Applications of Linear Algebra in Mean-Variance Portfolio Theory and Linear Regression

Asher Reedy

University of Illinois at Urbana Champaign

ASRM 406: Linear Algebra with Financial Applications

Eric Icaza

12/12/2025

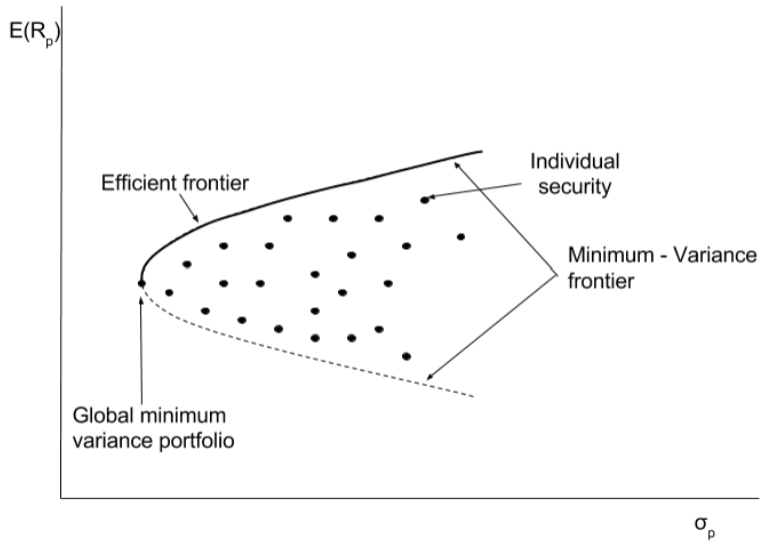
Linear algebra is a foundational branch of mathematics that has a variety of applications across many disciplines. In this paper, we explore two key applications: Mean-Variance Portfolio Theory, and Linear Regression. The first chapter focuses on Mean-Variance Portfolio Theory. We express the components of portfolios as matrices and use matrix operations and calculus to derive key formulas and apply Modern Portfolio Theory concepts. The second chapter explains the application of vector-subspace projections and distance calculations to the topic of linear regression and least squares. The goal of this paper is to give an introductory presentation of two significant applications of linear algebra.

1.1 Modern Portfolio Theory and CAPM

Modern Portfolio Theory is a framework created in 1952 by Harry Markowitz in the celebrated paper *Portfolio Selection*. This paper established a formal connection between risk and expected return, moving the focus away from the performance of individual stocks towards the idea of risk-minimization through diversification and the efficient frontier. Diversification is the central strategy of MPT and aims to reduce the overall volatility of a portfolio by selecting a mixture of assets with low or negative covariance. By choosing a low/negatively correlated portfolio of assets, a downturn in one asset will only slightly bring down the rest of the portfolio, it may even cause an increase in the negatively correlated assets. This effect softens the impact of loss and provides stability to the portfolio. When selecting assets, investors are interested in securing the highest expected

return given the level of risk they are comfortable with. The efficient frontier curve illustrates this tradeoff between risk and returns by modeling the set of portfolios that offer the highest possible return given a specific risk level. Optimal portfolios lie on the efficient frontier; any portfolio that lies below the curve is suboptimal, offering less return for the same risk level.

Figure 1.1



Modern Portfolio Theory differentiates between two types of risk: systematic risk which describes the overall risk of the entire market, and unsystematic risk which describes the unique risk of a specific asset or industry. Diversification can be used to mitigate specific unsystematic risks but is unable to address market risk. The Capital Asset Pricing Model (CAPM) provides a model to relate systematic market risks to the expected return of an investment by considering the asset's beta, risk-free rate, and market risk premium.

$$E(R_i) = R_{rf} + \beta_i \times \big(E(R_m) - R_{rf}\big)$$

Where:

- $E(R_i)$ = expected return of investment
- R_{rf} = risk-free rate
- β_i = beta, the investment's volatility in relation to the overall market
- $\big(E(R_m) - R_{rf}\big)$ = market risk premium, additional return investors expect from investing in the risky market instead of a risk-free asset

Theoretically, CAPM assumes the existence of a fully diversified, efficient market portfolio. Because it is fully diversified, the market portfolio holds no unsystematic risks and is only exposed to market risk. In this model, investor portfolios contain a combination of the market portfolio and a risk-free asset, according to their desired risk-level.

The Capital Asset Pricing Model makes several simplifying assumptions, for example it assumes all investors have the same expectations about risk and return and will all arrive at the same set of efficient portfolios. In reality, individual investors have differing perceptions and goals. CAPM also assumes investor rationality and risk-aversion, which often conflict with human emotion. There are several more simplifying assumptions such as: idealistic markets, equal access to information, and linear relation between risk and return. Overall CAPM provides a simple but useful model for evaluating investments and models the MPT principle that higher returns can only be achieved by taking on more risk.

The following sections will apply linear algebra concepts to the MPT concepts of diversification and the efficient frontier to solve problems in Mean-Variance Portfolio Analysis.

1.2 Derivations and Portfolio with Given Target Return

In this section, we define matrices to describe a portfolio, we then derive a formula for the variance of a portfolio. After deriving key formulas, we apply them to find the weights and variance of the minimum variance portfolio for a given target return.

Assume a portfolio with n assets, expected return vector of $\vec{\mu}$, covariance matrix $\sum$, return vector $\vec{R}$, asset price vector $\vec{S}$, weight vector $\vec{w}$, and portfolio rate of return for a given weight vector $R_p(\vec{w})$.

$$\vec{S}(t) = \begin{bmatrix} S_1(t) \\ \vdots \\ S_n(t) \end{bmatrix}, \qquad \vec{R} = \begin{bmatrix} R_1 \\ \vdots \\ R_n \end{bmatrix} = \begin{bmatrix} \frac{S_1(1)-S_1(0)}{S_1(0)} \\ \vdots \\ \frac{S_n(1)-S_n(0)}{S_n(0)} \end{bmatrix}, \qquad \vec{\mu} = E[\vec{R}] = \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}$$

$$\vec{w} = [w_1 \quad \dots \quad w_n]^T, where \ \sum_{n=1}^{N} w_n = 1$$

$$R_p(\vec{w}) = \sum_{n=1}^{N} w_n R_n = \vec{w}^T \vec{R}$$

$$\Sigma = E\left[\left(\vec{R}-\vec{\mu}\right)\left(\vec{R}-\vec{\mu}\right)^T\right] = E\left[\begin{bmatrix} R_1-\mu_1 \\ \vdots \\ R_n-\mu_n \end{bmatrix}[R_1-\mu_1 \quad \dots \quad R_n-\mu_n]\right]$$

$$= E\left[\begin{bmatrix} (R_1-\mu_1)(R_1-\mu_1) & \cdots & (R_1-\mu_1)(R_n-\mu_n) \\ \vdots & \ddots & \vdots \\ (R_n-\mu_n)(R_1-\mu_1) & \cdots & (R_n-\mu_n)(R_n-\mu_n) \end{bmatrix}\right] = \begin{bmatrix} Cov(R_1,R_1) & \cdots & Cov(R_1,R_n) \\ \vdots & \ddots & \vdots \\ Cov(R_n,R_1) & \cdots & Cov(R_n,R_n) \end{bmatrix}$$

$$= \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n_1} & \cdots & \sigma_{nn} \end{bmatrix}$$

$\sigma_{ij} = \sigma_{ji}\ and\ \sigma_{ii} = Var[R_1]$

For each weight vector $\vec{w}$ we can create a formula for the variance of the portfolio's return σ_P^2.

$$\sigma_P^2 = Var\left(R_p(\vec{w})\right) = Var(R_1w_1 + \cdots + R_nw_n) = \sum_{i=1}^n \sum_{j=1}^n w_i w_j \sigma_{ij} = \vec{w}^T \begin{bmatrix} \sum_{j=1}^n w_j\sigma_{1j} \\ \vdots \\ \sum_{j=1}^n w_j\sigma_{nj} \end{bmatrix}$$

$$= \vec{w}^T \begin{bmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n_1} & \cdots & \sigma_{nn} \end{bmatrix} \vec{w} = \vec{w}^T \Sigma \vec{w}$$

Now that these concepts are defined as matrices, we can use linear algebra and optimization to derive formulas for the optimal weights and smallest variance given a specific target return. This is a minimization problem where we aim to minimize σ_P^2 with respect to two boundary conditions.

$$\min \vec{w}^T \Sigma \vec{w}$$
$$such\ that\ \vec{w}^T\mu = r_p\ and\ \vec{w}^T\vec{1} = 1$$

Where:

- r_p = target return
- $\vec{w}^T$ = portfolio weight vector
- Σ = covariance matrix
- $\vec{\mu}$ = expectation of the return vector

The first condition dictates that the expected portfolio return $\vec{w}^T\mu$ is equal to the investor's target return. The second condition ensures the weights sum to 1. In dimensions higher than 1, we need to use Lagrange multipliers to solve this problem. This approach

uses the Lagrangian function, where $f(x)$ is the objective function and $g(x) = 0$ are the boundary constraints.

$$L(x, \lambda) \overset{def}{=} f(x) + \lambda \times g(x)$$

We can create a system of equations by setting the partial derivatives of $L = 0$.

$$\frac{\partial L}{\partial x} = 0 \; and \; \frac{\partial L}{\partial \lambda} = 0$$

In the context of our problem for a portfolio with n assets,

$$f(w) = Var(R_p) = \vec{w}^T \Sigma \vec{w}$$
$$g_1 = \vec{w}^T \mu - r_p = 0$$
$$g_2 = \vec{w}^T \vec{1} - 1 = 0$$
$$L(w, \lambda, \gamma) = \vec{w}^T \Sigma \vec{w} + \lambda \times (\vec{w}^T \vec{1} - 1) + \gamma(\vec{w}^T \mu - r_p)$$

After applying solving the system created from the Lagrange multiplier process, we obtain a compact formula for the weight vector of the optimal portfolio.

$$\vec{w}^* = \Sigma^{-1}[\vec{\mu} \;\; \vec{1}] A^{-1} \begin{bmatrix} r_P \\ 1 \end{bmatrix}$$
$$\text{Where } A = \begin{bmatrix} \vec{\mu}^T \Sigma^{-1} \vec{\mu} & \vec{\mu}^T \Sigma^{-1} \vec{1} \\ \vec{1}^T \Sigma^{-1} \vec{\mu} & \vec{1}^T \Sigma^{-1} \vec{1} \end{bmatrix}$$

This weight vector will provide a point on the efficient frontier curve corresponding to the given target rate. We can find the optimal variance for a specific return by plugging the weight vector into the general portfolio return variance equation.

$$\sigma_P^2(\vec{w}^*) = (\vec{w}^*)^T \Sigma \, \vec{w}^* = \frac{1}{\vec{1}^T \Sigma^{-1} \vec{1}}$$

To illustrate this process, we provide an example. Consider a three-asset portfolio with expected return vector $\vec{\mu}$, covariance matrix Σ, and target return r_P.

$$\text{Where } \vec{\mu} = \begin{bmatrix} 0.20 \\ 0.10 \\ 0.08 \end{bmatrix}, \Sigma = \begin{bmatrix} 0.3 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.2 \end{bmatrix}, and \; r_P = 0.16$$

Using the formulas we derived from the minimization problem; we can solve for the weight vector and variance for the target return.

$$\Sigma^{-1} = \frac{Adj(\Sigma)}{Det(\Sigma)} = \frac{1}{0.006}\begin{bmatrix} 0.02 & 0 & 0 \\ 0 & 0.06 & 0 \\ 0 & 0 & 0.03 \end{bmatrix} = \begin{bmatrix} \frac{10}{3} & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \text{(Cramer's Rule)}$$

$$A = \begin{bmatrix} [.20 \quad .10 \quad .08]\begin{bmatrix} 3.33 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 5 \end{bmatrix}\begin{bmatrix} .20 \\ .10 \\ .08 \end{bmatrix} & [.20 \quad .10 \quad .08]\begin{bmatrix} 3.33 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 5 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \\ [1 \quad 1 \quad 1]\begin{bmatrix} 3.33 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 5 \end{bmatrix}\begin{bmatrix} .20 \\ .10 \\ .08 \end{bmatrix} & [1 \quad 1 \quad 1]\begin{bmatrix} 3.33 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 5 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \end{bmatrix} =$$

$$\begin{bmatrix} 0.265 & 2.067 \\ 2.067 & 18.333 \end{bmatrix}$$

$$A^{-1} = \frac{1}{(0.265)(18.333) - (2.067)^2}\begin{bmatrix} 18.333 & -2.067 \\ -2.067 & 0.265 \end{bmatrix} = \begin{bmatrix} 31.298 & -3.529 \\ -3.529 & 0.452 \end{bmatrix}$$

Finally, we can solve the weight vector and the variance:

$$\vec{w}^* = \Sigma^{-1}\begin{bmatrix} \vec{\mu} & \vec{1} \end{bmatrix} A^{-1}\begin{bmatrix} r_P \\ 1 \end{bmatrix} = \begin{bmatrix} \frac{10}{3} & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 5 \end{bmatrix}\begin{bmatrix} \begin{bmatrix} 0.20 \\ 0.10 \\ 0.08 \end{bmatrix} & \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \end{bmatrix}\begin{bmatrix} 31.298 & -3.529 \\ -3.529 & 0.452 \end{bmatrix}\begin{bmatrix} 0.16 \\ 1 \end{bmatrix}$$

$$= \begin{bmatrix} 0.610 \\ 0.352 \\ 0.028 \end{bmatrix}$$

$$\sigma_P^2(\vec{w}^*) = (\vec{w}^*)^T \Sigma\, \vec{w}^* = [0.610 \quad 0.352 \quad 0.028]\begin{bmatrix} 0.3 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.2 \end{bmatrix}\begin{bmatrix} 0.610 \\ 0.352 \\ 0.028 \end{bmatrix}$$

$$= 0.124$$

If we evaluated this portfolio for every possible rate of return, we would exactly trace the efficient frontier curve. We can see that expressing portfolio information as matrices allows us to efficiently determine the optimal weights for a given target return. The use of linear algebra is far cleaner than a substitution / algebraic approach, especially for portfolios with many assets.

1.3 Global Minimum Variance Portfolio

In this section we apply the approach used in Section 1.2 to derive formulas for the weights and variance of the global minimum variance portfolio. As the name implies, this is the portfolio that can be crafted with the least possible variance, this idea is depicted in Figure 1.1. To solve this problem, we take a similar minimization approach as in Section 1.2, however we no longer need to consider the $\vec{w}^T\mu = r_p$ constraint in the Lagrangian.

$$\min \vec{w}^T \Sigma \vec{w}$$
$$such\ that\ \vec{w}^T\vec{1} = 1$$

Where:

- $\vec{w}^T$ = portfolio weight vector
- Σ = covariance matrix

For a portfolio with n assets:

$$f(w) = Var(R_p) = \vec{w}^T\Sigma\vec{w}$$
$$g_2 = \vec{w}^T\vec{1} - 1 = 0$$
$$L(w, \lambda, \gamma) = \vec{w}^T\Sigma\vec{w} + \lambda \times (\vec{w}^T\vec{1} - 1)$$

After applying solving the system created from the Lagrange multiplier process, we obtain formulas for the weight vector and variance of the global minimum variance portfolio.

$$\vec{w}^* = \frac{\Sigma^{-1}\vec{1}}{\vec{1}^T\Sigma^{-1}\vec{1}}$$

$$\sigma_P^2(\vec{w}^*) = (\vec{w}^*)^T\Sigma\,\vec{w}^* = \frac{1}{\vec{1}^T\Sigma^{-1}\vec{1}}$$

To demonstrate the process, we can solve the previous three-asset portfolio for the global minimum weight vector and variance.

$$\text{Where } \vec{\mu} = \begin{bmatrix} 0.20 \\ 0.10 \\ 0.08 \end{bmatrix}, \Sigma = \begin{bmatrix} 0.3 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.2 \end{bmatrix}, \Sigma^{-1} = \begin{bmatrix} \frac{10}{3} & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

$$\vec{w}^* = \frac{\begin{bmatrix} \frac{10}{3} & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}{\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{10}{3} & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}} = \frac{1}{18.\overline{33}} \times \begin{bmatrix} 10 \\ 3 \\ 10 \\ 5 \end{bmatrix} = \begin{bmatrix} 0.182 \\ 0.545 \\ 0.273 \end{bmatrix}$$

$$\sigma_P^2(\vec{w}^*) = \frac{1}{\begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} \frac{10}{3} & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 5 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}} = 0.0545$$

1.4 Portfolio for Given Target Risk, Introducing a Risk-Free Asset

Most commonly, investors will focus on investments with a specified target return and worry about the volatility associated with the target return afterward. However, we can also create a formula to find the optimal portfolio weights and returns, given a target risk level. Here, expected returns are maximized under a constrained variance.

Consider a portfolio of n assets where:

- $\Sigma =$ covariance matrix
- $\vec{\mu} =$ expectation of the return vector
- $\sigma_P = target\ portfolio\ risk$
- $\mu_P^* = optimal\ return\ for\ given\ risk$
- $w_P^* = optimal\ weights\ for\ portfolio\ of\ given\ risk$

Problem Statement:

$$\max \vec{\mu}^T \vec{w}$$
$$such\ that\ \vec{w}^T \Sigma \vec{w} = \sigma_P\ and\ \vec{w}^T \vec{1} = 1$$

Solving the optimization problem yields:

$$\mu_P^* = \frac{B}{C} + \frac{\sqrt{\Delta}}{C}\sqrt{C\sigma_P - 1}$$

$$w_P^* = \frac{A - B\mu_P^*}{\Delta}\Sigma^{-1}\vec{1} + \frac{B\mu_P^* - C}{\Delta}\Sigma^{-1}\vec{\mu}$$

Where:

$$A = \mu^T \Sigma^{-1} \mu$$
$$B = \mu^T \Sigma^{-1} \vec{1}$$
$$C = \vec{1}^T \Sigma^{-1} \vec{1}$$
$$\Delta = AC - B^2$$

To demonstrate, we consider the same three-asset portfolio from Section 1.2 and Section 1.3 with a target risk of $\sigma_P = 0.124$.

$$\vec{\mu} = \begin{bmatrix} 0.20 \\ 0.10 \\ 0.08 \end{bmatrix}, \Sigma = \begin{bmatrix} 0.3 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 0.2 \end{bmatrix}, \Sigma^{-1} = \begin{bmatrix} \frac{10}{3} & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 5 \end{bmatrix}, \sigma_P = 0.124$$

$$\mu_P^* = 0.16$$
$$w_P^* = \begin{bmatrix} 0.610 \\ 0.352 \\ 0.028 \end{bmatrix}$$

Applying the target risk formulas results in an optimal return of 0.16. This is consistent with the example from Section 1.2. Where 0.16 was the target return for the same portfolio that resulted in an optimal risk of $\sigma_P = 0.124$. The formulas agree with each other and verify the efficient frontier.

Introducing a Risk-Free Asset

The problem of finding the optimal portfolio for a target risk becomes easier when a risk-free asset is introduced. The approach follows the assumptions of CAPM where we consider a single optimal risky portfolio and a risk-free asset, the investor chooses how much they want to invest in each according to their specified risk-level.

This optimal risky portfolio is called the tangency portfolio and is significant because its weights maximize the Sharpe ratio. The Sharpe ratio describes the amount of return gained for each unit of risk.

$$Sharpe\ Ratio: \frac{E[R_P] - E[R_f]}{\sigma_P} = \frac{\vec{w}^T \vec{\mu} - R_f}{\sqrt{\vec{w}^T \Sigma \vec{w}}}$$

Where R_f is the risk-free portfolio.

Our goal is to construct the Capital Market Line (CML) which depicts the most efficient trade-off between risk and return. The slope of the CML is the maximized Sharpe ratio

from the tangency portfolio. Maximizing the Sharpe Ratio results in the normalized tangency portfolio:

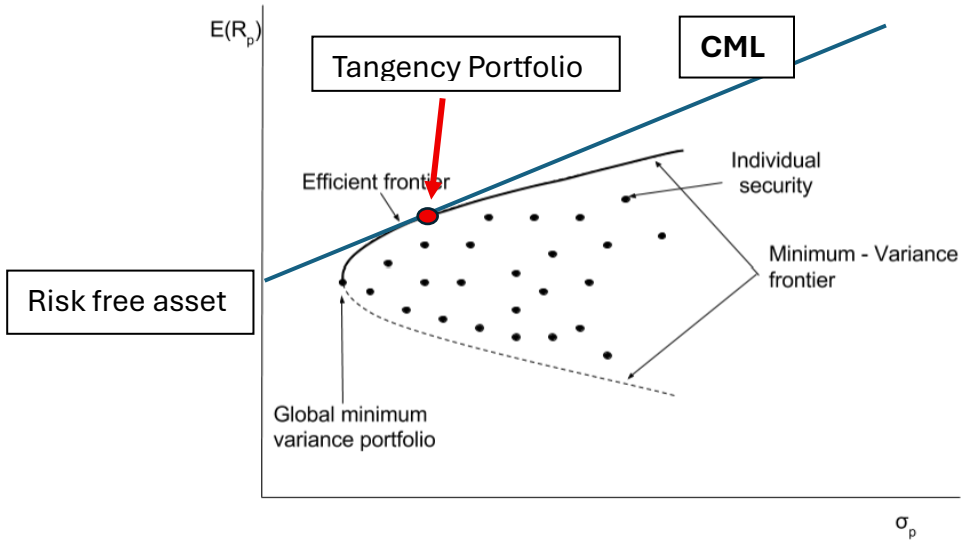
$$\vec{w}_{tan} = \frac{\Sigma^{-1}(\vec{\mu} - R_f\vec{1})}{1^T\Sigma^{-1}(\vec{\mu} - R_f\vec{1})}$$

With return $\vec{\mu}_{tan} = \vec{w}_{tan}^T\vec{\mu}$ and standard deviation $\sigma_{tan} = \vec{w}_{tan}^T\Sigma\overrightarrow{w_{tan}}$. The resulting portfolio has expected return $E(R_p) = R_f + x(\vec{\mu}_{tan} - R_f)$, and standard deviation $\sigma_P = x\sigma_{tan}$. The investor chooses to invest x in the risky portfolio and $1 - x$ in the risk-free asset. We eliminate x to find the CML.

The Capital Market Line is Given by:

$$E[R_P] = R_f + Maximized\ Sharpe\ Ratio \times \sigma_P = R_f + \left(\frac{\vec{\mu}_{tan} - R_f}{\sigma_{tan}}\right)\sigma_P$$

Figure 1.2



We can see from Figure 1.2. Once a risk-free asset is introduced, the efficient frontier becomes the Capital Market Line. The CML offers higher returns for the same levels of risk compared to portfolios containing only risky assets.

1.5 Extension of Mean-Variance Portfolio Analysis

Fischer Black and Robert Litterman created the Black-Litterman (BL) model in 1990 with the goal of incorporating subjective investor sentiments into the MPT portfolio allocation model. This approach starts by considering the returns given by the optimal

market portfolio described in CAPM, next the model incorporates subjective investor insights about the performance of each asset through a process called reverse optimization. The process results in a portfolio optimized according to the investor's risk-level and subjective market outlook. The BL model has several advantages and disadvantages. Firstly, traditional Mean-Variance analysis often yields portfolios that are highly sensitive to small changes in inputs. The BL model is less prone to input sensitivity because it starts with the market portfolio and makes small adjustments based on the investor's views. In comparison, this approach is much more stable than Mean-Variance analysis. Additionally, BL offers a mathematically sound way to incorporate subjective views into portfolio allocation, though these views may lead to losses if they are poorly chosen. BL is also subjected to limitations the model inherits by incorporating the CAPM market portfolio and its unrealistic assumptions. While it has some limitations, the BL model is a significant extension of Modern Portfolio theory and is used widely in the finance industry among institutional investors and asset management firms.

2.1 Distances and Orthogonal Projections

In chapter 2, we discuss linear algebra's applications to linear regression models and least squares. The goal of linear regression is to construct a "best fit" line using past data to approximate the general relationship between the dependent and independent variables. Least squares is a common method of fitting a linear regression model and aims to minimize the sum of the squared errors between the predictors and responses in the data. Before diving into these topics and their relation to linear algebra, it's important to review the relevant linear algebra ideas of distance and orthogonal projections.

In $\mathbb{R}$ The distance of two real numbers a and b is defined as $|b-a|$. The distance between two vectors u and v in $\mathbb{R}^n$ is the length of vector $u-v$ or $v-u$.

$$dist(u,v) = \|u-v\| = \sqrt{(u_1-v_1)^2+\cdots+(u_n-v_n)^2}$$

If S is a subspace in $\mathbb{R}^n$ spanned by a set of basis vectors, the distance from a vector y and the subspace is the smallest possible distance between y and any vector v in S. In

order to find the distance, we must calculate the orthogonal projection of y onto S, the distance we are interested in is the magnitude of the orthogonal component of this projection.

$$Distance = \|y - proj_S y\|$$

Figure 2.1

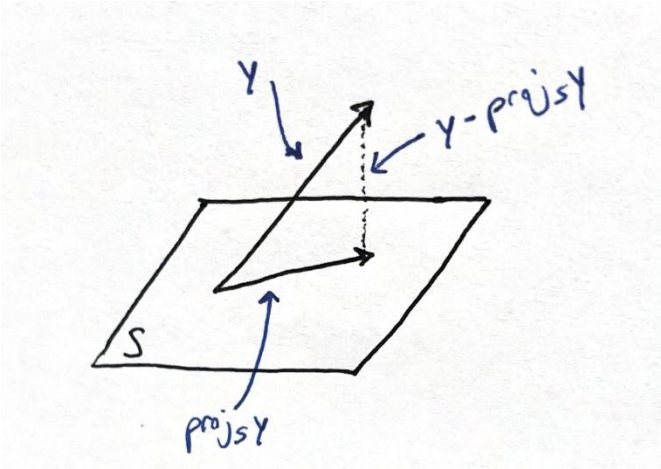


Figure 2.1 shows the orthogonal projection and orthogonal component of a vector y onto a two-dimensional subspace.

In the context of linear regression, $proj_S y$ is the model's approximation for y, the distance is interpreted as the residual or error between the observed data and the best fit line, the least squares approach aims to minimize these errors, we'll explore this more in later sections.

Conceptually, a projection of a vector onto a subspace is the vector inside the subspace that is closest to the original vector, geometrically, the projection can be thought of as the shadow of the vector cast onto the subspace. Which formula we use for the projection depends on whether the bases for the subspace are orthogonal, that is, the dot product of every combination of basis vectors equals zero. In the case that the subspace has orthogonal bases, the projection is calculated as the sum of the projections of the vector y onto each basis vector in the subspace.

$$proj_S y = \sum_{i=1}^{n} \frac{v_i \cdot y}{v_i \cdot v_i} v_i$$

$$Where\ vectors\ \{v_1, v_2, \dots, v_n\}\ are\ an\ orthogonal\ basis\ for\ S$$

In the general case where the basis vectors are not orthogonal, the projection formula is not as clean and relies on the projection matrix P.

$$P = A(A^T A)^{-1} A^T$$

$$proj_S y = Py = A(A^T A)^{-1} A^T y$$

$$Where\ A\ is\ a\ matrix\ whose\ columns\ form\ the\ basis\ for\ S$$

These ideas of distance and projection form the foundation for linear regression and solving the least-squares problem.

2.2 The Linear Regression Model

Linear regression models are used to describe the relationship between a dependent/ response variable y and one or more independent/predictor variables $x_1, \ldots, x_n$ by creating a best fit line through the data. The "simplest" linear regression model is simple linear regression which considers the relationship between one independent variable y and one dependent variable x.

$$y = \beta_0 + \beta_1 x$$

Figure 2.2

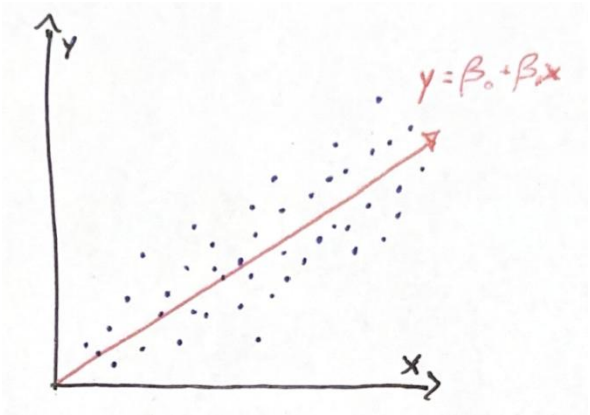


Figure 2.2 shows a simple linear regression model.

However, there are many situations where there is more than one predictor variable that influences a response variable. The simple linear regression model can be generalized by the general linear model which allows for multiple linear regression.

$$y = X\beta + \varepsilon$$

Where:

- X is an $n\ x\ p$ matrix whose columns represent predictor variables
- β is a $p\ x\ 1$ vector of coefficients
- y is an $n\ x\ 1$ vector of observed responses
- ε is an $n\ x\ 1$ vector of random errors

Each row of X corresponds to a single observation, the model for a single data point is:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_{ip} x_{ip} + \varepsilon_i$$

To construct the best possible line of best fit for our data, we need to estimate the coefficient vector β such that it minimizes the sum of squared residuals, i.e., differences between the observed responses y and the model's prediction for y. This is a least-squares problem and is very similar to the vector-subset distance problem from Section 2.1. We will solve this problem in the next section.

Before moving on, it's important to understand several assumptions that linear regression models make. Firstly, linear regression assumes linearity between the predictor and response variables, the response changes proportionally with the predictor. Consequently, linear regression struggles to describe non-linear patterns, in this case it would be better to use non-linear regression where we can choose functions that better capture the general shape of the data.

$$y_i = \beta_0 + \beta_1 f_1(x) + \cdots + \beta_n f_n(x)$$

Another assumption is the homoscedasticity of residuals which assumes that the variance of the residuals is constant for all independent variables. If the variance of residuals has a strong increasing or decreasing pattern, the estimations for coefficients may be unreliable. The model also assumes that the residuals follow a normal distribution and are independent of one another, data that violates these assumptions will receive a less reliable model from linear regression. Lastly, the multi linear regression model assumes there is no high correlation between predictors (multicollinearity), high correlation makes it difficult to determine the individual effect of each predictor.

2.3 Least Squares Approximation

It is well known that a linear system of equations $Ax = b$ can have either one solution, infinitely many solutions, or no solutions for x. Least squares is a method that allows us to approximate the solution to a linear system by finding an x that makes the distance between Ax and b as small as possible. In Section 2.1 we introduced the formula to find the distance between a vector and a subspace. Here we apply this idea to find the vector $\hat{x}$ in the column space of A ($Col\, A$) that is closest to b. Here, the least squares approximation x minimizes the squared distance between Ax and b.

$$Min_{\hat{x}} \|A\hat{x} - b\|^2$$

In the context of linear regression, the least squares approximation vector x represents the regression coefficients β, $Col\, A$ represents all the possible fitted values for the model, and $A\hat{x}$ is the orthogonal projection of b onto $Col\, A$.

To find x we can take the derivative with respect to x and set the problem equal to zero.

$$\|Ax - b\|^2 = \ Ax - b\ ^T\ Ax - b\ = f(x)$$

$$\frac{\partial f}{\partial \hat{x}} = 2A^T(A\hat{x} - b) = 0$$

This simplifies yielding the normal equations:

$$A^T A\hat{x} = A^T b$$

The normal equations are the result of the least squares problem; we can rearrange the normal equations to find a solution for the coefficient vector x that minimizes the sum of squared residuals.

$$\hat{x} = (A^T A)^{-1} A^T b$$

Let's illustrate this process with a linear regression example.

Consider:

$$b = \begin{bmatrix} 2.3 \\ -0.6 \\ 11.7 \\ 0.1 \\ 3.2 \\ 1.3 \end{bmatrix}, A = \begin{bmatrix} 1 & 1 & 2 \\ 1 & 2 & 1 \\ 1 & 3 & 6 \\ 1 & 4 & 3 \\ 1 & 5 & 4 \\ 1 & 6 & 5 \end{bmatrix}$$

Where:

- b is the observation data
- A is the prediction data with two independent variables
- Each observation for the system can be expressed: $b_i = \beta_0 a_{i0} + \beta_1 a_{i1} + \beta_2 a_{i2}$

Next, we must compute the components of the normal equation and solve for x.

$$A^T A = \begin{bmatrix} 6 & 21 & 21 \\ 21 & 91 & 78 \\ 21 & 78 & 90 \end{bmatrix}$$

$$A^T b = \begin{bmatrix} 18 \\ 67.2 \\ 64.8 \end{bmatrix}$$

$$(A^T A)^{-1} = \begin{bmatrix} 1.317 & -0.474 & -0.242 \\ -0.474 & 0.413 & -0.151 \\ -0.242 & -0.151 & 0.163 \end{bmatrix}$$

$$\hat{x} = (A^T A)^{-1} A^T b = \begin{bmatrix} 1.317 & -0.474 & -0.242 \\ -0.474 & 0.413 & -0.151 \\ -0.242 & -0.151 & 0.163 \end{bmatrix} \begin{bmatrix} 18 \\ 67.2 \\ 64.8 \end{bmatrix}$$

$$= \begin{bmatrix} -0.525 \\ -1.882 \\ 2.889 \end{bmatrix}$$

Now that we've solved the coefficients of our model, we can interpret them. The first coefficient $\beta_0 = -0.525$ is a constant in the model and is the predicted value when a_1 and a_2 equal zero. The second coefficient β_1 indicates that the predicted value will decrease by 1.882 for every single unit increase in a_1. Similarly, the third coefficient a_2 indicates that the predicted value will increase by 2.889 for every single unit increase in a_2. All together, these coefficients determine the best fitted regression plane for the data.

Figure 2.3

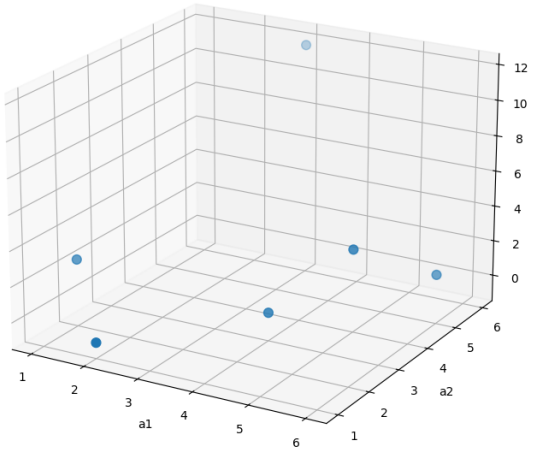


Figure 2.4

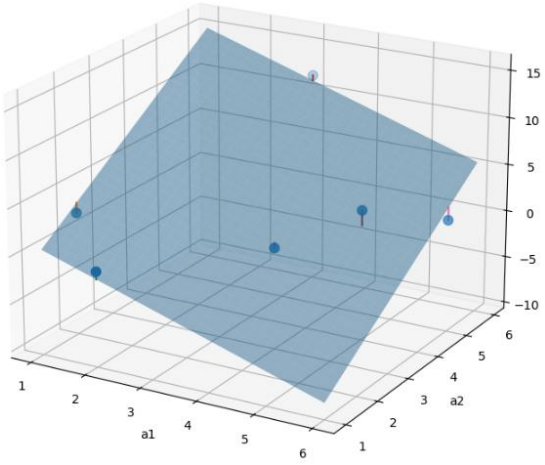


Figure 2.3 shows the plotted example data. Figure 2.4 demonstrates the linear regression plane created with the calculated coefficient vector $\boldsymbol{x}$

2.4 Extension of Linear Regression: Ridge Regression

As mentioned in Section 2.2, the standard linear regression model assumes low multicollinearity (low correlation) between predictors. In cases where predictors have high multicollinearity, a standard regression may produce unstable coefficient estimates that overfit the model. Ridge regression is a technique used to regularize coefficient estimations and penalize large coefficient values. Ridge regression modifies the standard linear regression optimization problem by adding the penalty term $\lambda\|\beta\|^2$. The addition of the penalty term increases the minimized residual sum of squares, introducing a small amount of bias for a large decrease in variance.

$$Min_{\beta}\ \|y - \beta X\|^2 + \lambda\|\beta\|^2 = \hat{\beta}$$

We can see that higher coefficients will be more impacted by this penalty term due to the squaring of $\|\beta\|$. $\lambda\|\beta\|^2$ expands to $\lambda \times (\beta_0^2 + \beta_1^2 + \cdots + \beta_n^2)$, small coefficients like $\beta = 0.10$ have a penalty term of $\lambda \times 0.01$, large coefficients will have a much higher penalty, for example $\beta = 4$ becomes $\lambda \times 16$. Additionally, the choice of λ determines the impact of the penalty term, a λ of zero results in the standard linear regression model, a high choice of λ results in coefficients shrinking towards zero. The solution for the minimization problem is:

$$\hat{\beta}_{ridge} = (X^T X - \lambda I) X^T y$$

Ridge regression is a valuable technique because it cleanly addresses the standard linear regression assumption of low multicollinearity. By simply adding a penalty term to the minimization problem, the model's bias-variance tradeoff can be adjusted as needed.

References

Axehedge. (2023). *The Black–Litterman model in stock trading: The portfolio optimizer.* Medium. https://medium.com/@axehedge/the-black-litterman-model-in-stock-trading-the-portfolio-optimizer-6e6301fe40f6

CliffsNotes. (n.d.). *Projection onto a subspace.* https://www.cliffsnotes.com/study-guides/algebra/linear-algebra/real-euclidean-vector-spaces/projection-onto-a-subspace

Columbia University Mailman School of Public Health. (n.d.). *Ridge regression.* https://www.publichealth.columbia.edu/research/population-health-methods/ridge-regression

Financial Edge. (n.d.). *The Black–Litterman model.* https://www.fe.training/free-resources/portfolio-management/black-litterman-model/

GeeksforGeeks. (n.d.). *Assumptions of linear regression.* https://www.geeksforgeeks.org/machine-learning/assumptions-of-linear-regression/

GuidedChoice. (n.d.). *Modern portfolio theory.* https://guidedchoice.com/modern-portfolio-theory/

Hoerl, A. E., & Kennard, R. W. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics, 12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

IBM. (n.d.). *Ridge regression.* https://www.ibm.com/think/topics/ridge-regression

Investopedia. (n.d.). *Capital asset pricing model (CAPM).* https://www.investopedia.com/terms/c/capm.asp

Investopedia. (n.d.). *Capital market line (CML).* https://www.investopedia.com/terms/c/cml.asp

Investopedia. (n.d.). *What's the difference between linear regression and multiple regression?*

https://www.investopedia.com/ask/answers/060315/what-difference-between-linear-regression-and-multiple-regression.asp

Lay, D. C., Lay, S. R., & McDonald, J. J. (2021). *Linear algebra and its applications* (6th ed., ch. 6). Pearson.

MachineLearningNoJutsu. (2022). *Linear regression: Ordinary least squares.* Medium. https://medium.com/@machinelearningnojutsu/linear-regression-ordinary-least-square-cabe0d97117

Markowitz, H. M. (1952). Portfolio selection. *The Journal of Finance, 7*(1), 77–91. https://www.math.hkust.edu.hk/~maykwok/courses/ma362/07F/markowitz_JF.pdf

MathWorks. (n.d.). *What is linear regression?* https://www.mathworks.com/help/stats/what-is-linear-regression.html

Massachusetts Institute of Technology. (2011). *Projections and least squares (Lecture notes).* MIT OpenCourseWare. https://ocw.mit.edu/courses/18-06sc-linear-algebra-fall-2011/00e9c8f0eafedeab21a3d079a17ed3d8_MIT18_06SCF11_Ses2.2sum.pdf

Icaza, E. (2025). *ASRM 406 lecture slides.* University of Illinois at Urbana–Champaign.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *The Journal of Finance, 19*(3), 425–442. https://doi.org/10.2307/2977928

StableBread. (n.d.). *Capital asset pricing model explained.* https://stablebread.com/capital-asset-pricing-model/

Städler, N. (2022). *Ridge regression and shrinkage.* In *High-dimensional statistics* (Bookdown). https://bookdown.org/staedler_n/highdimstats/ridge-regression-and-shrinkage.html

Vandenberghe, L. (2013). *Least squares (Lecture notes).* University of California, Los Angeles. http://www.seas.ucla.edu/~vandenbe/133A/lectures/ls.pdf

Verum Partnership. (n.d.). *Understanding the efficient frontier in portfolio construction.* https://verumpartnership.com/understanding-the-efficient-frontier-in-portfolio-construction/

Zivot, E. (2006). *Portfolio theory with matrix algebra (Lecture notes).* University of Washington. https://faculty.washington.edu/ezivot/econ424/portfolioTheoryMatrix.pdf