))\}$$

The value function we're looking for is a fixed point of this operator

$$T(V) = V$$

Contraction Mapping

Theorem (Blackwell's sufficient conditions)

Let $X \subseteq \mathbb{R}^l$ and let $B(X)$ be the space of bounded functions, $f : X \to \mathbb{R}$, with the sup norm. Let $T : B(X) \to B(X)$ be an operator satisfying:

a) *(monotonicity) $f, g \in B(X)$ with $f(x) \leq g(x)$ $\forall x$ implies $(Tf)(x) \leq (Tg)(x)$ $\forall x$.*

b) *(discounting) there exists some $\beta \in (0, 1)$ such that $T(f + a)(x) \leq Tf(x) + \beta a$.*

then T is a contraction mapping.

The Bellman operator is a contraction mapping. We can start from any V^0, and by iterating the operator will arrive at the correct value function.

Good Luck!