Lecture 4: Optimization

Naz Koont

26 August 2021

Today: Optimization

1. Aside to Multivariable Calculus: partial derivatives
2. Aside to Linear Algebra: Quadratic forms
3. Finding extreme values of a function f
    - Necessary and sufficient conditions
    - Convexity and concavity
4. Constrained optimization
    - Optimization over an interval
    - Lagrangian method
5. Principal Component Analysis

Motivation

Many economic models are based on the idea that an individual decision maker makes an *optimal choice* from some set of alternatives.

- To formalize this idea, we interpret optimal choice as maximizing or minimizing the value of some function.
- For example, a firm is assumed to minimize costs of producing each level of output and maximize profit, a consumer maximizes their utility, a policy maker maximizes welfare or GDP, etc.
- Thus, the mathematics of optimization are of central importance in economics.

Luckily, we have almost finished developing the machinery that will be used to optimize functions.

Multivariable Differential Calculus

In our study of analysis, so far we have focused on functions of one variable. In practice, it is often necessary to deal also with functions depending on two, three, or more variables.

Definition (Vector Valued $\gamma : \mathbb{R} \to \mathbb{R}^n$)
The first and simplest generalization of the derivative comes from looking at vector valued functions $\gamma : \mathbb{R} \to \mathbb{R}^n$. These functions are differentiable at a point if each of their component functions are differentiable, and the derivative may be computed by differentiating each component separately.

$$f(x) = (x, x^2) \implies f'(x) = (1, 2x)$$

Multivariable Function: Partial Derivatives

The simplest notion of derivative for a function of several variables $f : \mathbb{R}^n \to \mathbb{R}$ is that of *partial derivatives*, which are just the derivatives of the function with respect to each of its variables when the others are held fixed.

- Once we fix the values of other variables, then our function of multiple variables becomes again a single variable function.

Definition (Partial Derivative)
The partial derivative of a function $f(x_1, ..., x_n)$ with respect to the variable x_j is as below, provided that the limit exists.

$$\lim_{h \to 0} \frac{f(x_1, ..., x_j - h, ..., x_n) - f(x_1, ..., x_j, ..., x_n)}{h}$$

The partial derivative is denoted as $\dfrac{\partial f}{\partial x_j}$, f_{x_j}, f_j, $\partial_{x_j} f$, or $\partial_j f$.

Partial Derivatives

Further, just as in differentiation of a single variable, we can consider *higher order* partial derivatives.

- We can consider differentiating with respect to the same variable again *(pure partial derivatives)*, or differentiating with respect to a different variable *(mixed partial derivatives)*.
- In general, given a function $f : \mathbb{R}^n \to \mathbb{R}$, there are n^2 different second-order partial derivatives.

Theorem (Young; Schwarz)

If a function f is twice-differentiable at x, then for any $i, j \in \{1, ..., n\}$, both $\frac{\partial^2 f}{\partial x_j \partial x_i}(x)$ and $\frac{\partial^2 f}{\partial x_i \partial x_j}$ exist and further are equal.

Multivariable Function: Partial Derivatives

As an example, lets find all the (first-order) partial derivatives of the following function $f : \mathbb{R}^n \to \mathbb{R}$:

$$\begin{aligned} f(x, y, z) &= xyz + z^{-2}e^{y} \\ \frac{\partial f}{\partial x} &= yz \\ \frac{\partial f}{\partial y} &= xz + z^{-2}e^{y} \\ \frac{\partial f}{\partial z} &= xy - 2z^{-3}e^{y} \end{aligned}$$

Now lets analyze the second order partial derivatives

Multivariable Function: Partial Derivatives

$$f(x, y, z) = xyz + z^{-2}e^{y}$$

$$\frac{\partial f}{\partial x} = yz \qquad \frac{\partial f}{\partial y} = xz + z^{-2}e^{y} \qquad \frac{\partial f}{\partial z} = xy - 2z^{-3}e^{y}$$

Now lets analyze the second order partial derivatives:

$$\frac{\partial^2 f}{\partial x \partial y} = \frac{\partial^2 f}{\partial y \partial x} = z$$

$$\frac{\partial^2 f}{\partial x \partial z} = \frac{\partial^2 f}{\partial z \partial x} = y$$

$$\frac{\partial^2 f}{\partial y \partial z} = \frac{\partial^2 f}{\partial z \partial y} = x - 2z^{-3}e^{y}$$

$$\frac{\partial^2 f}{\partial x^2} = 0$$

$$\frac{\partial^2 f}{\partial y^2} = z^{-2}e^{y}$$

$$\frac{\partial^2 f}{\partial z^2} = 6z^{-4}e^{y}$$

Gradient of Partial Derivatives

Definition (Hessian Matrix)
For a function $f : \mathbb{R}^n \to \mathbb{R}$, the gradient is the vector-valued function ∇f ($\nabla =$ nabla or grad) whose value at x is the vector whose components are the partial derivatives of f at x

$$\nabla f(x) \in \mathbb{R}^n = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x) \\ \frac{\partial f}{\partial x_2}(x) \\ \dots \\ \frac{\partial f}{\partial x_n}(x) \end{bmatrix}$$

The gradient vector can be interpreted as the direction and rate of fastest increase.

Hessian Matrix of Second Derivatives

Definition (Hessian Matrix)
For a function $f : \mathbb{R}^n \to \mathbb{R}$, if all second partial derivatives of f exist and are continuous over the domain of the function, then the Hessian matrix H of f is a square $n \times n$ matrix:

$$H_f(x) \in \mathbb{R}^{n\times n} = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(x) & \frac{\partial f}{\partial x_1 \partial x_2}(x) & \dots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(x) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(x) & \frac{\partial^2 f}{\partial x_2^2}(x) & \dots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(x) \\ \dots & \dots & \dots & \dots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(x) & \dots & \dots & \frac{\partial^2 f}{\partial x_n^2}(x) \end{bmatrix}$$

By Shwarz' theorem, we know that the Hessian matrix is symmetric!

Quadratic Forms

Definition (Quadratic Form)
A quadratic form on $\mathbb{R}^n$ is a function $Q : \mathbb{R}^n \to \mathbb{R}$ that can be represented by a unique symmetric matrix A as follows:

$$Q(x) = x^T A x = \sum_{i=1}^{n} x_i (Ax)_i = \sum_{i=1}^{n} x_i \left(\sum_{j=1}^{n} A_{ij} x_j \right) = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} x_i x_j$$

For example, given the matrix:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}$$

We can define the quadratic form $q(x) : \mathbb{R}^4 \to \mathbb{R}$

$$q(x) = x^T A x = \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} x_1 a_{11} + x_2 a_{21} & x_1 a_{12} + x_2 a_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$q(x) = a_{11} x_1^2 + a_{21} x_1 x_2 + a_{12} x_1 x_2 + a_{22} x_2^2$$

If A is not symmetric ($a_{21} \neq a_{12}$), we can always define a symmetric matrix A^* that yields the same quadratic form as A, $a_{ij}^* = a_{ji}^* = \frac{1}{2}(a_{ij} + a_{ji})$

Quadratic Forms

Definition (Positive/Negative Definite)

1. If $q(x) = x^T Ax > 0$ for all $x \neq 0$, $q(x)$ is said to be positive definite and A a positive definite matrix.
2. If $q(x) = x^T Ax \geq 0$ for all $x \neq 0$, $q(x)$ is said to be positive semidefinite and A a positive semidefinite matrix.
3. If $q(x) = x^T Ax < 0$ for all $x \neq 0$, $q(x)$ is said to be negative definite and A a negative definite matrix.
4. If $q(x) = x^T Ax \leq 0$ for all $x \neq 0$, $q(x)$ is said to be negative semidefinite and A a negative semidefinite matrix.

If a quadratic form is positive for some x and negative for some other x, it is said to be **indefinite**, as is the matrix that defines it.

Theorem (Positive/Negative Semi-Definite Matrix)

Let $A \in \mathbb{R}^{n \times n}$ a symmetric matrix. The matrix A is:

1. *Positive definite iff all of its eigenvalues are positive.*
2. *Negative definite iff all of its eigenvalues are negative.*
3. *Positive semi-definite iff all of its eigenvalues are non-negative.*
4. *Negative semi-definite iff all of its eigenvalues are non-positive.*
5. *Indefinite iff it has both positive and negative eigenvalues.*

Optimizing a function f

Given some function $f(x)$, we optimize it by finding a value of x at which it takes on a maximum or minimum value, i.e. an *extreme value*.

- If we are optimizing over the entire real line $\mathbb{R}$, we say that we are *unconstrained*.
- If we focus on some strict subset $[a, b] \subset \mathbb{R}$, we say that we are conducting *constrained* optimization.

It is possible that a function does not have a minimum or maximum value.

- An example is a linear function $y = a + bx$ for $a, b > 0$.
- Similarly, a parabola $y = x^2$ has a minimum at $x = 0$ but no maximum.

Optimizing a function f

Definition (Global and Local Maximum)
At a global maximum x^*,

$$f(x^*) \geq f(x), \forall x \in \mathbb{R}$$

Whereas at a local maximum x^*,

$$f(x^*) \geq f(x), \forall x \in B_\epsilon(x) \text{ for } \epsilon > 0$$

While we are generally interested in finding a global maximum, the methods we have for optimizing functions will only guarantee that we are able to find local maximum (if they exist).

In practice, many economic problems have built in assumptions to ensure that there is only one local maximum, in which case it is also the global maximum.

Finding Extreme Values of f

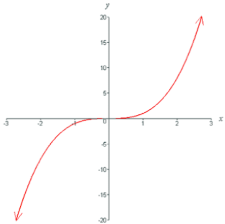
Using the first derivative of a function, we can determine some necessary conditions for the existence of a local extreme point.

Theorem (First Order Conditions)

Let f a function defined on (a, b). If x is a maximum or minimum point for f on (a, b), and f is differentiable at x, then $f'(x) = 0$. (Converse is not true)

The intuition here is that if the derivative is not 0, then on some small interval containing x, $B_\epsilon(x)$, the function will be strictly increasing or decreasing, contradicting that x can be an extreme point of f.

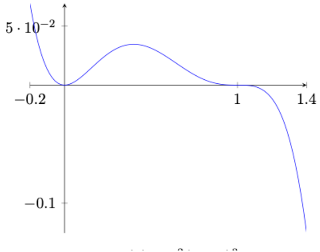
To see however that first order conditions are not sufficient to guarantee the existence of an extreme value, consider the function $f(x) = x^3$, the cubic function. We have that $f'(x)|_{x=0} = 3x^2 = 0$, however 0 is an *inflection point* for f, not a local minima or maxima.



Finding Extreme Values of f

Lets look at another example:

$$f(x) = x^2(1-x)^3$$
$$\implies f'(x) = 2x(1-x)^3 + 3x^2(1-x)^2(-1) = 0$$
$$2x(1-x)^3 = 3x^2(1-x)^2 \text{ thus } x = 0, 1 \text{ are solutions}$$
$$2(1-x) = 3x \implies x = \frac{2}{5} \text{ is another}$$



However, only $0, \frac{2}{5}$ are local minima/maxima of the function.

Finding Extreme Values of f

So far we have only used information about the first derivative of f. By bringing in information about higher order derivatives, namely the second derivative, we can distinguish whether a given point is or is not an extreme value of the function.

Theorem (Second Order Conditions)

Suppose $f'(a) = 0$. If $f''(a) > 0$, then f has a local minimum at a, and if $f''(a) < 0$ then f has a local maximum at a.

Finding Extreme Values of f

Lets revisit our earlier example:

$$f(x) = x^2(1-x)^3$$

$$\implies f'(x) = 2x(1-x)^3 - 3x^2(1-x)^2 = 0 \text{ has solutions } 0, 1, \frac{2}{5}$$

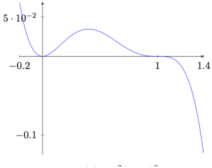
Let's take another derivative to impose our sufficient second-order conditions:

$$f''(x) = 2(1-x)^3 - 12x(1-x)^2 + 6x^2(1-x)$$

$$f''(0) = 2 > 0 \implies \text{ local minimum}$$

$$f''(\frac{2}{5}) = -\frac{149}{125} < 0 \implies \text{ local maximum}$$

$$f''(1) = 0 \implies \text{ indeterminate}$$



Convex Functions

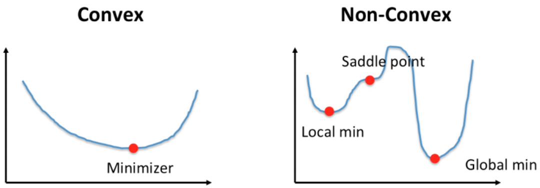
Definition (Convex Function)
A function f is convex on an interval, iff its graph lies above all of its tangent lines: $f(x) \geq f(x) + f'(y)(x - y)$ iff its derivative is monotonically non-decreasing on that interval.

Theorem (Convex Functions)

If f is differentiable and f' is increasing, then f is convex.

In economic applications, convex functions are used to ensure that local minima are global.

- Any local minimum of a convex function is also a global minimum.
- A strictly convex function will have at most one global minimum.



Finding Extreme Values of f: Multivariable

For multivariable functions, the first and second order conditions have analogs in terms of partial derivatives.

Definition (First Order Conditions for $\mathbb{R}^n$)
A stationary value of the function f over the domain $\mathbb{R}^n$ occurs at a point $x \in \mathbb{R}^n$ at which the n equalities below hold simultaneously:

$$f_1(x_1, ..., x_n) = 0$$
$$...$$
$$f_n(x_1, ..., x_n) = 0$$

i.e., $\nabla f(x) = 0$. As before, this is a necessary but not sufficient condition to find a local maxima or minima.

Finding Extreme Values of f: Multivariable

Lets look at an example and find the critical points of the following multivariate function:

$$
\begin{aligned}
f(x, y) &= x^4 - 2x^2 + y^3 - 6y \\
&\implies \nabla f(x, y) = (4x^3 - 4x, 3y^2 - 6) \\
4x^3 &= 4x \implies x = 0 \\
&\implies x^2 = 1 \implies x = \pm 1 \\
&\implies 3y^2 = 6 \implies y = \pm\sqrt{2}
\end{aligned}
$$

Thus we have 6 critical points: $(0, \pm\sqrt{2}), (\pm 1, \pm\sqrt{2})$. These points may be local minima, maxima, or neither.

Finding Extreme Values of f: Multivariable

Similarly, we can think of second order conditions for multivariable functions by requiring that small changes in the x vector along any dimension from this point must reduce the value of f if x is to be a maximum.

Definition (Second Order Conditions for $\mathbb{R}^n$)
It is sufficient for x to yield a local maximum of the continuous function f that $f_i(x) = 0, \forall i 1, ..., n$, and the quadratic form

$$d^2 y(x) = \sum_i \sum_j f_{ij}(x) dx_i dx_j < 0, \quad i, j = 1, ..., n$$

That is, $d^2 y$ is negative definite.

Notes that $d^2 y = x^T H x$; the Hessian matrix H must be negative definite.

Finding Extreme Values of f: Multivariable

Lets apply our sufficient second order conditions to our earlier example:

$$f(x,y) = x^4 - 2x^2 + y^3 - 6y$$
$$\implies \nabla f(x,y) = (4x^3 - 4x, 3y^2 - 6)$$

And we have 6 critical points: $(0, \pm\sqrt{2}), (\pm 1, \pm\sqrt{2})$. Let us compute the Hessian:

$$H_f(x,y) = \begin{bmatrix} 12x^2 - 4 & 0 \\ 0 & 6y \end{bmatrix}$$

The eigenvalues of a diagonal matrix are the diagonal entries. Recall the equivalence between eigenvalues and positive/negative definiteness.

$$12x^2 - 4 > 0 \implies x > \frac{1}{\sqrt{3}} \text{ or } x < -\frac{1}{\sqrt{3}}$$
$$6y > 0 \implies y > 0$$

$(0, -\sqrt{2})$ is a local maximum, $(\pm 1, \sqrt{2})$ are local minima, and the rest are indeterminate/saddle points.

Finding Extreme Values of f: Multivariable

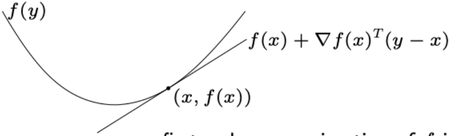
Theorem
Suppose that $f(x)$ is a strictly convex function defined on $x \in \mathbb{R}^n$. If at $x = x^*$ all first derivatives vanish, $f_i(x^*) = 0$, $i = 1, ..., n$ then x^* yields a unique global minimum.

A multivariable function $f : \mathbb{R}^n \to \mathbb{R}$ is convex if and only if:

1. $f(x) \geq f(y) + \nabla f(y)(x - y)$ for all x, y in its domain. (graph of f lies above all of its tangent planes).
2. If twice differentiable, its Hessian matrix is positive semidefinite.

1st-order condition: differentiable f with convex domain is convex iff

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) \quad \text{for all } x, y \in \mathbf{dom}\, f$$



first-order approximation of f is global underestimator

Finding Extreme Values of f: Multivariable

Lets see an example of a strictly convex function over $\mathbb{R}^2$:

$$f(x, y) = x^2 + y^2$$
$$\frac{\partial f}{\partial x} = 2x$$
$$\frac{\partial f}{\partial y} = 2y$$
$$H_f(x, y) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

Positive definite Hessian $\forall (x, y) \in \mathbb{R}^2$

Constrained Optimization

So far we have been dealing with unconstrained optimization, in which the solution to the optimization can lie anywhere on the real line $\mathbb{R}$.

- Often in economics however, this is unacceptably general.
- A key idea of economics is the idea of *scarcity of resources*– the gap between limited resources and theoretically limitless wants.
- The optimization of interest for real world applications is that of allocating finite resources efficiently, i.e. in a way the maximizes utility or welfare for the agents involved.

In order to formalize this mathematically, we consider optimization over a constrained set, such as an interval in $\mathbb{R}$ for functions of one variable.

Optimization over an Interval

When we optimize a function f subject to the constraint that the value lies within some closed interval $[a, b]$, it is no longer a necessary condition that the derivative $f'(x) = 0$ at the extreme value.

This is because the extreme value can additionally occur at the endpoints of the interval, a or b, rather than an *interior solution*.

In fact, there are now three possibilities for a local maximum (or minimum) x:

1. $x = a$. In this case, we must have $f'(x) \leq 0$. *Why?*
2. $a < x < b$. In this case we must have $f'(x) = 0$
3. $x = b$. In this case, we must have $f'(x) \geq 0$.

Lagrangian Method

Suppose that we wish to maximize a function $f(x_1, x_2)$ for strictly concave and continuous f.

- And further, that we impose the constraint $g(x_1, x_2) = 0$ for continuous g.
- This means that we are only allowed to consider as possible solutions to the problem, pairs of x_1, x_2 that satisfy the equation $g = 0$.

We proceed by introducing a new variable λ, the **Lagrange Multiplier**, and by forming the **Lagrangian**:

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda g(x_1, x_2)$$

Lagrangian Method

$$L(x_1, x_2, \lambda) = f(x_1, x_2) + \lambda g(x_1, x_2)$$

We then carry out the unconstrained maximization of L with respect to x_1, x_2, λ to get the following first order conditions:

$$\begin{aligned}
\frac{\partial L}{\partial x_1} &= f_1(x_1, x_2) + \lambda g_1(x_1, x_2) = 0 \\
\frac{\partial L}{\partial x_2} &= f_2(x_1, x_2) + \lambda g_2(x_1, x_2) = 0 \\
\frac{\partial L}{\partial \lambda} &= g(x_1, x_2) = 0
\end{aligned}$$

Thus we have three equations to solve for the three unknowns, x_1, x_2, λ.

The Lagrange multiplier procedure is a way of defining an unconstrained optimization problem, which delivers in a routine way the tangency conditions for an optimal solution.

Lagrangian Method Example

Example Lagrangian Problem: Solve the constrained maximization problem:

$$\max y = x_1^{0.25} x_2^{0.75} \quad \text{such that } 100 - 2x_1 - 4x_2 = 0$$

We write out the Lagrangian function for this problem and take the first-order conditions:

$$L = x_1^{0.25} x_2^{0.75} + \lambda(100 - 2x_1 - 4x_2)$$
$$\frac{\partial L}{\partial x_1} : 0.25 x_1^{-0.75} x_2^{0.75} - 2\lambda = 0$$
$$\frac{\partial L}{\partial x_2} : 0.75 x_1^{0.25} x_2^{-0.25} - 4\lambda = 0$$
$$\frac{\partial L}{\partial \lambda} : 100 - 2x_1 - 4x_2 = 0$$

Lagrangian Method Example

Let us solve this system of equations via substitution:

$$\frac{1}{8}\left(\frac{x_2}{x_1}\right)^{0.75} = \lambda = \frac{3}{16}\left(\frac{x_1}{x_2}\right)^{0.25}$$

$$\frac{x_2}{x_1} = \frac{3}{2}$$

$$x_2 = \frac{3}{2}x_1$$

$$100 - 2x_1 - 6x_1 = 0 \implies \frac{100}{8} = x_1 \implies x_1 = \frac{25}{2} \text{ and } x_2 = \frac{75}{4}$$

How to interpret λ?

Interpretation of λ

We apparently introduced λ just as a placeholder to help generate the conditions which give us the solution to the original constrained optimization problem.

- It turns out however, that λ has a very important and interesting economic interpretation in constrained optimization problems.
- The value of the Lagrange multiplier λ at the optimal solution tells us the effect of the optimized value of the function f of a small relaxation of the constraint
- i.e. the *shadow price* of the constraint.

Lagrange Method with Multiple Constraints

We can extend the Lagrange method to optimization problems with multiple constraints:

$$\max f(x) \text{ such that: } g_1(x) = \ldots = g_m(x) = 0$$

Then we have:

$$L(x, \lambda) = f(x) + \sum_{j=1}^{m} \lambda_j g_j(x)$$

Which gives the $n + m$ conditions:

$$\frac{\partial L}{\partial x_i} = f_i(x^*) + \sum_j \lambda_j^* \frac{\partial g_j}{\partial x_i}(x^*) = 0$$
$$\frac{\partial L}{\partial \lambda_j} = g_j(x^*) = 0$$

Where one restriction we place is that $n > m$ (otherwise there may be no point at which all constraints are simultaneously satisfied).

Existence of Global Maxima in Constrained Optimization

Recall that in Calculus we found sufficient conditions for extreme values of continuous functions:

Theorem (Bolzano-Weierstrass)
If f is a continuous function, and X is a nonempty, closed, and bounded set, then f has both a maximum and a minimum on X.

- The set X is denoted by our constraints.
- Such maxima can be attained in the interior of the set X **such that the constraints do not bind**.
- Or attained on the boundary points **in which case the constraints will bind** (think about a linear function over a closed interval).

In general, the set X can be denoted by inequality constraints (e.g., profit π greater than or equal to \$100) , not just equality constraints. If the constraint does not bind at some x, $x \in X$ is an interior point of the set X.

Existence of Global Maxima in Constrained Optimization

Definition (Kuhn Tucker Conditions)
Assuming that the conditions of Bolzano-Weierstrass hold, when will our Lagrangian method give us the solution with k linearly independent constraints? It must be that x^* satisfies the FOC and further the **complementary slackness condition:**

$$\lambda_j \geq 0, g_j(x) \geq 0, \text{ and } \lambda_j g_j(x) = 0, \forall j = 1, ..., k$$

If a constraint binds with equality ($g_j(x) = 0$), it will be that $\lambda_j > 0$ and if a constraint is slack ($g_j(x) \geq 0$), it will be that $\lambda_j = 0$. Thus there is **complementary** slackness, and always $\lambda_j g_j(x) = 0$.

However often in economic applications we will deal with equality constraints ("more is better") so we will not cover practice problems with inequality constraints

Principal Component Analysis

Principal Component Analysis (PCA) is one of a family of techniques for taking high-dimensional data, and using the dependencies between the variables to represent it in a more tractable, lower dimensional form, without loosing too much information, i.e. to reduce the dimension of the data.

- We begin with p-dimensional vectors with mean 0, and want to summarize them by projecting down into a q-dimensional shape, for $p > q$.
- Our summary will be the projection of the original vectors on to q directions, the **principal components**, which span the subspace.

There are multiple equivalent ways to find these principal components:

1. We can find the projections that maximize the variance.
2. We can find the projection with the smallest mean-squared distance between the original vectors and their projections onto the principal components.

Principal Component Analysis

Via maximizing variance: If we stack our n data vectors into an $n \times p$ matrix x, then the projections are given by $xw \in \mathbb{R}^{n\times 1}$. The variance is:

$$\sigma_\omega^2 = \frac{1}{n}(xw)^T(xw) = \frac{1}{n}w^T x^T xw = w^T \frac{x^T x}{n} w = w^T vw$$

Thus we want to choose a unit vector w to maximize σ_ω^2. Thus our optimization problem is the following:

$$\max_w w^T vw \quad \text{such that} w^T w = 1$$

So we can set up our Lagrangian:

$$L(w, \lambda) = w^T vw + \lambda(1 - w^T w)$$

$$\frac{\partial L}{\partial w} : 2vw - 2\lambda w = 0 \implies vw = \lambda w$$

$$\frac{\partial L}{\partial \lambda} : w^T w = 1$$

Thus, the desired vector w is an **eigenvector** of the covariance matrix v, and the maximizing vector will be the one associated with the largest **eigenvalue** λ. Since $v \in \mathbb{R}^{p\times p}$, we know that it will have p different eigenvectors. The eigenvectors of v are the **principal components** of the data.

Portfolio Optimization

A key topic in financial economics is that of an investor's portfolio choice across assets.

- An investor may wish to maximize the average return of their portfolio while minimizing the corresponding risk, or volatility of their returns.
- We can combine the tools that we have learned so far to solve the classic Markowitz (1952) model of choosing N risky assets to minimize variance for a given level of mean return.

We define $R \in \mathbb{R}^n$ as the vector of mean returns on the n risky assets, Σ as the variance-covariance matrix of returns, ω as the vector of portfolio weights, and 1 as a vector of ones. The problem we wish to solve is the following optimization problem:

$$\min_{\omega} \frac{1}{2}\omega'\Sigma\omega \quad \text{such that } R'\omega = A \text{ and } 1'\omega = 1$$

Portfolio Optimization

$$\min_{\omega} \frac{1}{2}\omega'\Sigma\omega \quad \text{such that } R'\omega = A \text{ and } 1'\omega = 1$$

We set up the Lagrangian:

$$\begin{aligned}
L(\omega, \lambda_1, \lambda_2) &= \frac{1}{2}\omega'\Sigma\omega + \lambda_1(A - R'\omega) + \lambda_2(1 - 1'\omega) \\
\frac{\partial L}{\partial \omega} &= \Sigma\omega - \lambda_1 R - \lambda_2 1 = 0 \\
\frac{\partial L}{\partial \lambda_1} &= A - R'\omega \\
\frac{\partial L}{\partial \lambda_2} &= 1 - 1'\omega \\
\implies \omega &= \lambda_1 \Sigma^{-1} R + \lambda_2 \Sigma^{-1} 1
\end{aligned}$$

We can then solve for the Lagrange multipliers using the two constraints to solve for two unknowns:

$$\begin{aligned}
A = R'\omega &= \lambda_1 R'\Sigma^{-1}R + \lambda_2 R'\Sigma^{-1}1 \\
1 = 1'\omega &= \lambda_1 1'\Sigma^{-1}R + \lambda_2 1'\Sigma^{-1}1
\end{aligned}$$

Portfolio Optimization

$$
\begin{aligned}
A &= R'\omega = \lambda_1 R'\Sigma^{-1}R + \lambda_2 R'\Sigma^{-1}1 \\
1 &= 1'\omega = \lambda_1 1'\Sigma^{-1}R + \lambda_2 1'\Sigma^{-1}1
\end{aligned}
$$

This gives:

$$
\begin{aligned}
\lambda_1 &= \frac{(1'\Sigma^{-1}1)A - R'\Sigma^{-1}1}{(R'\Sigma^{-1}R)(1'\Sigma^{-1}1) - (R'\Sigma^{-1}1)^2} \\
\lambda_2 &= \frac{(R'\Sigma^{-1}R - (R'\Sigma^{-1}1)A}{(R'\Sigma^{-1}R)(1'\Sigma^{-1}1) - (R'\Sigma^{-1}1)^2}
\end{aligned}
$$

Today: Optimization

1. Aside to Multivariable Calculus: partial derivatives
2. Aside to Linear Algebra: Quadratic forms
3. Finding extreme values of a function f
   - Necessary and sufficient conditions
   - Convexity and concavity
4. Constrained optimization
   - Optimization over an interval
   - Lagrangian method
5. Principal Component Analysis