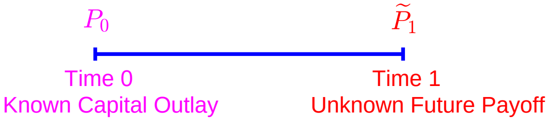
Quantitative Finance
Part II - Portfolio Choice

Dr. Luo, Dan

May 30, 2020

The Single Period Setup



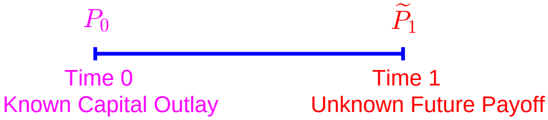
Two most important things in asset pricing:

1. Time value
2. Uncertainty

When is a single period model a good approximation?

1. A zero coupon bond held to maturity
2. A physical project providing no dividends until completed

Returns



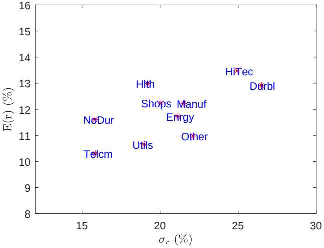
If we have a long position, then

- Initial outlay is X_0. Final receipt is X_1.
- The total return is $R = X_1/X_0$.

If we short sell the asset, then

- Initial outlay is $-X_0$. Final receipt is $-X_1$.
- The total return is $R = (-X_1)/(-X_0) = X_1/X_0$.

Mean-Standard Deviation Diagram



1 NoDur Consumer Nondurables -- Food, Tobacco, Textiles, Apparel, Leather, Toys
2 Durbl Consumer Durables -- Cars, TVs, Furniture, Household Appliances
3 Manuf Manufacturing -- Machinery, Trucks, Planes, Chemicals, Off Furn, Paper, Com Printing
4 Enrgy Oil, Gas, and Coal Extraction and Products
5 HiTec Business Equipment -- Computers, Software, and Electronic Equipment
6 Telcm Telephone and Television Transmission
7 Shops Wholesale, Retail, and Some Services (Laundries, Repair Shops)
8 Hlth Healthcare, Medical Equipment, and Drugs
9 Utils Utilities
10 Other Other -- Mines, Constr, BldMt, Trans, Hotels, Bus Serv, Entertainment, Finance

A simple method to represent randomness in returns is through a two-dimensional mean-standard deviation diagram. (Why not mean-variance?)

The above figure shows the 10 industry portfolios of the US equity market. Here we use monthly returns from July 1926 to January 2020 to do the calculation. We report means and standard deviations on an annualized basis.

Recap

We invest a total amount of X_0 in N assets, with amount X_{0i} for asset i, $i = 1, 2, ..., N$. Let $w_i = X_{0i}/X_0$ be the weight of asset i. Obviously, $\sum_{i=1}^N w_i = 1$.

The, the portfolio return is

$$R_p = \frac{\sum_{i=1}^N R_i w_i X_0}{X_0} = \sum_{i=1}^N w_i R_i.$$

Equivalently,

$$r_p = \sum_{i=1}^N w_i r_i, \tag{1}$$

because $\sum_{i=1}^N w_i = 1$.

Portfolio Mean and Variance

Let $\bar{r}_i = E(r_i)$.

Write $\bar{\boldsymbol{r}} = (\bar{r}_1, \bar{r}_2, ..., \bar{r}_N)^\top$ and $\boldsymbol{w} = (w_1, w_2, ..., w_N)^\top$.

Let $\boldsymbol{\Sigma} = [\sigma_{ij}]$ be the $N \times N$ variance-covariance matrix, where $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$.

Then, the portfolio mean and variance are

$$\bar{r}_p = \sum_{i=1}^{N} w_i \bar{r}_i = \boldsymbol{w}^\top \bar{\boldsymbol{r}}, \tag{2}$$

$$\sigma_p^2 = E\left[(r_p - \bar{r}_p)^2\right] = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \sigma_{ij} = \boldsymbol{w}^\top \boldsymbol{\Sigma} \boldsymbol{w}. \tag{3}$$

Notice that all vectors and matrices are shown in **bold**.

Exercise 1

Calculate the mean and variance of an equally weight portfolio on the 10 industry portfolios. Use monthly returns provided in *10_Industry_Portfolios.xlsx*. (First estimate the means and covariances for the 10 industry portfolios.)

Excel provides built-in functions for matrix manipulations.

- *MMULT(A, B)* for matrix multiplications;
- *TRANSPOSE(A)* for matrix transpose;
- *MUNIT(N)* to generate a $N \times N$ identity matrix;
- $A * B$ gives the element by element multiplication;
- $A * MUNIT(N)$ returns the diagonal matrix of A.

Press Ctrl+Shift+Enter, instead of Enter, to get results involving matrix functions.

Diversification

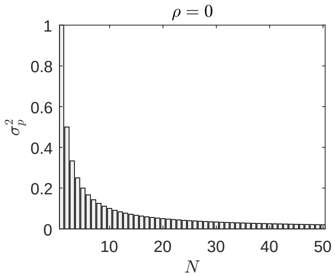
Equation (3) allows us to examine the effect of diversification.
Consider the simple case of an equally weighted portfolio consisting of N assets with the same mean $\bar{r}$ and same variance σ^2. Hence, $r_p = \frac{1}{N}\sum_{i=1}^{N} r_i$ and $\bar{r}_p = \bar{r}$.
If the assets are pairwise uncorrelated, then

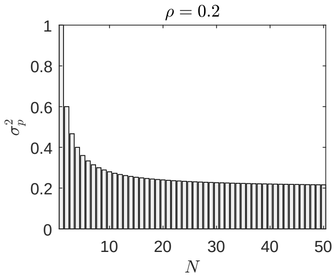
$$\sigma_p^2 = \frac{1}{N^2}\sum_{i=1}^{N}\sigma^2 = \frac{1}{N}\sigma^2. \tag{4}$$

If the assets are pairwise correlated with the same correlation coefficient ρ, then

$$\begin{aligned}\sigma_p^2 &= E\left[(r_p - \bar{r})^2\right] = \frac{1}{N^2}E\left[\sum_{i=1}^{N}(r_i - \bar{r})\right]^2 \\ &= \frac{1}{N^2}\left(\sum_{i=1}^{N}\sigma^2 + \sum_{j\neq i, j=1}^{N}\sum_{i=1}^{N}\rho\sigma^2\right) = \frac{1}{N}\sigma^2 + \left(1 - \frac{1}{N}\right)\rho\sigma^2 = \rho\sigma^2 + \frac{1}{N}(1-\rho)\sigma^2. \end{aligned} \tag{5}$$

Diversification





$$\sigma_p^2 \propto N^{-1}.$$

$$\sigma_p^2 \to \rho\sigma^2, \text{ as } N \to \infty.$$

The Markowitz Model

There are N assets with expected returns $\bar{r}_1, \bar{r}_2, ..., \bar{r}_N$, and covariances $\sigma_{ij} = \rho_{ij}\sigma_i\sigma_j$, $i, j = 1, 2, ..., N$.

A portfolio formed with the above assets is defined by the weights $w_1, w_2, ..., w_N$.

The quest is to find the minimum-variance portfolio for any (feasible) desired level of expected return.

A Mathematical Formulation

Fix the expected return of a portfolio at $\bar{r}_p$. We need to find the minimum-variance portfolio that achieves $\bar{r}_p$.

$$\begin{aligned} \min_{w_1, w_2, \ldots, w_N} \quad & \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} w_i \sigma_{ij} w_j, \\ s.t. \quad & \sum_{i=1}^{N} w_i \bar{r}_i = \bar{r}_p, \\ & \sum_{i=1}^{N} w_i = 1. \end{aligned} \tag{6}$$

The 1/2 is innocuous and just works to make the solution neater.

Solution

The Markowitz Model provides the foundation for single-period investment decisions by explicitly addressing the tradeoff between expected return and variance of a portfolio.

We solve it using the Lagrangian method. We form the Lagrangian

$$\mathfrak{L} = \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} w_i \sigma_{ij} w_j - \lambda \left(\sum_{i=1}^{N} w_i \bar{r}_i - \bar{r}_p \right) - \zeta \left(\sum_{i=1}^{N} w_i - 1 \right), \quad (7)$$

where λ and ζ are the Lagrangian Multipliers.

Solution - continued

Differentiating the Lagrangian w.r.t. the weights and the multipliers, we get the following first order conditions (F.O.C.s).

$$\sum_{j=1}^{N} \sigma_{ij} w_j - \lambda \bar{r}_i - \zeta = 0, \ \textit{for } i = 1, 2, ..., N, \tag{8}$$

$$\sum_{i=1}^{N} w_i \bar{r}_i = \bar{r}_p, \tag{9}$$

$$\sum_{i=1}^{N} w_i = 1. \tag{10}$$

We use the fact $\sigma_{ij} = \sigma_{ji}$ in (8). We have $N + 2$ linear equations for $N + 2$ unknowns. We can in principle solve the model with linear algebra methods.

A Simple Case with Two Assets

Before we present the general solution, we first consider the case of two individual assets with expected returns $\bar{r}_1$ and $\bar{r}_2$ ($\bar{r}_1 \neq \bar{r}_2$), and covariances σ_1^2, σ_{12} and σ_2^2. The F.O.C.s are:

$$\begin{aligned} w_1\sigma_1^2 + w_2\sigma_{12} - \lambda\bar{r}_1 - \zeta &= 0, \\ w_1\sigma_{12} + w_2\sigma_2^2 - \lambda\bar{r}_2 - \zeta &= 0, \\ w_1\bar{r}_1 + w_2\bar{r}_2 &= \bar{r}_p, \\ w_1 + w_2 &= 1. \end{aligned}$$

The last two equations give

$$w_1^* = \frac{\bar{r}_p - \bar{r}_2}{\bar{r}_1 - \bar{r}_2}, \; w_2^* = \frac{\bar{r}_1 - \bar{r}_p}{\bar{r}_1 - \bar{r}_2}.$$

An Illustration

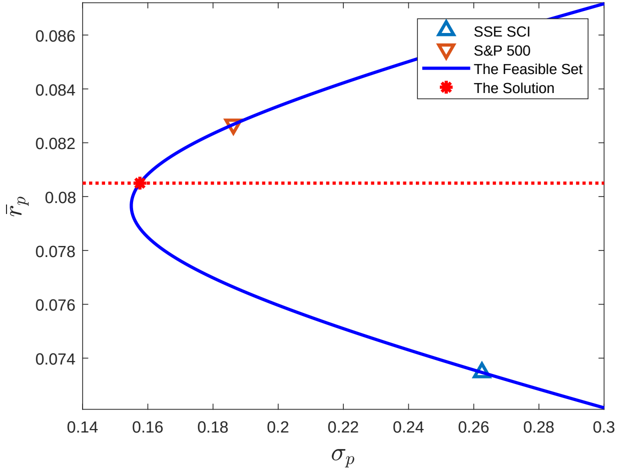
For the simple case of two assets, only one combination of the assets achieves the required portfolio expected return $\bar{r}_p$. The minimization problem degenerates.

We use the SSE Stock Composite Index and S&P 500 Index as two aggregate stocks (or ETFs), for illustration.

	$\bar{r}_i$	σ_i^2	σ_{12}	ρ
SSE SCI	0.07347	0.06891	0.002088	4.27%
S&P 500	0.08268	0.03469	0.002088	4.27%

All quantities are annualized when possible.

The Feasible Set



A Case of Three Uncorrelated Assets

Now we consider a slightly more complicated case of three uncorrelated assets, with expected returns $\bar{r}_1$, $\bar{r}_2$ and $\bar{r}_3$, and the same variance σ^2. ($\sigma_{ij} = 0,\ i \neq j$.) The F.O.C.s are:

$$\begin{aligned} w_1\sigma^2 - \lambda\bar{r}_1 - \zeta &= 0, \\ w_2\sigma^2 - \lambda\bar{r}_2 - \zeta &= 0, \\ w_3\sigma^2 - \lambda\bar{r}_3 - \zeta &= 0, \\ w_1\bar{r}_1 + w_2\bar{r}_2 + w_3\bar{r}_3 &= \bar{r}_p, \\ w_1 + w_2 + w_3 &= 1. \end{aligned}$$

We could no longer solve for the w_is from the last two equations.

A Further Simplification

Let $\bar{r}_1 = 1$, $\bar{r}_2 = 2$ and $\bar{r}_3 = 3$, and the same variance $\sigma^2 = 1$. The F.O.C.s are:

$$w_1 - \lambda - \zeta = 0, \tag{11}$$

$$w_2 - 2\lambda - \zeta = 0, \tag{12}$$

$$w_3 - 3\lambda - \zeta = 0, \tag{13}$$

$$w_1 + 2w_2 + 3w_3 = \bar{r}_p, \tag{14}$$

$$w_1 + w_2 + w_3 = 1. \tag{15}$$

Solution

Now, (11)+(13)-2×(12), we obtain:

$$w_1 - 2w_2 + w_3 = 0, \tag{16}$$

$$w_1 + 2w_2 + 3w_3 = \bar{r}_p, \tag{17}$$

$$w_1 + w_2 + w_3 = 1. \tag{18}$$

Then, (18)-(16) gives

$$w_2^* = 1/3.$$

And

$$w_1^* = -\bar{r}_p/2 + 4/3,$$
$$w_3^* = \quad \bar{r}_p/2 - 2/3.$$

Discussion on the Effect of $\bar{r}_p$ and $\bar{r}_i$

We can write the solution as:

$$w_2^* = 1/3, \tag{19}$$

$$w_1^* = -\bar{r}_p/2 + 4/3 = 1/3 - (\bar{r}_p/2 - 1), \tag{20}$$

$$w_3^* = \bar{r}_p/2 - 2/3 = 1/3 + (\bar{r}_p/2 - 1). \tag{21}$$

Note the interesting special case of $\bar{r}_p = 2$ $(= (\bar{r}_1 + \bar{r}_2 + \bar{r}_3)/3)$, which results in $w_1^* = w_2^* = w_3^* = 1/3$.

- ◇ A higher $\bar{r}_p > 2$ increases investment in Asset 3 while decreases investment in Asset 1.
- ◇ A lower $\bar{r}_p < 2$ decreases investment in Asset 3 while increases investment in Asset 1.
- ◇ The level of $\bar{r}_p$ does not affect investment in Asset 2 because of the symmetry in $\bar{r}_i$. No shorting is needed for $4/3 < \bar{r}_p < 8/3$.

An Illustration

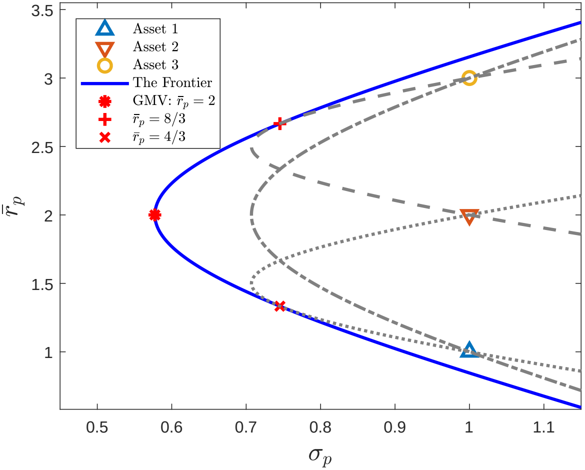
The minimum variance at $\bar{r}_p$ is

$$\sigma_p^2 = w_1^2 + w_2^2 + w_3^2 = \frac{\bar{r}_p^2}{2} - 2\bar{r}_p + \frac{7}{3}.$$

The Global Minimum Variance portfolio is computed at:

$$\bar{r}_p^G = 2,$$
$$\sigma_p^G = \frac{\sqrt{3}}{3},$$
$$w_1^G = w_2^G = w_3^G = \frac{1}{3}.$$

The Efficient Frontier



The Maximum Variance

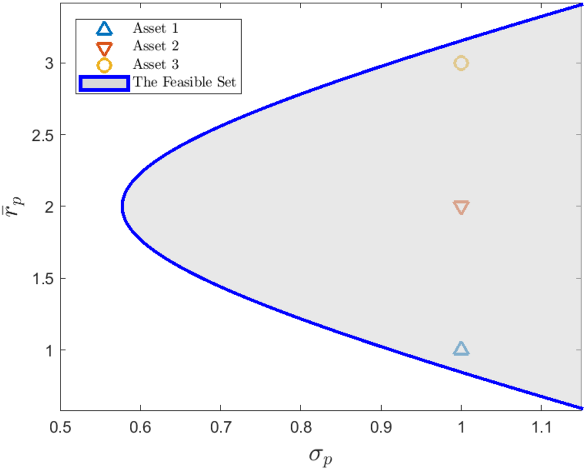
We ask the opposite question: what is the maximum variance of a portfolio with expected return $\bar{r}_p$?

1. For the case of two assets, any $\bar{r}_p$ is paired with a single σ_p.
2. For the case of three assets, we can show that, from (17) and (18),

$$
\begin{aligned}
w_2 &= -2w_1 - \bar{r}_p + 3 \\
w_3 &= w_1 + \bar{r}_p - 2, \\
\sigma_p^2 &= w_1^2 + w_2^2 + w_3^2 = 6w_1^2 + 2(3\bar{r}_p - 8)w_1 + 2\bar{r}_p^2 - 10\bar{r}_p + 13.
\end{aligned}
$$

Obviously, σ_p^2 is unbounded if short selling is allowed. That is, σ_p^2 could be any value in $\left[\frac{1}{2}\bar{r}_p^2 - 2\bar{r}_p + \frac{7}{3}, +\infty\right)$, at any fixed $\bar{r}_p$.

The Feasible Set



No Short Selling Constraint

$$\min_{w_1, w_2, \ldots, w_N} \quad \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} w_i \sigma_{ij} w_j, \tag{22}$$

$$s.t. \quad \sum_{i=1}^{N} w_i \bar{r}_i = \bar{r}_p,$$

$$\sum_{i=1}^{N} w_i = 1,$$

$$w_i \geq 0, \;\; for \;\; i = 1, 2, \ldots, N.$$

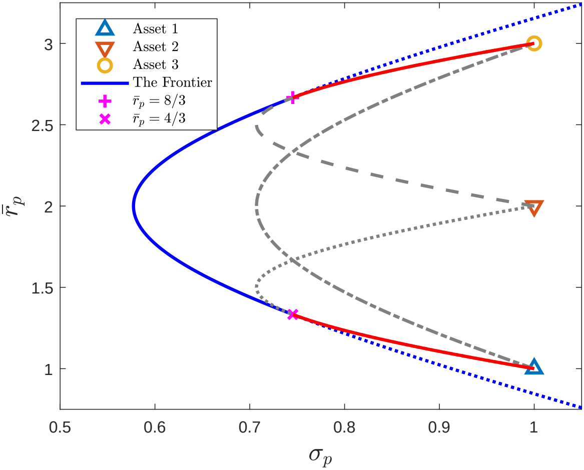
This is a **Quadratic Program** with a quadratic objective function and linear equality and inequality constraints.
We can solve it numerically, e.g., using Excel solver for a relatively small number of assets and other professional programs for hundreds or thousands of assets.

The Efficient Frontier - No Shorting

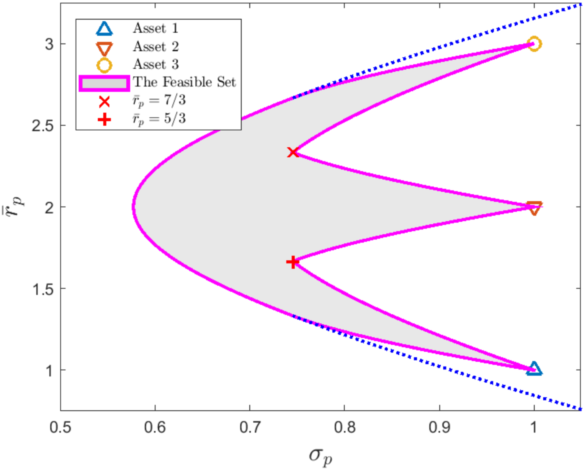
For the simple case of three assets above, we can find the efficient frontier explicitly through the solution in (19)-(21).

	$1 \leq \bar{r} < 4/3$	$4/3 \leq \bar{r} \leq 8/3$	$8/3 < \bar{r} \leq 3$
w_1	$2 - \bar{r}_p$	$1/3 - (\bar{r}_p/2 - 1)$	0
w_2	$\bar{r}_p - 1$	$1/3$	$3 - \bar{r}_p$
w_3	0	$1/3 - (\bar{r}_p/2 - 1)$	$\bar{r}_p - 2$
σ_p	$\sqrt{2\bar{r}_p^2 - 6\bar{r}_p + 5}$	$\sqrt{\bar{r}_p^2/2 - 2\bar{r}_p + 7/3}$	$\sqrt{2\bar{r}_p^2 - 10\bar{r}_p + 13}$
No Shorting	Binding	Not Binding	Binding

The Efficient Frontier - No Shorting

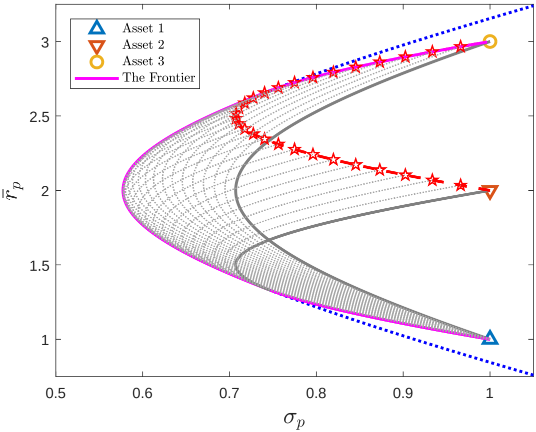


The Feasible Set - No Shorting



How to Deal with Shorting Constraints?

The trick is to watch the area swept by the curve!



Unconstrained Optimization

A general mathematical formulation of multivariate unconstrained optimization is as follows.

$$\min_{x \in \mathbb{R}^n} \quad f(x), \tag{23}$$

where $f(x)$ is the objective function.

Here, we assume all functions are well-behaved[1], that is, sufficiently smooth, e.g., twice-continuously differentiable.

Maximization problems can be transformed into equivalent minimizations by simply putting a negative sign before $f(x)$.

$$\max_{x \in \mathbb{R}^n} \quad f(x) \quad \Leftrightarrow \quad \min_{x \in \mathbb{R}^n} \quad -f(x). \tag{24}$$

[1] Or the functions can be well approximated by well-behaved ones.

Univariate Minimization

A univariate unconstrained optimization problem considers an objective function in one dimension.

$$\min_{x \in \mathbb{R}} \quad f(x) \tag{25}$$

Importance of univariate optimization:

1. Many real problems corresponds to finding the optimum of univariate functions (e.g., optimal hedge ratios of derivatives).
2. Multivariable optimization methods in commercial use today mostly contain a line search step.
3. Fundamental ideas are best illustrated with univariate cases, and they usually carry over to multivariate optimization.

Solution

Definition of a solution x^*:

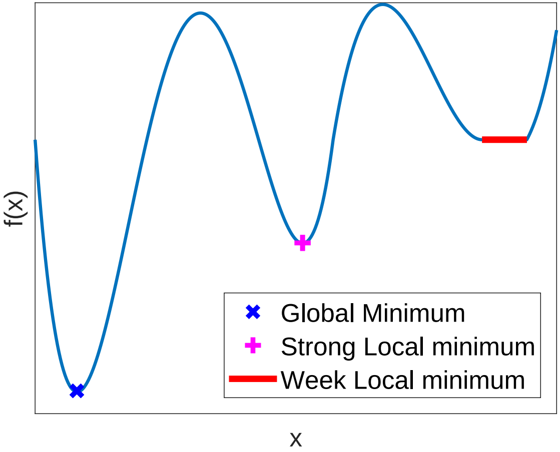
(i) Global Minimum: There exists a point x^* *s.t.* $f(x^*) \leq f(x), \forall x \in \mathbb{R}$.

(ii) Strong Local minimum: There exists a point x^* *s.t.* $f(x^*) < f(x)$, $\forall x \in U(x^*)$, where $U(x^*)$ is a neighbourhood of x^*.

(iii) Weak Local Minima: There exists a neighbourhood $U(x^*)$ of x^* *s.t.* $f(x^*) \leq f(x), \forall x \in U(x^*)$.

Clearly, (i) $\not\Rightarrow$ (ii), (ii) $\not\Rightarrow$ (i), (ii) $\Rightarrow$ (iii), (iii) $\not\Rightarrow$ (ii), (i) $\Rightarrow$ (iii), and (iii) $\not\Rightarrow$ (i).

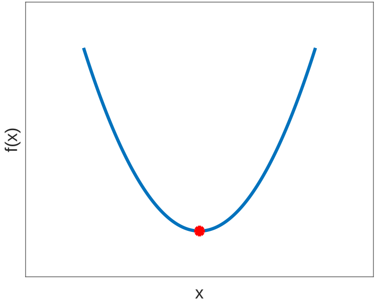
An Illustration



Convexity

If the objective function is convex, that is, for any $y, z \in \mathbb{R}$,

$$f(wy + (1-w)z) \leq wf(y) + (1-w)f(z),\ 0 \leq w \leq 1, \tag{26}$$



then any local minimum is also a global minimum.

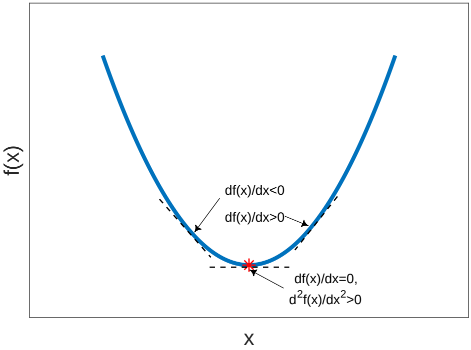
Discussion on Convexity

Convexity is a general concept used in optimization theory. It describes the property of having an optimum for a function.

Convexity combines both stationarity (stable points) and curvature into a single concept.

However, it is inconvenient to use for a specific function. For the well-behaved functions considered here, first derivatives give us a measure of the rate of change of the function. Second derivatives give us a measure of curvature of the function or the rate of change of the first derivatives.

Discussion on Convexity



For a (locally) convex function, the first derivative starts out negative and becomes positive, with the turning point x^*. Put differently, x^* is a (local) minimum of the function $f(x)$.

Necessary and Sufficient Conditions for an Optimum

For a well-behaved twice continuously differentiable function $f(x)$, the point x^* is an optimum iff:

$$\left.\frac{df(x)}{dx}\right|_{x^*} = 0 \; (\textit{stationarity}),$$

and

$$\left.\frac{d^2 f(x)}{dx^2}\right|_{x^*} > 0 \; (\textit{minimum}),$$

or

$$\left.\frac{d^2 f(x)}{dx^2}\right|_{x^*} < 0 \; (\textit{maximum}).$$

Necessary and Sufficient Conditions for an Optimum

What if $\frac{df(x)}{dx}\Big|_{x^*} = \frac{d^2f(x)}{dx^2}\Big|_{x^*} = 0$?

Two examples:

1. For the function $f(x) = x^4$, we know that it have a minimum at $x^* = 0$. And $\frac{df(x)}{dx}\Big|_{x^*} = \frac{d^2f(x)}{dx^2}\Big|_{x^*} = \frac{d^3f(x)}{dx^3}\Big|_{x^*} = 0$, while $\frac{d^4f(x)}{dx^4}\Big|_{x^*} > 0$.
2. For the function $f(x) = x^3$, we know that $x^* = 0$ is neither a maximum nor a minimum. And $\frac{df(x)}{dx}\Big|_{x^*} = \frac{d^2f(x)}{dx^2}\Big|_{x^*} = 0$, while $\frac{d^3f(x)}{dx^3}\Big|_{x^*} > 0$.

Necessary and Sufficient Conditions for an Optimum

Generally, we need to check higher-order derivatives in case of a stationary point x^* with a zero second derivative.

1. If the first non-zero derivative is of an odd order, that is:

$$\left.\frac{d^n f(x)}{dx^n}\right|_{x^*} \neq 0, \; n \geq 3 \textit{ and } n \textit{ is odd},$$

then x^* is a saddle/inflection point, not an extremum point.

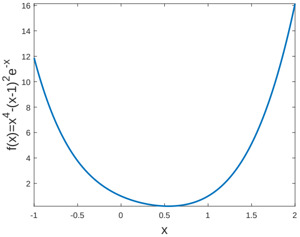
2. If the first non-zero derivative is of an even order, that is:

$$\left.\frac{d^n f(x)}{dx^n}\right|_{x^*} \neq 0, \; n \geq 4 \textit{ and } n \textit{ is even},$$

then x^* is a minimum point if $\left.\frac{d^n f(x)}{dx^n}\right|_{x^*} > 0$, and a maximum point if $\left.\frac{d^n f(x)}{dx^n}\right|_{x^*} < 0$.

An Example

$$\min_{x\in\mathbb{R}} \quad f(x) = x^4 - (x-1)^2 e^{-x} \tag{27}$$



F.O.C.: $4(x^*)^3 - 2(x^*-1)e^{-x^*} + (x^*-1)^2 e^{-x^*} = 0. \Rightarrow x^* = ?$

S.O.C.: $12(x^*)^2 + 2(2x^*-3)e^{-x^*} - (x^*-1)^2 e^{-x^*}$

$= 4(x^*)^3 + 12(x^*)^2 + 2(x^*-2)e^{-x^*} \gtrless 0?$

Discussions

Three observations:

1. The F.O.C. may be a nonlinear equation which is often as difficult to solve as the original optimization problem.
2. The sign of higher-order derivatives may be difficult to determine.
3. In practical problems, the objective functions and the derivatives may only be computed numerically.

In sum, unconstrained univariate optimization problems may not be as straightforward as you thought.

Multivariate Optimization

Now we extend to unconstrained minimization of a function in n dimensions.

$$\min_{x \in \mathbb{R}^n} \quad f(x), \tag{28}$$

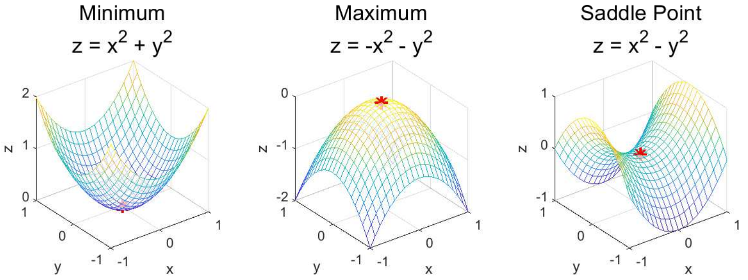
where $f(x)$ is the twice-continuously differentiable objective function.

We can likewise define global, weak local, and strong local minima, as well as convexity, in n dimensions.

Stationarity

A stationary point x^* of the function $f(x)$ in n dimensions is one satisfying $\nabla f(x^*) = 0$, where $\nabla = \frac{\partial}{\partial x}$ is the Nabla operator with $\nabla f(x) = \left(\frac{\partial f(x)}{\partial x_1}, \frac{\partial f(x)}{\partial x_2}, \ldots, \frac{\partial f(x)}{\partial x_n} \right)$.

There are three types of stationary points.



Necessary Conditions for an Optimum

The conditions are simply multivariate extensions of univariate ones.

Necessary conditions for a weak local minimum

C1. $\nabla f(x^*) = 0$ (stationarity).
C2. $\nabla^2 f(x^*)$ is positive semi-definite. That is, $v^\top \nabla^2 f(x^*) v \geq 0$, for all $n \times 1$ vector $v \neq 0$.

$$\text{Here, } \nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f(x)}{\partial x_1^2} & \frac{\partial^2 f(x)}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f(x)}{\partial x_2 \partial x_1} & \frac{\partial^2 f(x)}{\partial x_2^2} & \cdots & \frac{\partial^2 f(x)}{\partial x_2 \partial x_n} \\ \cdots & \cdots & \cdots & \cdots \\ \frac{\partial^2 f(x)}{\partial x_n \partial x_1} & \frac{\partial^2 f(x)}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f(x)}{\partial x_n^2} \end{bmatrix}.$$

It is called the Hessian Matrix after German mathematician Ludwig Otto Hesse. It describes the local curvature of $f(x)$.

Sufficient Conditions for an Optimum

The conditions are again simply multivariate extensions of univariate ones.

Sufficient conditions for a strong local minimum

C1. $\nabla f(x^*) = 0$ (stationarity).
C2. $\nabla^2 f(x^*)$ is positive definite. That is, $v^\top \nabla^2 f(x^*) v > 0$, for all $n \times 1$ vector $v \neq 0$.

The Recipe

1. Correctly formulate the optimization problem.
2. Find a point x^* that could potentially be a solution (satisfying the necessary conditions).
3. Verify this point x^* is certainly a solution (satisfying the sufficient conditions).

Example

Consider the point $(x^*, y^*) = (0, 0)$.

1. For $f(x, y) = x^2 + y^2$, $\nabla f(x^*, y^*) = 0$, and $\nabla^2 f(x^*, y^*) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$, which is positive definite.
2. For $f(x, y) = -x^2 - y^2$, $\nabla f(x^*, y^*) = 0$, and $\nabla^2 f(x^*, y^*) = \begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}$, which is negative definite. (Consider $-f(x, y)$. Then $\nabla^2(-f(x^*, y^*))$ is positive definite.)
3. For $f(x, y) = x^2 - y^2$, $\nabla f(x^*, y^*) = 0$, and $\nabla^2 f(x^*, y^*) = \begin{bmatrix} 2 & 2x^* - 2y^* \\ 2x^* - 2y^* & -2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & -2 \end{bmatrix}$, which is indefinite.

A Different Perspective

Before we turn to constrained optimization, we provide another intuitive discussion of the methods we covered for unconstrained optimization. Again, it would help us understand the unconstrained approach, and more importantly, it can be easily extended to the constrained situation.

We would understand what the second order conditions for optimality are and also why they are. We will use an algebraic approach, exploiting the matrix of second derivatives called the Hessian matrix. Then we will see later that the conditions for the constrained case can be easily stated in terms of a bordered Hessian matrix[2].

[2]Many applied mathematics students use it for a long time without knowing its relevance.

The Univariate Case

The 2nd order approximation to $f(x)$ near $\hat{x}$ is

$$f(x) = f(\hat{x}) + f_x(\hat{x})(x - \hat{x}) + \frac{1}{2} f_{xx}(\hat{x})(x - \hat{x})^2. \tag{29}$$

Hence,

$$df = f_x(\hat{x})dx + \frac{1}{2} f_{xx}(\hat{x})(dx)^2, \tag{30}$$

where $df = f(x) - f(\hat{x})$ and $dx = x - \hat{x}$.

The Univariate Case

For a critical point x^* satisfying $f_x(x^*) = 0$,

$$df = \frac{1}{2} f_{xx}(x^*)(dx)^2. \tag{31}$$

Then, clearly, because $(dx)^2 > 0$, $df > 0$ if $f_{xx}(x^*) > 0$. That is to say, x^* is a minimum point since f increases regardless of the direction of change in x around x^*.

Similarly, $df < 0$ if $f_{xx}(x^*) < 0$. That is to say, x^* is a maximum point since f decreases regardless of the direction of change in x around x^*.

Note that the Hessian matrix for the univariate case is just the 1×1 matrix $[f_{xx}(x^*)]$.

The Two-Variable Case

The 2nd order approximation to $f(x, y)$ near $(\hat{x}, \hat{y})$ is

$$df = f_x dx + f_y dy + \frac{1}{2}[f_{xx}(dx)^2 + f_{xy} dxdy + f_{yx} dydx + f_{yy}(dy)^2] \quad (32)$$

$$= f_x dx + f_y dy + \frac{1}{2}(dx, dy)H(dx, dy)^\top, \quad (33)$$

where $H = \begin{bmatrix} f_{xx} & f_{xy} \\ f_{yx} & f_{yy} \end{bmatrix}$ is the Hessian matrix.

The Two-Variable Case

For a critical point (x^*, y^*) satisfying $f_x(x^*, y^*) = f_y(x^*, y^*) = 0$,

$$df = \frac{1}{2}(dx, dy)H(dx, dy)^\top. \tag{34}$$

Then (x^*, y^*) is a minimum point if H is positive definite such that $(dx, dy)H(dx, dy)^\top > 0$ for all $(dx, dy) \neq (0, 0)$.

Similarly, (x^*, y^*) is a maximum point if H is negative definite such that $(dx, dy)H(dx, dy)^\top < 0$ for all $(dx, dy) \neq (0, 0)$.

Exercise 2

Do the analysis for the three-variable case.

Exercise 2

Do the analysis for the three-variable case.

The 2nd order approximation to $f(x, y, z)$ near $(\hat{x}, \hat{y}, \hat{z})$ is

$$df = f_x dx + f_y dy + f_z dz + \frac{1}{2}(dx, dy, dz)H(dx, dy, dz)^\top, \tag{35}$$

where $H = \begin{bmatrix} f_{xx} & f_{xy} & f_{xz} \\ f_{yx} & f_{yy} & f_{yz} \\ f_{zx} & f_{zy} & f_{zz} \end{bmatrix}$ is the Hessian matrix.

For a critical point (x^*, y^*, z^*) satisfying $f_x = f_y = f_z = 0$,

$$df = \frac{1}{2}(dx, dy, dz)H(dx, dy, dz)^\top. \tag{36}$$

Hence (x^*, y^*, z^*) is a minimum point if H is positive definite, and a maximum point if H is negative definite.

The n-Variable Case

The 2nd order approximation to $f(x_1, x_2, ..., x_n)$ is

$$df = \sum_{i=1}^{n} f_i dx_i + \frac{1}{2}(dx_1, dx_2, ..., dx_n)H(dx_1, dx_2, ..., dx_n)^\top, \tag{37}$$

where $H = \begin{bmatrix} f_{11} & f_{12} & \cdots & f_{1n} \\ f_{21} & f_{22} & \cdots & f_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ f_{n1} & f_{n2} & \cdots & f_{nn} \end{bmatrix}$ is the Hessian matrix.

For a critical point $(x_1^*, x_2^*, ..., x_n^*)$ satisfying $f_1 = f_2 = ... = f_n = 0$,

$$df = \frac{1}{2}(dx_1, dx_2, ..., dx_n)H(dx_1, dx_2, ..., dx_n)^\top. \tag{38}$$

Hence $(x_1^*, x_2^*, ..., x_n^*)$ is a minimum point if H is positive definite, and a maximum point if H is negative definite.

The Lagrangian Method

The **method of Lagrange multipliers** is a strategy for finding the local *maxima* and *minima* of a function subject to *equality* constraints.

This Lagrangian (or Lagrange) method converts a constrained optimization problem into an unconstrained one to which the derivative test applies.

1. Identify the stationary points from the first-order necessary conditions.
2. Determine whether the stationary points are maxima, minima, or saddle points, through the definiteness of the bordered Hessian matrices.

The Lagrange Theorem

You may find more general and rigorous proof of the Lagrange Theorem elsewhere. Here we focus on the following version.

The Lagrange Theorem

Consider the optimization problem of maximizing the function $f(x)$ in n dimensions subject to m equality constraints ($m < n$).

$$\max_{x \in \mathbb{R}^n} \; f(x), \tag{39}$$

$$\text{s.t.} \; g_j(x) = 0, \; j = 1, 2, ..., m, \tag{40}$$

where $f(x)$ and $g_j(x)$ are twice continuously differentiable.
Then at a maximum point x^*, there are scalars λ_1, λ_2, ..., λ_m, also called Lagrange Multipliers, such that

$$\nabla f(x^*) = \sum_{j=1}^{m} \lambda_j \nabla g_j(x^*) \Longleftrightarrow \nabla f(x^*) - \sum_{j=1}^{m} \lambda_j \nabla g_j(x^*) = 0. \tag{41}$$

The gradient of the function is a linear combination of the gradients of the constraints.

Intuition

Consider a slightly simpler case of maximizing $f(x)$ in $n = 3$ dimensions subject to an equality constraint $g(x) = 0$. The Lagrange Theorem says that at the maximum x^*, $\nabla f(x^*) = \lambda \nabla g(x^*)$.
So,

1. Why should the gradient of the objective function be proportional to that of the constraint? Why are the two gradients related at all?
2. Where does the λ come from?

Intuition: For the Case of a Single Constraint

1. The set of points satisfying $g(x) = 0$ is a surface in 3 dimensions, or a (maybe oddly shaped) balloon.
2. The set of points satisfying $f(x) = k$ is another surface in 3 dimensions, or another (maybe oddly shaped) balloon.
3. Take k to be a very large number, larger than the maximum of $f(x)$ under the constraint. Then the Balloon $g(x) = 0$ lies inside the balloon $f(x) = k$.

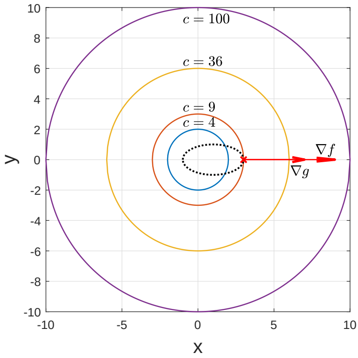
Intuition: For the Case of a Single Constraint

4. Now gradually shrink k, by leaking the air in the outer balloon. At some point, the outer balloon will touch the inner one at the maximum under constraint.
5. At the touching point of the maximum, the two balloons should be tangent, hence their normal vectors, given by their gradients, should be both perpendicular to the same tangent plane thus parallel to each other.
6. Recall the Geometrical interpretation of a gradient: The direction of the gradient is the direction of fastest increase of the function at a certain point, and its magnitude is the rate of increase in that direction.
7. Hence, in contrast to the direction of increase, $f(x)$ and $g(x)$ need not have the same rate of increase since the two balloons may have different shapes. The scaling constant λ results.

This explanation works for any n if you think of *hyper*-balloons.

A Pictorial Illustration

Let $n = 2$. Consider $z = f(x, y) = x^2 + y^2$ and $g(x, y) = (x - 1)^2 + 4y^2 - 4 = 0$. The level curves of f, defined by $x^2 + y^2 = c$, are circles. The constraint is an ellipse.



Intuition: For the Case of Multiple Constraints

Once we understood the case with one constraint, extension to multiple constraints are straightforward.

Consider the case of maximizing $f(x)$ in n dimensions subject to m equality constraints $g_j(x) = 0$.

1. The Lagrange Theorem says that at the maximum x^*, $\nabla f(x^*) = \sum_{j=1}^{m} \lambda_j \nabla g_j(x^*)$.
2. This tells us that any direction of change that is perpendicular to all the gradients ∇g_j must be perpendicular to the gradient ∇f as well.
3. This in turn means that you cannot increase f without violating at least one of the constraints.

The Lagrange Method

We can summary everything by defining a single Lagrangian

$$\mathfrak{L}(x_1, x_2, ..., x_n; \lambda_1, \lambda_2, ..., \lambda_n) = f(x_1, x_2, ..., x_n) - \sum_{j=1}^{m} \lambda_j g_j(x_1, x_2, ..., x_n).$$

F.O.C.s:

$$\frac{\partial \mathfrak{L}}{\partial x_i} = \frac{\partial f}{\partial x_i} - \sum_{j=1}^{m} \lambda \frac{\partial g_j}{\partial x_i} = 0, \ i = 1, 2, ..., n,$$

$$\frac{\partial \mathfrak{L}}{\partial \lambda_j} = -g_j(x_1, x_2, ..., x_n) = 0, \ j = 1, 2, ..., m.$$

In one shot, we obtain $n + m$ equations, the first n from the Lagrange Theorem and the next m for constraints, with $n + m$ unknowns.

Second Order Conditions

Now we have identified the stationary points, candidates for extrema. How could we determine they are actually maxima, minima, or saddle points?

Naturally, we may wonder wether the argument with Hessian matrix of a unconstrained optimization problem could be transplanted here. After all, we have made great efforts to transform a constrained optimization into an unconstrained one. The answer is "Yes" and "No".

- ◇ "Yes": We can basically apply a Hessian-based approach to perform such a job.
- ◇ "No": Now the optimum points must satisfy the constraints which dictate that for each j, $\nabla g_j \cdot (dx_1, dx_2, ..., dx_n) = 0$. Hence $(dx_1, dx_2, ..., dx_n)$ cannot be arbitrary but confined by all the constraints. We only need to check optimality of f along the directions perpendicular to all the gradients of the constraints.

A Motivational Example

Consider the following maximization problem.

$$\max_{x,y} \; -x^2 - y^2, \qquad (42)$$
$$\text{s.t.} \;\; x + y - 2 = 0.$$

The Lagrangian is

$$\mathfrak{L}(x, y, \lambda) = -x^2 - y^2 - \lambda(x + y - 2).$$

F.O.C.s:

$$-2x^* - \lambda^* = 0,$$
$$-2y^* - \lambda^* = 0,$$
$$-x^* - y^* + 2 = 0.$$

$$\Longrightarrow \qquad x^* = 1, \; y^* = 1, \; \lambda^* = -2.$$

The Hessian

We can calculate the Hessian for the Lagrangian using the F.O.C.s.

$$H = \begin{bmatrix} \mathfrak{L}_{xx} & \mathfrak{L}_{xy} & \mathfrak{L}_{x\lambda} \\ \mathfrak{L}_{yx} & \mathfrak{L}_{yy} & \mathfrak{L}_{y\lambda} \\ \mathfrak{L}_{\lambda x} & \mathfrak{L}_{\lambda y} & \mathfrak{L}_{\lambda\lambda} \end{bmatrix} = \begin{bmatrix} -2 & 0 & -1 \\ 0 & -2 & -1 \\ -1 & -1 & 0 \end{bmatrix}.$$

Note that we have already converted the problem into an unconstrained one.

If we directly apply the second order conditions for unconstrained problems, we only need to check whether H is negative definite to ensure that (x^*, y^*) is a strong local maximum.

The Hessian

A matrix is negative definite iff the determinants of its leading principal minors alternate in sign, with the first being negative, that is, using H_k to denote the kth leading principal minor of H, then $(-1)^k \det(H_k) > 0$.

$$\det(H_1) = \det([\mathfrak{L}_{xx}]) = \det([-2]) = -2 < 0,$$

$$\det(H_2) = \det\left(\begin{bmatrix} \mathfrak{L}_{xx} & \mathfrak{L}_{xy} \\ \mathfrak{L}_{yx} & \mathfrak{L}_{yy} \end{bmatrix}\right) = \det\left(\begin{bmatrix} -2 & 0 \\ 0 & -2 \end{bmatrix}\right) = 4 > 0,$$

$$\det(H_3) = \det\left(\begin{bmatrix} \mathfrak{L}_{xx} & \mathfrak{L}_{xy} & \mathfrak{L}_{x\lambda} \\ \mathfrak{L}_{yx} & \mathfrak{L}_{yy} & \mathfrak{L}_{y\lambda} \\ \mathfrak{L}_{\lambda x} & \mathfrak{L}_{\lambda y} & \mathfrak{L}_{\lambda\lambda} \end{bmatrix}\right)$$

$$= \det\left(\begin{bmatrix} -2 & 0 & -1 \\ 0 & -2 & -1 \\ -1 & -1 & 0 \end{bmatrix}\right) = 4 > 0.$$

The Hessian is NOT negative definite.

The Bordered Hessian

Although we have converted the problem into an unconstrained one, we cannot directly use the Hessian test since $\nabla g(x, y) \cdot (dx, dy) = dx + dy = 0$. The usual argument fails.

The right way to do it is to consider the border Hessian matrices.

$$\det(H_1) = \det\left(\begin{bmatrix} \mathfrak{L}_{xx} & \mathfrak{L}_{x\lambda} \\ \mathfrak{L}_{\lambda x} & \mathfrak{L}_{\lambda\lambda} \end{bmatrix}\right) = \det\left(\begin{bmatrix} -2 & -1 \\ -1 & 0 \end{bmatrix}\right) = -1 < 0,$$

$$\det(H_2) = \det\left(\begin{bmatrix} \mathfrak{L}_{xx} & \mathfrak{L}_{xy} & \mathfrak{L}_{x\lambda} \\ \mathfrak{L}_{yx} & \mathfrak{L}_{yy} & \mathfrak{L}_{y\lambda} \\ \mathfrak{L}_{\lambda x} & \mathfrak{L}_{\lambda y} & \mathfrak{L}_{\lambda\lambda} \end{bmatrix}\right)$$

$$= \det\left(\begin{bmatrix} -2 & 0 & -1 \\ 0 & -2 & -1 \\ -1 & -1 & 0 \end{bmatrix}\right) = 4 > 0.$$

The "border" is shown in "blue".

The Bordered Hessian - General Result

Theorem for the Bordered Hessian

Let $f, g_1, g_2, ..., g_m$ be twice continuously differentiable functions on $x_1, x_2, ..., x_n$, and $(x_1^*, x_2^*, ..., x_n^*)$ is a critical/stationary point for $\mathfrak{L}(x_1, x_2, ..., x_n; \lambda_1, \lambda_2, ..., \lambda_m) = f - \sum_{j=1}^{m} \lambda_j g_j$.

Suppose that the vectors $\nabla g_j, j = 1, 2, ..., m$, are linearly independent (no redundant constraint(s) at the critical point).

If the last $n - m$ principal minors of the bordered Hessian H (the Hessian of $\mathfrak{L}$ at the critical point) is such that the smallest minor has sign $(-1)^{m+1}$ and are alternating in sign, then $(x_1^*, x_2^*, ..., x_n^*)$ is a local maximum of f subject to the constraints $g_j = 0$.

Proof. See *Introduction to Mathematical Programming*, 3rd edition by Russell C. Walker.

The Bordered Hessian - General Result

Now we write the bordered Hessian equivalently in the following form.

$$H = \begin{bmatrix} \mathfrak{L}_{\lambda_1\lambda_1} & \mathfrak{L}_{\lambda_1\lambda_2} & \cdots & \mathfrak{L}_{\lambda_1\lambda_m} & \mathfrak{L}_{\lambda_1 x_1} & \mathfrak{L}_{\lambda_1 x_2} & \cdots & \mathfrak{L}_{\lambda_1 x_n} \\ \mathfrak{L}_{\lambda_2\lambda_1} & \mathfrak{L}_{\lambda_2\lambda_2} & \cdots & \mathfrak{L}_{\lambda_2\lambda_m} & \mathfrak{L}_{\lambda_2 x_1} & \mathfrak{L}_{\lambda_2 x_2} & \cdots & \mathfrak{L}_{\lambda_2 x_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathfrak{L}_{\lambda_m\lambda_1} & \mathfrak{L}_{\lambda_m\lambda_2} & \cdots & \mathfrak{L}_{\lambda_m\lambda_m} & \mathfrak{L}_{\lambda_m x_1} & \mathfrak{L}_{\lambda_m x_2} & \cdots & \mathfrak{L}_{\lambda_m x_n} \\ \mathfrak{L}_{x_1\lambda_1} & \mathfrak{L}_{x_1\lambda_2} & \cdots & \mathfrak{L}_{x_1\lambda_m} & \mathfrak{L}_{x_1 x_1} & \mathfrak{L}_{x_1 x_2} & \cdots & \mathfrak{L}_{x_1 x_n} \\ \mathfrak{L}_{x_2\lambda_1} & \mathfrak{L}_{x_2\lambda_2} & \cdots & \mathfrak{L}_{x_2\lambda_m} & \mathfrak{L}_{x_2 x_1} & \mathfrak{L}_{x_2 x_2} & \cdots & \mathfrak{L}_{x_2 x_n} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \mathfrak{L}_{x_n\lambda_1} & \mathfrak{L}_{x_n\lambda_2} & \cdots & \mathfrak{L}_{x_n\lambda_m} & \mathfrak{L}_{x_n x_1} & \mathfrak{L}_{x_n x_2} & \cdots & \mathfrak{L}_{x_n x_n} \end{bmatrix}.$$

The Bordered Hessian - An Illustration

For example, when $n = 2$ and $m = 1$, we only need to check the last $n - m = 1$ principal minor

$$H_3 = \begin{bmatrix} \mathfrak{L}_{\lambda_1\lambda_1} & \mathfrak{L}_{\lambda_1 x_1} & \mathfrak{L}_{\lambda_1 x_2} \\ \mathfrak{L}_{x_1\lambda_1} & \mathfrak{L}_{x_1 x_1} & \mathfrak{L}_{x_1 x_2} \\ \mathfrak{L}_{x_2\lambda_1} & \mathfrak{L}_{x_2 x_1} & \mathfrak{L}_{x_2 x_2} \end{bmatrix},$$

to see if $\det(H_3) > 0$.

For the problem in (42),

$$\det(H_3) = \det\left(\begin{bmatrix} 0 & -1 & -1 \\ -1 & -2 & 0 \\ -1 & 0 & -2 \end{bmatrix}\right) = 4,\ sign(\det(H_3)) = (-1)^{(1+1)}.$$

We have found a local maximum.

The Bordered Hessian - Another Illustration

Consider the following maximization problem.

$$\max_{x,y,z,w} \; -x^2 - y^2 - z^2 - w^2, \tag{43}$$
$$\text{s.t.} \;\; x - z - 2w + 2 = 0.$$
$$y + 2z + 3w - 6 = 0.$$

The Lagrangian is

$$\mathfrak{L}(x, y, \lambda) = -x^2 - y^2 - z^2 - w^2 - \lambda_1(x - z - 2w + 2) - \lambda_2(y + 2z + 3w - 6).$$

F.O.C.s: $\Longrightarrow \quad x^* = 1,\; y^* = 1,\; z^* = 1,\; w^* = 1,\; \lambda_1^* = -2,\; \lambda_2^* = -2.$

The Bordered Hessian - Another Illustration

The bordered Hessian is

$$H = \begin{bmatrix} \mathfrak{L}_{\lambda_1\lambda_1} & \mathfrak{L}_{\lambda_1\lambda_2} & \mathfrak{L}_{\lambda_1 x} & \mathfrak{L}_{\lambda_1 y} & \mathfrak{L}_{\lambda_1 z} & \mathfrak{L}_{\lambda_1 w} \\ \mathfrak{L}_{\lambda_2\lambda_1} & \mathfrak{L}_{\lambda_2\lambda_2} & \mathfrak{L}_{\lambda_2 x} & \mathfrak{L}_{\lambda_2 y} & \mathfrak{L}_{\lambda_2 z} & \mathfrak{L}_{\lambda_2 w} \\ \mathfrak{L}_{x\lambda_1} & \mathfrak{L}_{x\lambda_2} & \mathfrak{L}_{xx} & \mathfrak{L}_{xy} & \mathfrak{L}_{xz} & \mathfrak{L}_{xw} \\ \mathfrak{L}_{y\lambda_1} & \mathfrak{L}_{y\lambda_2} & \mathfrak{L}_{yx} & \mathfrak{L}_{yy} & \mathfrak{L}_{yz} & \mathfrak{L}_{yw} \\ \mathfrak{L}_{z\lambda_1} & \mathfrak{L}_{z\lambda_2} & \mathfrak{L}_{zx} & \mathfrak{L}_{zy} & \mathfrak{L}_{zz} & \mathfrak{L}_{zw} \\ \mathfrak{L}_{w\lambda_1} & \mathfrak{L}_{w\lambda_2} & \mathfrak{L}_{wx} & \mathfrak{L}_{wy} & \mathfrak{L}_{wz} & \mathfrak{L}_{ww} \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0 & -1 & 0 & 1 & 2 \\ 0 & 0 & 0 & -1 & -2 & -3 \\ -1 & 0 & -2 & 0 & 0 & 0 \\ 0 & -1 & 0 & -2 & 0 & 0 \\ 1 & -2 & 0 & 0 & -2 & 0 \\ 2 & -3 & 0 & 0 & 0 & -2 \end{bmatrix}$$

The Bordered Hessian - Another Illustration

we need to check the last $n - m = 4 - 2 = 2$ principal minors.

$$H_5 = \begin{bmatrix} 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & -1 & -2 \\ -1 & 0 & -2 & 0 & 0 \\ 0 & -1 & 0 & -2 & 0 \\ 1 & -2 & 0 & 0 & -2 \end{bmatrix}, \quad \begin{array}{c} sign(\det(H_5)) \\ = sign(-12) \\ = (-1)^{(2+1)}. \end{array}$$

$$H_6 = \begin{bmatrix} 0 & 0 & -1 & 0 & 1 & 2 \\ 0 & 0 & 0 & -1 & -2 & -3 \\ -1 & 0 & -2 & 0 & 0 & 0 \\ 0 & -1 & 0 & -2 & 0 & 0 \\ 1 & -2 & 0 & 0 & -2 & 0 \\ 2 & -3 & 0 & 0 & 0 & -2 \end{bmatrix}, \quad \begin{array}{c} sign(\det(H_6)) \\ = sign(80) \\ = (-1)^{(2+2)}. \end{array}$$

A Final Word

1. Intuition for the bordered Hessian method: As we have discussed before, because of the m (linearly independent) equality constraints, we can only check for directions of change for the critical point in the subspace $\mathbf{R}^{n-m}$ that is perpendicular to all ∇g_j. This behaviour is captured by the last $n-m$ minors of the Hessian.
2. If the bordered Hessian does not satisfy the conditions for a maximizer, we cannot generally determine whether it is a minimizer or a saddle point.
3. To find a minimizer, say x^*, of f, use the fact that x^* is a maximizer of $-f$. Then use the above algorithm.
4. If there are also inequality constraints ($h_j(x) \leq 0$), e.g., no short selling, we need to invoke the Karush-Kuhn-Tucker Theorem.[3]

[3] We will not cover it for now.

Minimum Variance Portfolio

Recall the problem of finding the minimum variance portfolio of three uncorrelated assets, with expected returns $\bar{r}_1 = 1$, $\bar{r}_2 = 2$ and $\bar{r}_3 = 3$, and the same variance $\sigma^2 = 1$. ($\sigma_{ij} = 0,\ i \neq j$.)

$$\mathfrak{L} = \frac{1}{2}\sum_{i=1}^{3} w_i^2 - \lambda\left(\sum_{i=1}^{3} i w_i - \bar{r}_p\right) - \zeta\left(\sum_{i=1}^{3} w_i - 1\right),$$

where λ and ζ are the Lagrangian Multipliers.

Minimum Variance Portfolio

We convert the problem into finding a maximum.

$$\mathfrak{L} = -\frac{1}{2}\sum_{i=1}^{3} w_i^2 - \lambda\left(\sum_{i=1}^{3} i w_i - \bar{r}_p\right) - \zeta\left(\sum_{i=1}^{3} w_i - 1\right).$$

The F.O.C.s are:

$$\begin{aligned} -w_1^* - \lambda^* - \zeta^* &= 0, \\ -w_2^* - 2\lambda^* - \zeta^* &= 0, \\ -w_3^* - 3\lambda^* - \zeta^* &= 0, \\ -(w_1^* + 2w_2^* + 3w_3^* - \bar{r}_p) &= 0, \\ -(w_1^* + w_2^* + w_3^* - 1) &= 0. \end{aligned}$$

$$\Longrightarrow \quad w_1^* = 1/3 - (\bar{r}_p/2 - 1),\ w_2^* = 1/3,\ w_3^* = 1/3 + (\bar{r}_p/2 - 1),$$
$$\lambda^* = -(\bar{r}_p/2 - 1),\ \zeta^* = -1/3 + (\bar{r}_p - 2).$$

Minimum Variance Portfolio

The bordered Hessian is

$$H = \begin{bmatrix} 0 & 0 & -1 & -2 & -3 \\ 0 & 0 & -1 & -1 & -1 \\ -1 & -1 & -1 & 0 & 0 \\ -2 & -1 & 0 & -1 & 0 \\ -3 & -1 & 0 & 0 & -1 \end{bmatrix}.$$

$$sign(\det(H_5)) = sign(\det(H)) = sign(-6) = (-1)^{(2+1)}.$$

So (w_1^*, w_2^*, w_3^*) is indeed a maximizer to this problem, hence a minimizer to the original problem.

The General Problem

Now go back to the general mean-variance problem.

$$\mathfrak{L} = \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} w_i \sigma_{ij} w_j - \lambda\left(\sum_{i=1}^{N} w_i \bar{r}_i - \bar{r}_p\right) - \zeta\left(\sum_{i=1}^{N} w_i - 1\right).$$

The F.O.C.s are

$$\sum_{j=1}^{N} \sigma_{ij} w_j^* - \lambda^* \bar{r}_i - \zeta^* = 0, \;\; for\ i = 1, 2, ..., N,$$

$$\sum_{i=1}^{N} w_i^* \bar{r}_i = \bar{r}_p,$$

$$\sum_{i=1}^{N} w_i^* = 1.$$

Two-Fund Separation

Note that the F.O.C.s are all linear functions of $(w_1, w_2, w_3, \lambda, \zeta)$. If $x^1 = (w_1^1, w_2^1, w_3^1, \lambda^1, \zeta^1)$ is a known solution for $\bar{r}_p^1$, and $x^2 = (w_1^2, w_2^2, w_3^2, \lambda^2, \zeta^2)$ is a known solution for $\bar{r}_p^2$, then, $\forall \alpha$, $\alpha x^1 + (1-\alpha)x^2$ is a solution for $\alpha \bar{r}_p^1 + (1-\alpha)\bar{r}_p^2$. (Check this yourself.)

The Two-Fund Theorem

Once two efficient funds (minimum variance portfolios) are established, investors seeking efficient investment in funds, in terms of mean-variance, can duplicate any efficient portfolio by simply combining the two.

It is also referred to as the Mutual Fund Separation Theorem. Investors need not purchase individual securities. Two mutual funds would be a complete service.

Two-Fund Separation - Discussions

The Two-Fund Theorem relies crucially on the following assumptions.

1. Investors only care mean and variance.
2. Investors have the same assessment of the means, variances and covariances.
3. A single-period investment horizon.

All the three assumptions are tenuous in reality. But the theorem provides a good way to understand the investment process.

- For instance, if investors care more than mean and variance and invest for multiple periods, the two-fund separation is no long attained. However, we may have a three-fund separation.[4]

[4] See Merton, Robert C., 1973. An intertemporal capital asset pricing model, Econometrica 41(5), 867-887. We will not cover it in this course.

Adding a Riskfree Asset

So far we have focused on risky assets with $\sigma > 0$.

Introduction of a riskfree asset with $\sigma = 0$ enables borrowing and lending at the riskfree rate r_f.

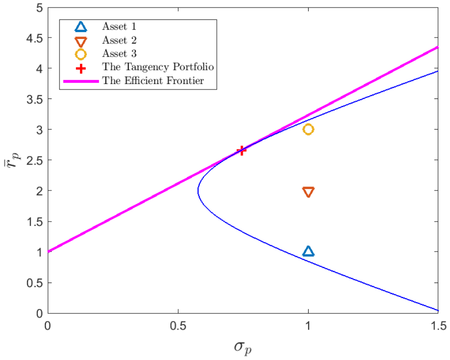
Perhaps surprisingly, adding one more riskfree asset causes a mathematical degeneracy hence greatly simplifies the shape of the efficient frontier.

For a portfolio with a weight α ($\alpha \leq 1$) in the riskfree asset and a weight $1 - \alpha$ in a risky asset with mean return of $\bar{r}$ and standard deviation σ,

$$\bar{r}_p = \alpha r_f + (1 - \alpha)\bar{r}, \ \sigma_p = (1 - \alpha)\sigma.$$

The Efficient Frontier

Let $r_f = 1$. The frontier with three independent risky assets but without the riskfree asset is described by $\sigma_p = \sqrt{\bar{r}_p^2/2 - 2\bar{r}_p + 7/3}$. Calculate the weights of the tangency portfolio by yourself. (Answer: (1/9,1/3,5/9))



The One-Fund Theorem

Any efficient portfolio is a combination of riskfree borrowing/lending and a single fund of risky assets.

Note that on the aggregate level, borrowing and lending cancel out. The weights would be 100% in the risky fund, and 0% in the riskfree asset.

We are ready to study market equilibrium, for instance, the Capital Asset Pricing Model.

Question

Question on the Tangency Portfolio

What are the weights of the tangency portfolio?

Question

Question on the Tangency Portfolio

What are the weights of the tangency portfolio?

The frontier of risky assets is the curve $\sigma_p^2 = \bar{r}_p^2/2 - 2\bar{r}_p + 7/3$. Differentiate both sides, we obtain:

$$2\sigma_p d\sigma_p = \bar{r}_p d\bar{r}_p - 2d\bar{r}_p, \quad \Longrightarrow \quad \frac{d\bar{r}_p}{d\sigma_p} = \frac{2\sigma_p}{\bar{r}_p - 2}$$

The slope of the tangency line is $\frac{\bar{r}_p - r_f}{\sigma_p}$. Let the two slopes equal.

$$\Longrightarrow \quad \bar{r}_p = \frac{14/3 - 2r_f}{2 - r_f}.$$

Then $w_1 = 1/3 - (\bar{r}_p/2 - 1)$, $w_2 = 1/3$, $w_3 = 1/3 + (\bar{r}_p/2 - 1)$.