# Sharpe Ratio Optimization Via Gradient Descent
## Theory, Proof, & Implementation

Joe Nunez

Harvey Mudd College

May 2, 2018

# Portfolio

A portfolio can be regarded as a set of assets $a_i$, each with proportion $\pi_i$ between 0 and 1 corresponding to the portion of the portfolio allocated to that asset, and where the sum of all $\pi_i$ is 1. There is no upper bound on the number of assets theoretically in a portfolio.

Immediately, we can see that the set of possible portfolios is a subset of the Statistic Manifold, $S$, in that the set is infinite-dimensional and has elements summing to exactly 1. This will allow us to apply methods from information geometry to the analysis of portfolios.

# Simplex

The Statistic Manifold $S$ is the set of all probability distributions. When $S$ is restricted to $n$ dimensions, it creates an $(n-1)$-dimensional hyperplane based on the restriction that the elements sum to 1.

$$\{\pi \in \mathbb{R}^n \mid \pi_1 + \pi_2 + \cdots + \pi_n = 1\}$$

# Basic Portfolio Theory

Consider the returns $R_i$ of assets to be random variables (they have been empirically found to resemble a log normal distribution. The risks $\sigma_i$ of assets can be represented as the standard deviation in the returns, and the rates of return $r_i$ can be represented with mean returns (typically found using daily log continuous growth rates for the past year or two). (Wong)

# Portfolio Returns

For a portfolio of $n$ assets with portfolio weights $\pi_1, \pi_2, \ldots, \pi_n$ and return rates $\mu_1, \mu_2, \ldots, \mu_n$, the return of the portfolio overall is

$$\mu_P = \pi_1\mu_1 + \pi_2\mu_2 + \cdots + \pi_n\mu_n = \sum_{i=1}^{n} \pi_i\mu_i = \pi \cdot \mu$$

(Wong)

## Portfolio Risk

Financial assets often have correlated performance, so it is important to account for this when estimating the risk level in a portfolio. The variance of the returns for a portfolio with non-independent assets is calculated as

$$\sigma_P^2 = \sum_{i,j=1}^{n} \pi_i \pi_j \text{Cov}(R_i, R_j),$$

where $\text{Cov}(R_i, R_j)$ is the covariance of the returns for assets $i$ and $j$ (Wong). Note that if $C$ is the covariance matrix of the returns, then the formula becomes

$$\sigma_P^2 = \pi \cdot C\pi.$$

# Sharpe Surface

Define the Sharpe Surface as the range of the function $s : V \to \mathbb{R}$, where $V$ is the range set of all possible portfolio configurations, and $s$ is the function that calculates the Sharpe ratio of a portfolio.

For a portfolio vector $\pi \in V$, the Sharpe ratio associated with that vector is

$$s(\pi) = \frac{\pi \cdot \mu}{\sqrt{\pi \cdot C\pi}}$$

# Sharpe Surface

We'd like to know the gradient of this surface at a given point. To obtain this, we take the partial derivative with respect to $\pi_i$

$$\frac{\partial s}{\partial \pi_i} = \frac{\pi_i \mu_i \sqrt{\pi \cdot C\pi} - \frac{\pi \cdot \mu}{2\sqrt{\pi \cdot C\pi}}(\pi' \cdot C\pi + \pi \cdot C\pi')}{\pi \cdot C\pi}$$

where $\pi'$ is the derivative of $\pi$ with respect to $\pi_i$, i.e., a one-hot vector.

# Dimensionality of $V$

For a portfolio with $n$ assets, there are only $n - 1$ degrees of freedom, so we want the space of vectors which $\pi$ can come from to be $n - 1$ dimensions instead of $n$. To do so, consider the diffeomorphism $\phi : U \to V$

$$\phi(\pi_1, \pi_2, \ldots, \pi_{n-1}) \mapsto (\pi_1, \pi_2, \ldots, \pi_{n-1}, 1 - \pi_1 - \pi_2 - \cdots - \pi_{n-1})$$

Note that the linearity of this map combined with the restriction that $\sum_{i=1}^{n} \pi_i = 1$ makes the map invertible, and the linearity makes the map and its inverse differentiable.

# Sharpe Surface Parametrization

We now parametrize the Sharpe Surface by $\psi = s \circ \phi$. For $\pi \in U$, we then get

$$\psi(\pi) = \frac{\phi(\pi) \cdot \mu}{\sqrt{\phi(\pi) \cdot C\phi(\pi)}}$$

as the new parametrization of our surface.

# Sharpe Surface Parametrization

To account for the mapping from $U$ to $V$, we must establish a connection $\nabla : (U, V) \to \nabla_U V$. This connection gets associated with a matrix

$$D\phi = \begin{pmatrix} \frac{\partial \phi_1}{\partial \pi_1} & \frac{\partial \phi_1}{\partial \pi_2} & \cdots & \frac{\partial \phi_1}{\partial \pi_{n-1}} \\ \frac{\partial \phi_2}{\partial \pi_1} & \frac{\partial \phi_2}{\partial \pi_2} & \cdots & \frac{\partial \phi_2}{\partial \pi_{n-1}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \phi_n}{\partial \pi_1} & \frac{\partial \phi_n}{\partial \pi_2} & \cdots & \frac{\partial \phi_n}{\partial \pi_{n-1}} \end{pmatrix}$$

## Sharpe Surface Parametrization

Notice that for $\phi$, these values become

$$D\phi = \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \\ -1 & -1 & \ldots & -1 \end{pmatrix}$$

Then the gradient in terms of a weight vector from $U$ becomes

$$\nabla\psi = D\phi\left[\frac{\pi_i\mu_i\sqrt{\phi(\pi)\cdot C\phi(\pi)} - \frac{\phi(\pi)\cdot\mu}{2\sqrt{\phi(\pi)\cdot C\phi(\pi)}}(\pi' \cdot C\phi(\pi) + \phi(\pi) \cdot C\pi')}{\phi(\pi) \cdot C\phi(\pi)}\right]$$

where the [] denote the vector formed by all values $1 \leq i \leq n-1$.

# Convexity

Let $f : X \to \mathbb{R}$ and $x_1, x_2 \in X$. Let $[x_1, x_2]$ be the line segment from $x_1$ to $x_2$ and $g(x)$ be the line segment from $f(x_1)$ to $f(x_2)$ such that $g(x_1) = f(x_1)$ and $g(x_2) = f(x_2)$. A function $f$ is convex if for any $x \in [x_1, x_2]$, $g(x) \geq f(x)$.

Notice that if the opposite is true—$g(x) \leq f(x)$ for any $x \in [x_1, x_2]$—then $f(x)$ can be made convex by letting $h(x) = -f(x)$.

# Convexity

Let $f : X \to \mathbb{R}$ and $x_1, x_2 \in X$. Let $[x_1, x_2]$ be the line segment from $x_1$ to $x_2$ and $g(x)$ be the line segment from $f(x_1)$ to $f(x_2)$ such that $g(x_1) = f(x_1)$ and $g(x_2) = f(x_2)$. A function $f$ is convex if for any $x \in [x_1, x_2]$, $g(x) \geq f(x)$.

Notice that if the opposite is true—$g(x) \leq f(x)$ for any $x \in [x_1, x_2]$—then $f(x)$ can be made convex by letting $h(x) = -f(x)$. For our purposes, we will use this second definition.

# Convexity of Sharpe Surface

Consider two portfolios $\pi_a$ and $\pi_b$. Let $\alpha \in [0,1]$ and let $\pi_c = \alpha\pi_a + (1-\alpha)\pi_b$. Note that since $\phi$ is a linear function, $\phi(\alpha\pi_a + (1-\alpha)\pi_b) = \alpha\phi(\pi_a) + (1-\alpha)\phi(\pi_b)$. Now we'll calculate the Sharpe ratio for $\phi(pi_c)$. For simplicity of notation, we'll just use $\pi$ instead of $\phi(\pi)$ for $\pi = \pi_a, \pi_b, \pi_c$.

# Convexity of Sharpe Surface

$$s(\pi_c) = \frac{(\alpha\pi_a + (1-\alpha)\pi_b) \cdot \mu}{\sqrt{(\alpha\pi_a + (1-\alpha)\pi_b) \cdot C(\alpha\pi_a + (1-\alpha)\pi_b)}}$$

To show convexity, we wish to show that this value is greater than or equal to

$$\frac{\alpha\pi_a \cdot \mu}{\sqrt{\pi_a \cdot C\pi_a}} + \frac{(1-\alpha)\pi_b \cdot \mu}{\sqrt{\pi_b \cdot C\pi_b}}$$

# Convexity of Sharpe Surface

Since the terms in the square root represent norms under the Mahalanobis distance, they obey the triangle inequality. Therefore the denominator of the Sharpe ratio is less than or equal to

$$\sqrt{\alpha \pi_a \cdot C\alpha\pi_a} + \sqrt{(1-\alpha)\pi_b \cdot C(1-\alpha)\pi_b} = \alpha\sqrt{\pi_a \cdot C\pi_a} + (1-\alpha)\sqrt{\pi_b \cdot C\pi_b}$$

Therefore

$$s(\pi_c) \geq \frac{\alpha \pi_a \cdot \mu + (1-\alpha)\pi_b \cdot \mu}{\alpha\sqrt{\pi_a \cdot C\pi_a} + (1-\alpha)\sqrt{\pi_b \cdot C\pi_b}}$$

## Convexity of Sharpe Surface

We can get common denominators by multiplying by convenient ones, then ignore the common denominators, leaving

$$(\alpha \pi_a \cdot \mu + (1-\alpha)\pi_b \cdot \mu)\sqrt{\pi_a \cdot C\pi_a}\sqrt{\pi_b \cdot C\pi_b}$$

for the Sharpe ratio and

$$(\alpha \pi_a \cdot \mu \sqrt{\pi_b \cdot C\pi_b} + (1-\alpha)\pi_b \cdot \mu \sqrt{\pi_a \cdot C\pi_a})$$

$$(\alpha\sqrt{\pi_a \cdot C\pi_a} + (1-\alpha)\sqrt{\pi_b \cdot C\pi_b})$$

for the point on the line segment.

## Convexity of Sharpe Surface

The point on the line segment becomes

$$\alpha \pi_a \cdot \mu [\alpha \sqrt{\pi_b \cdot C \pi_b} \sqrt{\pi_a \cdot C \pi_a} + (1 - \alpha) \pi_b \cdot C \pi_b] +$$

$$(1 - \alpha) \pi_b \cdot \mu [(1 - \alpha) \sqrt{\pi_b \cdot C \pi_b} \sqrt{\pi_a \cdot C \pi_a} + \alpha \pi_a \cdot C \pi_a]$$

To obtain the desired inequality, we need

$$\alpha(1 - \alpha)[(\pi_b \cdot \mu)\pi_a \cdot C \pi_a + (\pi_a \cdot \mu)\pi_b \cdot C \pi_b] \leq$$

$$\alpha(1 - \alpha)(\pi_a \cdot \mu + \pi_b \cdot \mu)\sqrt{\pi_b \cdot C \pi_b} \sqrt{\pi_a \cdot C \pi_a}$$

# Convexity of Sharpe Surface

This simplifies to

$$(\pi_b \cdot \mu)\pi_a \cdot C\pi_a + (\pi_a \cdot \mu)\pi_b \cdot C\pi_b \leq$$

$$(\pi_a \cdot \mu + \pi_b \cdot \mu)\sqrt{\pi_b \cdot C\pi_b}\sqrt{\pi_a \cdot C\pi_a}$$

Without loss of generality, let $\pi_b \cdot \mu > \pi_a \cdot \mu$. Then this inequality holds if $\sqrt{\pi_b \cdot C\pi_b} > \sqrt{\pi_a \cdot C\pi_a}$, i.e. if portfolios with higher returns are riskier.

# Capital Asset Pricing Model

While the previous assumption may seem overly broad, it actually matches the core idea of the Capital Asset Pricing Model (CAPM) developed by Sharpe, Markowitz, and Miller (for which they won the 1990 Nobel prize in economics). The CAPM dictates that an asset should only be added to a portfolio if its expected return varies with its the asset's variance according to a market determined ratio.

The CAPM is widely used, so it is not unreasonable to assume that most stocks obey this rule. Therefore, as long as the assets we consider obey the efficient market hypothesis, the Sharpe surface is convex.

# Gradient Descent

Gradient descent, the process of finding a local minima through taking small steps along the gradient of a surface, is guaranteed to find a global a minima for convex surface.

Since the space of weight vectors is not fully $n$-dimensional, but instead $n - 1$ dimensions, we need to make sure that we stay on the surface while making steps along the gradient. To do so, after each step along the gradient, we need to return to the closest point on the surface. To do so, we need to make an orthogonal projection back into the simplex.

# Orthogonal Projection

For the hyperplane defined by

$$\{\pi \in \mathbb{R}^n \mid \pi_1 + \pi_2 + \cdots + \pi_n = 1\},$$

the set of normal vectors to this hyperplane is

$$\{c \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} \mid c \in \mathbb{R}\}$$

So to project back into the simplex, we let $c = \frac{1}{n} \sum_{i=1}^{n} \pi_i$

# Sources

Wong, T. 2016, 'Geometry and Optimization of Relative Arbitrage', PhD Thesis, University of Michigan, Ann Arbor MI.
http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=125867
https://stats.stackexchange.com/questions/7440/
kl-divergence-between-two-univariate-gaussians
https://www.pyimagesearch.com/2016/10/17/
stochastic-gradient-descent-sgd-with-python/
https://en.wikipedia.org/wiki/Capital_asset_pricing_model
https://docs.scipy.org/doc/numpy/reference/
**Data**
api.iextrading.com
**Images**:
https://en.wikipedia.org/wiki/Simplex
https://commons.wikimedia.org/wiki/File:
Error_surface_of_a_linear_neuron_with_two_input_weights.png