The Geometry of Risk and Reward

An application of Euclidean linear algebra

by

Victor T. Norton, Jr.
Department of Mathematics and Statistics
Bowling Green State University

Abstract

We adopt a geometric view of Markowitz's and Sharpe's mean-variance theory of portfolio choice. Our model posits that expected reward is a linear function of risk. This axiom is generally true after singular value reduction of data. The Sharpe-optimal long portfolio leads to an investment strategy that appears to have considerable merit.

Table of Contents

1. Return and reward

This is not a theoretical paper. Our techniques and results are data driven. We will demonstrate our ideas with quarterly total returns from seven Vanguard mutual funds over the nine-year period 1993–2001. These data are displayed in Table 1 (Appendix A). They were obtained online from

Vanguard Fund Finder <http://majestic5.vanguard.com/FP/DA/>

The funds are

Vanguard 500 Index (VFINX) – This fund tracks the performance of the Standard & Poor's 500 Index, which is dominated by the stocks of large U.S. companies.

Vanguard Growth Index (VIGRX) – This fund tracks the performance of the Standard & Poor's 500/BARRA Growth Index, which includes those stocks of the S&P 500 Index with higher-than-average price/book ratios.

Vanguard Value Index (VIVAX) – This fund tracks the performance of the Standard & Poor's 500/BARRA Value Index, which includes those stocks of the S&P 500 Index with lower-than-average price/book ratios.

Vanguard Total Stock Market Index (VTSMX) – This fund tracks the performance of the Wilshire 5000 Total Market Index, which consists of all the U.S. common stocks regularly traded on the New York and American Exchanges and the Nasdaq over-the-counter market.

Vanguard Extended Market Index (VEXMX) – This fund tracks the performance of the Wilshire 4500 Completion Index, which consists of all stocks in the Wilshire 5000 Total Market Index, except those included in the S&P 500 Index.

Vanguard Total Bond Market Index (VBMFX) – This fund seeks to track the performance of the Lehman Brothers Aggregate Bond Index, which measures the total universe of public investment-grade fixed income securities in the United States with maturity of over 1 year.

Vanguard Prime Money Market (VMMXX) – This fund invests in high-quality, short-term money market instruments—average maturity 90 days or less. It seeks to maintain a stable share price of $1.

We think of the first six funds as "risky" investments, while the money market fund is "risk free." Perhaps the periodic returns of a "risk-free" asset should have zero variance. This is not true of our money market fund. However the variance of its returns is essentially zero—relative to the variances of the six "risky" funds.

Reward

Let $X(q)$ denote the return of a risky fund in quarter q and $X_0(q)$ denote the risk free return in the same quarter. We define the *reward* of the risky fund in quarter q to be the differential return

$$r(q) = X(q) - X_0(q).$$

This is the reward for taking risk. If you put all your money in a good money market fund, you are taking no risk, and, by our definition, you earn no reward.

Summary statistics

Here are summary statistics for the above funds. The statistics cover the nine-year period, 1993–2001. They are annualizations of quarterly statistics. (Annualized means are 4 times quarterly means; annualized standard deviations are 2 times quarterly standard deviations.)

Return statistics	VFINX	VIGRX	VIVAX	VTSMX	VEXMX	VBMFX	VMMXX
mean return	13.97	14.14	13.68	13.29	12.96	7.05	4.85
stdv of return	15.46	18.72	14.40	16.65	22.54	4.15	0.53
wrong ratio	0.904	0.756	0.950	0.798	0.575	1.697	9.116
Reward statistics	VFINX	VIGRX	VIVAX	VTSMX	VEXMX	VBMFX	VMMXX
mean reward	9.12	9.29	8.82	8.44	8.11	2.20	0.00
stdv of reward	15.45	18.72	14.36	16.65	22.59	4.10	0.00
Sharpe ratio	0.590	0.496	0.614	0.507	0.359	0.536	0/0

The mean rewards are the corresponding mean returns less the money market mean return. (Deviations from this relationship are due to rounding.) If the "risk free" returns had zero variance, the standard deviations of corresponding "risky" returns and rewards would be equal. This is essentially true due to the negligible variance of money market returns.

The Sharpe ratio (Sharpe [1994]) is the ratio of mean reward to standard deviation of reward—the ratio of reward to risk. It provides a measure of the performance of a mutual fund, the higher the Sharpe ratio the better. The Sharpe ratio of the risk free fund is undefined. For comparison we have included the corresponding "wrong ratio," which uses returns rather than rewards.

2. Weighty Matters

Given a sample of m successive periodic rewards $r = (r_1, \ldots, r_m)$ from a single fund, with r_1 the latest reward, we posit that the expected reward for the next period should be given by an equation of form

$$\bar{r}_P = E_P(r) = \mu_1 r_1 + \ldots + \mu_m r_m,$$

where the weights μ_i $(i = 1, \ldots, m)$ are positive and sum to 1. What are the appropriate weights? This is the big question.

In much of the literature, e.g., Sharpe [1994], these *ex post* weights are of equal size and the expected reward is the arithmetic mean of the sample rewards. This is all well and good if one is primarily interested in measuring historical performance. However, the arithmetic average leaves much to be desired if one actually wants to estimate next quarter's reward from the last few years of data. For this purpose the performance over the past year should bear more weight than the performance of two or three years ago.

Our weight systems

We have experimented with weight systems constructed from the one parameter family of functions

$$y = K_w(1 - x^{1/w}) \quad (0 \le x < 1), \quad w \ge 0,$$

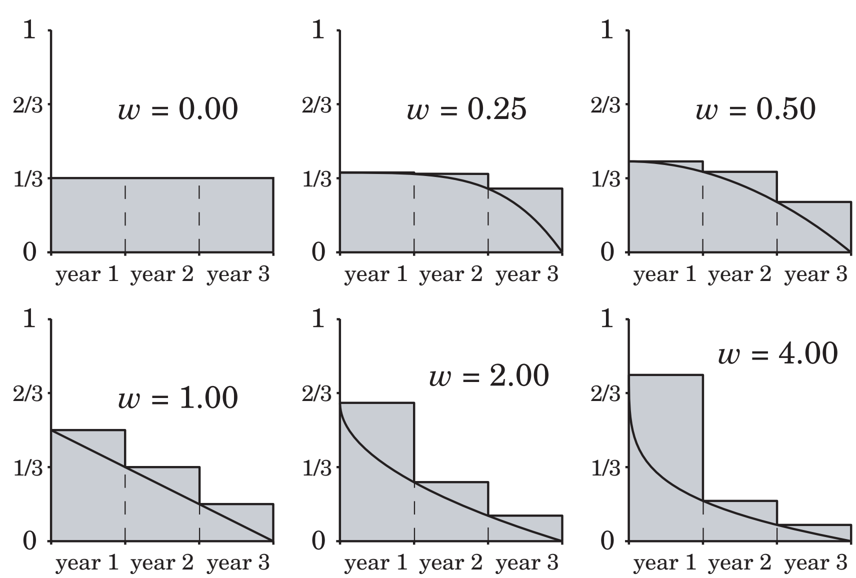
(Here we adopt the convention that $x^{1/0} = 0$ for $0 \le x < 1$).

For samples of n_Y years duration, the constants K_w are chosen so that the $y_k = y_k(w)$ corresponding to $x_k = (k-1)/n_Y,\ k = 1, \ldots, n_Y,$ sum to 1. The final weight system is defined by

$$\mu_i = y_k/n_P, \quad k = 1 + \left\lfloor \frac{i-1}{n_P} \right\rfloor, \quad i = 1, \ldots, m = n_P n_Y,$$

where n_P is the number of periods per year and m the number of periods in total. (The lower brackets denote the greatest integer function [floor].)

Weight Systems



The parameter value $w = 0.0$ produces the uniform weight system $\mu_i = 1/m$. In many of our examples we use the system corresponding to $w = 2.0$ with $n_P = 4$ and $n_Y = 3$ (three years of quarters). Here

$$\mu_{1:4} = 0.1557 \text{ with sum } y_1 = 0.623,$$
$$\mu_{5:8} = 0.0658 \text{ with sum } y_2 = 0.263,$$

$$\mu_{9:12} = 0.0286 \text{ with sum } y_3 = 0.114.$$

Annualized statistics

Throughout this paper we will work with annualized statistics. The (annualized) *expected reward*,

$$\bar{r} = E(r) = n_P \times \{\mu_1 r_1 + \ldots + \mu_m r_m\},$$

corresponds to the sum of n_P periodic rewards. The (annualized) *variance*,

$$v = V(r) = n_P \times \sum_{i=1}^{m} \mu_i \left(r_i - \bar{r}_P\right)^2,$$

derives from the assumption that the periodic rewards are independent and identically distributed.

The (annualized) *Sharpe ratio* (Sharpe [1994]) plays a crucial role in this paper. Roughly speaking it is the ratio of reward to risk. The formal definition is

$$s = S(r) = \frac{E(r)}{\sqrt{V(r)}}.$$

(When weighting is uniform [$w = 0$], our Sharpe ratio differs from the standard one by a factor of $\sqrt{m/(m-1)}$, because $m-1$ rather than m is the divisor in the definition of the standard deviation of r.)

3. Orthogonalization

From now on we assume that a sample size $m = n_P n_Y$ and a system of weights μ_i ($i = 1, \ldots, m$) has been chosen.

Let $r = (r_1, \ldots, r_m)$ be a sample of successive rewards from a single fund with r_1 the latest reward. We wish to view r geometrically, as an element of a Euclidean "risk–reward" space. For this purpose, define the unit vector $\mathbf{e}_R \in \mathbf{R}^m$ by

$$\mathbf{e}_R = [\sqrt{\mu_1}, \ldots, \sqrt{\mu_m}]^T$$

("unit" since $\|\mathbf{e}_R\|^2 = \mathbf{e}_R^T \mathbf{e}_R = \mu_1 + \ldots + \mu_m = 1$), and define the *reward vector* $\mathbf{r}$ corresponding to r by

$$\mathbf{r} = \left[r_1 \sqrt{n_P\, \mu_1}, \ldots, r_m \sqrt{n_P\, \mu_m}\right]^T.$$

Then

$$\bar{r} = E(\mathbf{r}) = \left(\mathbf{e}_R \sqrt{n_P}\right)^T \mathbf{r} = n_P \times \{\mu_1 r_1 + \ldots + \mu_m r_m\}$$

is the (annualized) expected reward of the last section.

Think of $\mathbf{e}_R$ as defining an "expected reward" axis in "risk–reward" space. From this point of view the expected reward $\bar{r}$ of $\mathbf{r}$ is just its expected reward coordinate, $\mathbf{e}_R^T \mathbf{r}$, scaled by the annualization factor $\sqrt{n_P}$. (If we weren't annualizing, there wouldn't be any n_P's at all.)

Risk

Risk is orthogonal to reward. The projection of $\mathbf{r}$ perpendicular to $\mathbf{e}_R$ is given by

$$\mathbf{f} = \mathbf{r} - \mathbf{e}_R \left(\mathbf{e}_R^T \mathbf{r} \right).$$

We will refer to $\mathbf{f}$ as the *risk vector*, or sometimes just the *risk*. corresponding to $\mathbf{r}$. The square length of $\mathbf{f}$ is the (annualized) variance of r,

$$v = V(\mathbf{r}) = \|\mathbf{f}\|^2 = n_P \times \sum_{i=1}^{m} \mu_i \left(r_i - \bar{r}_P \right)^2,$$

as defined in the last section.

The equation

$$\mathbf{r} = \mathbf{f} + \mathbf{e}_R \left(\frac{\bar{r}}{\sqrt{n_P}} \right)$$

expresses $\mathbf{r}$ as an orthogonal sum of risk and expected reward. The (annualized) Sharpe ratio can be expressed in terms of these components as

$$s = S(\mathbf{r}) = \frac{E(\mathbf{r})}{\sqrt{V(\mathbf{r})}} = \frac{\bar{r}}{\|\mathbf{f}\|}.$$

Risk and reward: sources of confusion

We use the words "risk" and "reward" to mean different things depending on the context. For example, we think of the Sharpe ratio, $s = \bar{r}/\|\mathbf{f}\|$, as the ratio of "reward" to "risk." Technically, the numerator is "expected reward", and the denominator should be called "scalar risk" to distinguish it from the multidimensional "risk" vector $\mathbf{f}$.

Most of the time "reward" will mean expected reward, as in this example. "Risk" is a bit more confusing. Sometimes it will mean scalar risk, $\|\mathbf{f}\|$, but, from our point of view, risk is fundamentally a multidimensional quantity. "Growth versus value" may be one dimension of risk, "large cap versus small cap" another, and certainly "bond versus stock" is a very fundamental dimension of risk.

Be this as it may, we will continue to use the words "risk" and "reward" in a variety of ways, hoping their meaning will be clear from the context.

4. Portfolios

We continue to assume that a sample size $m = n_P n_Y$ and a system of weights μ_i ($i = 1, \ldots, m$) has been chosen. Then the choice of m successive periods and n risky funds gives rise to n reward vectors $\mathbf{r}_1, \ldots, \mathbf{r}_n$ in $\mathbf{R}^m$, with corresponding expected rewards $\bar{r}_1, \ldots, \bar{r}_n$, and risks $\mathbf{f}_1, \ldots, \mathbf{f}_n$, that satisfy the equations

$$\mathbf{r}_i = \mathbf{f}_i + \mathbf{e}_R \left(\frac{\bar{r}_i}{\sqrt{n_P}} \right), \qquad i = 1, \ldots, n.$$

Portfolios of risky funds

We are interested in portfolios of risky funds. Given n funds as above, a *portfolio* of these funds is determined by the sequence, $p_1, \ldots, p_n$, of the proportions of each fund in the portfolio. The proportions must sum to one: $\sum_{i=1}^{n} p_i = 1$. This is the only requirement. However, we will be particularly concerned with *long* portfolios—portfolios for which all of the p_i are nonnegative.

The reward vector, $\mathbf{r}$, expected reward, $\bar{r}$, and risk, $\mathbf{f}$, of a portfolio of the risky funds are made up of the same proportions of the reward vectors, expected rewards, and risks of the component funds:

$$\mathbf{r} = \sum_{i=1}^{n} \mathbf{r}_i p_i, \quad \bar{r} = \sum_{i=1}^{n} \bar{r}_i p_i, \quad \mathbf{f} = \sum_{i=1}^{n} \mathbf{f}_i p_i.$$

The variance of the rewards of the portfolio is a quadratic function of these proportions:

$$v = \|\mathbf{f}\|^2 = \mathbf{f}^T \mathbf{f} = \sum_{i=1}^{n} \sum_{j=1}^{n} p_i \left(\mathbf{f}_i^T \mathbf{f}_j \right) p_j.$$

The portfolio flat

The *portfolio flat* (in risk space) plays a crucial role in this paper. It is the affine subspace $\mathcal{P}$ of $\mathbf{R}^m$ consisting of all risk vectors from portfolios of the risky funds under consideration. Thus

$$\mathcal{P} = \left\{ \mathbf{f} \in \mathbf{R}^m : \mathbf{f} = \sum_{i=1}^{n} \mathbf{f}_i p_i, \ \sum_{i=1}^{n} p_i = 1 \right\}.$$

The tangent space of the portfolio flat is the linear subspace $\mathrm{T}(\mathcal{P})$ of $\mathbf{R}^m$ defined by

$$\mathrm{T}(\mathcal{P}) = \left\{ \mathbf{v} \in \mathbf{R}^m : \mathbf{v} = \sum_{i=1}^{n} \mathbf{f}_i t_i, \ \sum_{i=1}^{n} t_i = 0 \right\}.$$

The difference of any two elements of $\mathcal{P}$ is an element of $\mathrm{T}(\mathcal{P})$. Consequently $\mathcal{P} = \mathbf{f} + \mathrm{T}(\mathcal{P})$ for any $\mathbf{f} \in \mathcal{P}$.

Examples

Suppose we want to invest \$100,000 in a portfolio of risky funds. We take \$100,000 out of our money market account, and put \$25,000 in Fund#1 and \$75,000 in Fund#2. This is what Sharpe [1994] calls a zero-investment strategy. We have simply transferred money from one type of investment to another. Hopefully we will be rewarded for the transfer. Hopefully the risky funds will return more than money market.

Our risky portfolio consists of $p_1 = 25\%$ of Fund#1 and $p_2 = 75\%$ of Fund#2. It is a long portfolio. Its risk vector $\mathbf{f}$ is an element of the portfolio flat $\mathcal{P}$: $\mathbf{f} = \mathbf{f}_1 p_1 + \mathbf{f}_2 p_2$.

In this example we took safe money and invested it in two risky funds in the hope that each would earn more than money market. Here is a different possibility.

Fund#2 looks good, but we are sure that the price of Fund#1 is going to drop significantly. So we take \$100,000 out of our money market fund, sell \$50,000 of Fund#1 short, and put \$150,000 in Fund#2. Another zero-investment strategy. This time our risky portfolio consists of $p_1 = -50\%$ of Fund#1 and $p_2 = 150\%$ of Fund#2. Again $p_1 + p_2 = 1$, but we are not working with a long portfolio anymore.

5. Components of risk (nonsingular case)

Continue to assume that a sample size $m = n_P n_Y$, a system of weights, m successive periods, and n risky funds have been chosen. Then the corresponding reward vectors $\mathbf{r}_1, \ldots, \mathbf{r}_n$ in $\mathbf{R}^m$, with expected rewards $\bar{r}_1, \ldots, \bar{r}_n$ and risks $\mathbf{f}_1, \ldots, \mathbf{f}_n$, satisfy the equations

$$\mathbf{r}_i = \mathbf{f}_i + \mathbf{e}_R \left(\frac{\bar{r}_i}{\sqrt{n_P}} \right), \qquad i = 1, \ldots, n.$$

Again let

$$\mathcal{P} = \left\{ \mathbf{f} \in \mathbf{R}^m : \mathbf{f} = \sum_{i=1}^{n} \mathbf{f}_i p_i, \; \sum_{i=1}^{n} p_i = 1 \right\}$$

denote the portfolio flat in $\mathbf{R}^m$.

The nonsingular case

Set $F = [\mathbf{f}_1, \ldots, \mathbf{f}_n] \in \mathbf{R}^{m \times n}$. Then the following conditions are equivalent:

1. $\mathbf{f}_1, \ldots, \mathbf{f}_n$ are linearly independent;
2. $\mathcal{F} = \text{range}(F)$, the linear subspace of $\mathbf{R}^m$ spanned by the $\mathbf{f}_i$, has dimension n;
3. $\text{rank}(F) = n$;
4. The variance-covariance matrix $V = F^T F$ is nonsingular;
5. The portfolio flat $\mathcal{P}$ has dimension $n - 1$ and $\mathbf{0} \notin \mathcal{P}$.

These conditions identify the *nonsingular* case. In this section we assume they are true.

Let $\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}$ be an orthonormal basis for the tangent space of the portfolio flat,

$$\mathrm{T}(\mathcal{P}) = \left\{ \mathbf{v} \in \mathbf{R}^m : \mathbf{v} = \sum_{i=1}^{n} \mathbf{f}_i t_i, \; \sum_{i=1}^{n} t_i = 0 \right\},$$

and (temporarily) set $U = [\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}] \in \mathbf{R}^{m\times(n-1)}$.

Now consider the mean risk vector $\bar{\mathbf{f}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{f}_i$. This is an element of $\mathcal{P}$. It can be decomposed into two orthogonal components,

$$\bar{\mathbf{f}} = UU^T\bar{\mathbf{f}} + (\bar{\mathbf{f}} - UU^T\bar{\mathbf{f}}),$$

with the first part parallel to $\mathcal{P}$ and the second part perpendicular to $\mathcal{P}$.

We refer to the second, normal component of $\bar{\mathbf{f}}$,

$$\mathbf{f}_N = \bar{\mathbf{f}} - UU^T\bar{\mathbf{f}},$$

as the *systemic risk* vector. It is an element of $\mathcal{P}$, and it doesn't depend on $\bar{\mathbf{f}}$ at all. Every portfolio of the risky funds has at least this risk. Indeed, every risk vector $\mathbf{f} \in \mathcal{P}$ has an orthogonal decomposition into parallel and perpendicular parts,

$$\mathbf{f} = (\mathbf{f} - \mathbf{f}_N) + \mathbf{f}_N,$$

and the square risk or variance of any such $\mathbf{f}$ has a corresponding parallel-normal decomposition: $\|\mathbf{f}\|^2 = \|\mathbf{f} - \mathbf{f}_N\|^2 + \|\mathbf{f}_N\|^2$.

Put $\mathbf{u}_N = \mathbf{f}_N/\|\mathbf{f}_N\|$ ($\|\mathbf{f}_N\| \neq 0$ since $\mathbf{0} \notin \mathcal{P}$). Then, for $\mathbf{f} \in \mathcal{P}$,

$$\mathbf{u}_N^T\mathbf{f} = \mathbf{u}_N^T\left(\mathbf{f} - \mathbf{f}_N\right) + \mathbf{u}_N^T\mathbf{f}_N = \mathbf{u}_N^T\mathbf{f}_N = \|\mathbf{f}_N\|.$$

Replace

$$U := [U, \mathbf{u}_N] = [\mathbf{u}_1, \ldots, \mathbf{u}_{n-1}, \mathbf{u}_N] \in \mathbf{R}^{m\times n}.$$

Now the columns of U form an orthonormal basis for $\mathcal{F}$.

Parametrization of $\mathcal{F}$

The mapping $\mathbf{x} \mapsto \mathbf{f} = U\mathbf{x},\ \mathbf{x} \in \mathbf{R}^n$, is an isometric parametrization of $\mathcal{F}$ by $\mathbf{R}^n$, and $\mathbf{x}$ can be recovered from $\mathbf{f} \in \mathbf{R}^n$ by $\mathbf{x} = U^T\mathbf{f}$. This means that geometric relationships in $\mathcal{F}$ are mirrored in $\mathbf{R}^n$ and vise versa. Also note that the portfolio flat $\mathcal{P}$ in $\mathcal{F}$ corresponds to the hyperplane $x_n = \|\mathbf{f}_N\|$ in $\mathbf{R}^n$.

The objects of our interest—points, vectors, flats—live in $\mathcal{F}$, an n-dimensional Euclidean space which, in turn, lies askew in the big space $\mathbf{R}^m$. We want to view these objects—take pictures if you will—from an $\mathbf{R}^n$ frame of reference. An orthonormal basis for $\mathbf{R}^n$, represented by the columns of an orthogonal $n \times n$ matrix, is just the camera for this purpose.

Suppose E_z is an orthonormal $n \times n$ matrix. Set $\mathbf{x} = E_z\mathbf{z}$ above. Then $\mathbf{z} \in \mathbf{R}^n$ is the E_z-coordinate vector of $\mathbf{f} = UE_z\mathbf{z} \in \mathcal{F}$. Moreover $\mathbf{z}$ comes from $\mathbf{f}$ via $\mathbf{z} = E_z^TU^T\mathbf{f}$. As a matter of notation we will let $\mathbf{f}_z$ denote the E_z-coordinate vector of $\mathbf{f}$, so that $\mathbf{f}_z = E_z^TU^T\mathbf{f}$ and $\mathbf{f} = UE_z\mathbf{f}_z$.

The correspondence $\mathbf{f} \leftrightarrow \mathbf{f}_z$ between $\mathcal{F}$ and $\mathbf{R}^n$ is an isometry. It preserves all metric information. In particular $F_z^T F_z = F^T F$, where F_z is the E_z-coordinate representation of $F: \; F_z = E_z^T U^T F$.

This has been a general discussion. We do have a particular orthogonal matrix, E_z, in mind, but, for the time being, we are thinking of the standard basis for $\mathbf{R}^n$ given by $E_x = [\mathbf{e}_1, \ldots, \mathbf{e}_n] = I_n$. Now the equations $\mathbf{f} = UE_x\mathbf{f}_x = U\mathbf{f}_x, \;\; \mathbf{f}_x = E_x^T U^T \mathbf{f} = U^T\mathbf{f}$ represent the isometry we started with. And the E_x-representation of F is

$$F_x = E_x^T U^T F = U^T F.$$

Reward as a linear function of risk

Put $R = [\bar{r}_1, \ldots, \bar{r}_n] \in \mathbf{R}^{1\times n}$ and define $\mathbf{r}_{Sx} = F_x^{-T} R^T \in \mathbf{R}^n$ and $\mathbf{r}_S = U\mathbf{r}_{Sx} \in \mathcal{F}$. Then $\mathbf{r}_S^T F = \mathbf{r}_{Sx}^T F_x = R$.

Now consider an arbitrary portfolio $\mathbf{f} = F\mathbf{p}$ with proportions $\mathbf{p} = [p_1, \ldots, p_n]^T$ of risky funds. Its expected reward and risk are related by

$$\bar{r} = R\mathbf{p} = \mathbf{r}_S^T F\mathbf{p} = \mathbf{r}_S^T \mathbf{f}.$$

In this way expected reward is a linear function of risk: $\bar{r} = \mathbf{r}_S^T\mathbf{f}$. Moreover the reward vector of any such portfolio can be recovered from its risk vector $\mathbf{f}$ via the linear relationship

$$\mathbf{r} = \mathbf{f} + \mathbf{e}_R \left(\frac{\bar{r}}{\sqrt{n_P}} \right), \quad \bar{r} = \mathbf{r}_S^T \mathbf{f}.$$

In summary, the risk vector $\mathbf{f}$ contains *all* the reward information about the portfolio over the selected range of periods. Indeed, even the specific proportions of the funds in the portfolio can be recovered by $\mathbf{p} = F_x^{-1}U^T\mathbf{f}$.

Systemic, productive, and incidental components of risk

Assume the expected rewards $\bar{r}_i$ are not all the same. Then $\mathbf{r}_S$ is not a multiple of $\mathbf{u}_N$, because $\mathbf{r}_S^T F = R$ varies from coordinate to coordinate whereas $\mathbf{u}_N^T F$ is constantly $\|\mathbf{f}_N\|$. As a consequence $\mathbf{u}_S = \mathbf{r}_S / \|\mathbf{r}_S\| \in \mathcal{F}$ and $\mathbf{u}_N$ are linearly independent.

For reasons that will be explained later we refer to $\mathbf{u}_S$ as the *unit Sharpe vector* and the axis through the origin in the $\mathbf{u}_S$ direction as the *Sharpe axis*. Define $\beta = \arcsin(\mathbf{u}_S^T\mathbf{u}_N)$. Then $-\frac{\pi}{2} < \beta < \frac{\pi}{2}$ since $|\mathbf{u}_S^T\mathbf{u}_N| < 1$. If $\beta = 0$, the Sharpe axis and the portfolio flat are parallel and never meet; otherwise β is the angle between the Sharpe axis and the portfolio flat, positive or negative according as the portfolio flat intersects the positive or negative axis.

We refer to the plane spanned by $\mathbf{u}_S$ and $\mathbf{u}_N$ as the *productive-systemic risk plane*. It has orthonormal bases $\{\mathbf{u}_N, \mathbf{u}_T\}$ and $\{\mathbf{u}_H, \mathbf{u}_S\}$, the first basis being a β-rotation of the second:

$$[\mathbf{u}_N \; \mathbf{u}_T] = [\mathbf{u}_H \; \mathbf{u}_S] \begin{bmatrix} \cos\beta & -\sin\beta \\ \sin\beta & \cos\beta \end{bmatrix},$$

where $\mathbf{u}_T$ and $\mathbf{u}_H$ are computed from

$$\mathbf{u}_T = \mathbf{u}_S \sec\beta - \mathbf{u}_N \sec\beta, \quad \mathbf{u}_H = \mathbf{u}_N \cos\beta - \mathbf{u}_T \sin\beta.$$

The vector $\mathbf{u}_T$, being in $\mathcal{F}$ and perpendicular to the normal vector $\mathbf{u}_N$, is tangent to the portfolio flat $\mathcal{P}$. The "T" subscript in $\mathbf{u}_T$ stands for "tangent" (to $\mathcal{P}$). The "H" in $\mathbf{u}_H$ stands for "horizontal"—for lack of a better choice. And, of course, the "N" and "S" in the other two unit vectors stand for "normal" (to $\mathcal{P}$) and "Sharpe", respectively.

Now consider any $\mathbf{f} \in \mathcal{P}$. We can decompose $\mathbf{f}$ into a part parallel to $\mathbf{u}_N$, a part parallel to $\mathbf{u}_T$, and a part perpendicular to both $\mathbf{u}_N$ and $\mathbf{u}_T$:

$$\mathbf{f} = \mathbf{f}_N + \mathbf{f}_T + \left(\mathbf{f} - \mathbf{f}_N - \mathbf{f}_T\right), \quad \mathbf{f}_T = \mathbf{u}_T\left(\mathbf{u}_T^T\mathbf{f}\right).$$

The part parallel to $\mathbf{u}_N$ is the systemic risk $\mathbf{f}_N$. It doesn't depend on $\mathbf{f}$ at all. It is the same for all portfolios.

The expected reward $\bar{r} = \mathbf{r}_S^T\mathbf{f}$ corresponding to $\mathbf{f}$ decomposes as

$$\bar{r} = \mathbf{r}_S^T\mathbf{f}_N + \mathbf{r}_S^T\mathbf{f}_T.$$

The first component is a systemic component. It has nothing to do with the individual portfolio represented by $\mathbf{f}$. It only depends on the system of funds chosen and the sequence of periods on which they are evaluated. The second component of $\bar{r}$ does depend on $\mathbf{f}$, but only on the tangential part $\mathbf{f}_T$. We refer to this tangential part of $\mathbf{f}$ as the *productive* component of risk. It is the part of risk that produces differing expected rewards from one portfolio to the next.

The third part of the risk, $\mathbf{f} - \mathbf{f}_N - \mathbf{f}_T$, is perpendicular to $\mathbf{r}_S$. It makes no contribution to expected reward. This is the *incidental* component of risk.

Finally note that the variable part of $\bar{r}$ can be rewritten as

$$\mathbf{r}_S^T\mathbf{f}_T = \left(\|\mathbf{r}_S\| \cos\beta\right) \mathbf{u}_T^T\mathbf{f}.$$

Since $\cos\beta > 0$, this implies that $\mathbf{u}_T$ is the direction of steepest increase of expected reward on $\mathcal{P}$.

Orthonormal bases (y and z coordinates)

We want to complete the orthonormal set $\left\{\mathbf{u}_T, \mathbf{u}_N\right\}$ to an orthonormal basis for $\mathcal{F}$. This can be accomplished by a Householder reflection of $\mathbf{R}^n$.

Put $\mathbf{u}_{Tx} = U^T\mathbf{u}_T \in \mathbf{R}^n$. Then $\mathbf{u}_{Tx}$ is perpendicular to $\mathbf{e}_n = \mathbf{u}_{Nx} = U^T\mathbf{u}_N$. Set $\mathbf{v} = \mathbf{u}_{Tx} - \mathbf{e}_{n-1} \in \mathbf{R}^n$ and define

$$E_z = I_n - \frac{2}{\mathbf{v}^T\mathbf{v}}\mathbf{v}\mathbf{v}^T.$$

The symmetric, orthogonal matrix E_z represents the reflection through the hyperplane $\mathbf{v}^T\mathbf{x} = 0$ in $\mathbf{R}^n$.

The transformation E_z leaves $\mathbf{e}_n$ fixed, since $\mathbf{v}^T\mathbf{e}_n = 0$, and it swaps $\mathbf{e}_{n-1}$ and $\mathbf{u}_{Tx}$ by construction (or by direct computation). Thus $E_z = E_z I_n$ is an orthogonal $n \times n$ matrix with next to last column $\mathbf{u}_{Tx}$ and last column $\mathbf{e}_n$: $E_z = [\ \ldots, \mathbf{u}_{Tx}, \mathbf{e}_n\ ]$. The columns of E_z, when lifted to $\mathcal{F}$, form the orthonormal basis, $UE_z = [\ \ldots, \mathbf{u}_T, \mathbf{u}_N]$, we are looking for.

To put $\{\mathbf{u}_H, \mathbf{u}_S\}$ into an orthonormal basis for $\mathcal{F}$ simply replace $\mathbf{u}_T$ by $\mathbf{u}_H$ and $\mathbf{u}_N$ by $\mathbf{u}_S$ in UE_z. This produces $UE_y = [\ \ldots, \mathbf{u}_H, \mathbf{u}_S]$, where

$$E_y = E_z \left[\begin{array}{c|cc} I_{n-2} & \multicolumn{2}{c}{0} \\ \hline 0 & -\sin\beta & \cos\beta \\ & \cos\beta & \sin\beta \end{array}\right].$$

Let $\mathbf{f}_y = E_y^T U^T \mathbf{f} = [y_1, \ldots, y_n]^T$ and $\mathbf{f}_z = E_z^T U^T \mathbf{f} = [z_1, \ldots, z_n]^T$ be the E_y and E_z coordinates of $\mathbf{f} \in \mathcal{F}$. Here are some properties of these coordinate systems.

1. $\mathbf{f} \in \mathcal{P}$ iff $z_n = \|\mathbf{f}_N\|$.
2. $\bar{r} = \mathbf{r}_S^T\,\mathbf{f} = \|\mathbf{r}_S\|\, y_n = \|\mathbf{r}_S\|(z_{n-1}\cos\beta + z_n \sin\beta)$.
3. $\|\mathbf{f}\| = \sqrt{y_1^2 + \cdots + y_n^2} = \sqrt{z_1^2 + \cdots + z_n^2}$.
4. The Sharpe ratio of $\mathbf{f}$ is given by

$$s = \frac{\bar{r}}{\|\mathbf{f}\|} = \|\mathbf{r}_S\| \left(\frac{y_n}{\sqrt{y_1^2 + \cdots + y_n^2}} \right) = \|\mathbf{r}_S\| \left(\frac{z_{n-1}\cos\beta + z_n \sin\beta}{\sqrt{z_1^2 + \cdots + z_n^2}} \right).$$

5. $[y_1, \ldots, y_{n-2}, 0, 0]^T = [z_1, \ldots, z_{n-2}, 0, 0]^T$ represents the incidental component of $\mathbf{f}$ in either coordinate system; whereas $\mathbf{e}_{n-1} z_{n-1}$ and $\mathbf{e}_n z_n$ are the E_z-coordinate vectors of the productive and systemic components of $\mathbf{f}$.

6. SVD reduction

In the last section we assumed that the risk vectors $\mathbf{f}_i$ were linearly independent. This certainly will not be the case if $n > m$, if there are more funds under consideration than periods of data. Even if the $\mathbf{f}_i$ are formally independent, it is likely that approximate dependencies will exist. If several funds invest in pretty much the same kind of securities, the variations in their returns from one period to the next may be highly synchronized. This will produce approximate linear dependency among the $\mathbf{f}_i$. Two funds may essentially be parts of a third, as is the case with the Vanguard Growth Index, Value Index, and 500 Index funds. The risk vectors of such funds will be approximately, if not formally, linearly dependent. In any case, whatever the cause, approximate dependency leads to numerical ill-conditioning, and extraneous factors may confuse meaningful inferences.

The linear independence of risk vectors is not essential to our theory. Everything proceeds pretty much as in the independent case if the following two axioms hold.

Axiom 1. There is no riskless portfolio of risky funds.
This is the $\mathbf{0} \notin \mathcal{P}$ condition of the nonsingular case.

Axiom 2. Expected reward is a linear function of risk.
More precisely, there exists an $\mathbf{r}_S \in \mathbf{R}^m$ such that $\bar{r} = \mathbf{r}_S^T \mathbf{f}$ for any portfolio of risky funds with expected reward $\bar{r}$ and risk $\mathbf{f}$.

As we have seen, these axioms are a consequence of independence.

The two axioms may very well not hold for a given set of data (the periodic returns from n risky funds over m successive periods). Nevertheless the returns can generally be adjusted by insignificant amounts so that the axioms do hold.

SVD reduction of risk space

Think of the n risk vectors $\mathbf{f}_i$ as points lying on the portfolio flat $\mathcal{P}$ in Euclidean space $\mathbf{R}^m$. The mean risk, $\bar{\mathbf{f}} = \frac{1}{n} \sum_{i=1}^{n}, \mathbf{f}_i$, also lies on $\mathcal{P}$, right in the middle of these points. Let $\widetilde{\mathbf{f}}_i = \mathbf{f}_i - \bar{\mathbf{f}}$ $(i = 1, \ldots, n)$. The vectors $\widetilde{\mathbf{f}}_i$ span the tangent space, $\mathrm{T}(\mathcal{P})$, of $\mathcal{P}$. They measure how the points $\mathbf{f}_i$ spread out from the central point $\bar{\mathbf{f}}$.

Suppose $\widetilde{F} = [\widetilde{\mathbf{f}}_1, \ldots, \widetilde{\mathbf{f}}_n] \in \mathbf{R}^{m \times n}$ has singular value decomposition $\widetilde{F} = U \Sigma V^T$, where the matrices

$$U = [\mathbf{u}_1, \ldots, \mathbf{u}_m] \in \mathbf{R}^{m \times m} \quad \text{and} \quad V = [\mathbf{v}_1, \ldots, \mathbf{v}_n] \in \mathbf{R}^{n \times n}$$

are orthogonal and

$$\Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbf{R}^{m \times n}, \qquad p = \min(m, n),$$

with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$. Then the $\mathbf{u}_j$ corresponding to nonzero σ_j form an orthonormal basis for $\mathrm{T}(\mathcal{P}) = \mathrm{range}(\widetilde{F})$.

Let $\mathbf{1}_n = [1, \ldots, 1]^T \in \mathbf{R}^n$. By construction $\widetilde{F}\mathbf{1}_n = \mathbf{0}$. It follows that $\mathbf{1}_n \in \mathrm{range}([\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n])$, $r = \mathrm{rank}(\widetilde{F})$. We will assume that $\mathbf{v}_n = \mathbf{1}_n/\sqrt{n}$. If this is not the case, simply replace

$$[\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n] := H[\mathbf{v}_{r+1}, \ldots, \mathbf{v}_n]$$

in V, where

$$H = I_n - 2\,\frac{\mathbf{w}\mathbf{w}^T}{\mathbf{w}^T\mathbf{w}}, \qquad \mathbf{w} = \mathbf{v}_n - \mathbf{1}_n/\sqrt{n}$$

is the Householder reflection of $\mathbf{R}^n$ that swaps $\mathbf{v}_n$ and $\mathbf{1}_n/\sqrt{n}$. Now $\mathbf{v}_n = \mathbf{1}_n/\sqrt{n}$, and $\widetilde{F} = U \Sigma V^T$ is again a singular value decomposition.

For a given $j \leq p$, the row vector $\mathbf{u}_j^T \widetilde{F} = \sigma_j \mathbf{v}_j^T$ lists the coordinates of the vectors $\widetilde{\mathbf{f}}_i$ in the $\mathbf{u}_j$-direction, and, since $\|\mathbf{v}_j\| = 1$, the sum of squares of these coordinates is σ_j^2. Thus

σ_j measures how much the points $\mathbf{f}_i$ spread out from the mean point $\bar{\mathbf{f}}$ in the $\mathbf{u}_j$-direction. For $j > \operatorname{rank}(\widetilde{F})$ there is no spread at all.

The $\mathbf{f}_i$ spread out most in the $\mathbf{u}_1$-direction since σ_1 is the largest singular value. They have less spread in other dimensions and no spread at all in directions orthogonal to $\mathcal{P}$. Fix an *svd-cutoff* ε, $0 \le \varepsilon < 0.1$. We will ignore $\mathbf{u}_j$-directions with $\sigma_j \le \varepsilon\,\sigma_1$, that is to say, ignore dimensions in which the $\mathbf{f}_i$ take up very little space. In our work with quarterly returns, where percentages are rounded to two decimal places, we have generally used ε's in the 0.00–0.05 range.

Let $d = \max\{j : \sigma_j > \varepsilon\,\sigma_1\}$. Set $U_d = [\mathbf{u}_1, \ldots, \mathbf{u}_d]$ and put $\widehat{\mathbf{f}}_i = \bar{\mathbf{f}} + U_d U_d^T \widetilde{\mathbf{f}}_i$ $(i = 1, \ldots, n)$. Then the $\widehat{\mathbf{f}}_i$ are the orthogonal projections of the $\mathbf{f}_i$ onto the d-dimensional flat

$$\widehat{\mathcal{P}} = \bar{\mathbf{f}} + \operatorname{range}(U_d) = \left\{ \mathbf{f} \in \mathbf{R}^m : \ \mathbf{f} = \sum_{i=1}^{n} \widehat{\mathbf{f}}_i\, p_i, \ \ \sum_{i=1}^{n} p_i = 1 \right\}.$$

This is the d-flat that is closest to the $\mathbf{f}_i$ in the least squares sense and

$$\sum_{i=1}^{n} \operatorname{distance}(\mathbf{f}_i, \widehat{\mathcal{P}})^2 = \sum_{i=1}^{n} \|\mathbf{f}_i - \widehat{\mathbf{f}}_i\|^2 = \sum_{j=d+1}^{p} \sigma_j^2.$$

Moreover, if $F = [\mathbf{f}_1, \ldots, \mathbf{f}_n]$ and $\widehat{F} = [\widehat{\mathbf{f}}_1, \ldots, \widehat{\mathbf{f}}_n]$ in $\mathbf{R}^{m\times n}$, then the 2-norm error of approximation is $\|\widehat{F} - F\| = \sigma_{d+1}$, where $\sigma_{d+1} = 0$ if $d = p$. Finally, if the $\mathbf{f}_i$ happen to be substantially independent to start off with, then $d = n - 1$ and $p = n$.

We will be working in the (approximate) risk space $\widehat{\mathcal{F}} = \operatorname{range}(\widehat{F})$, the linear subspace of $\mathbf{R}^m$ spanned by $\widehat{\mathbf{f}}_1, \ldots, \widehat{\mathbf{f}}_n$. The (approximate) portfolio space $\widehat{\mathcal{P}}$ is a hyperplane in $\widehat{\mathcal{F}}$. The point in $\widehat{\mathcal{P}}$ that is closest to the origin is $\widehat{\mathbf{f}}_N = \bar{\mathbf{f}} - UU^T\bar{\mathbf{f}}$. As in the nonsingular case, every risk vector $\widehat{\mathbf{f}} \in \widehat{\mathcal{P}}$ has an orthogonal decomposition

$$\widehat{\mathbf{f}} = \left(\widehat{\mathbf{f}} - \widehat{\mathbf{f}}_N\right) + \widehat{\mathbf{f}}_N$$

into parts parallel to and perpendicular to $\widehat{\mathcal{P}}$. We refer to the normal component, $\widehat{\mathbf{f}}_N$, as the *systemic risk* and assume that $\widehat{\mathbf{f}}_N \ne \mathbf{0}$ or, equivalently, that $\mathbf{0} \notin \widehat{\mathcal{P}}$.

Let $\mathbf{u}_N = \widehat{\mathbf{f}}_N / \|\widehat{\mathbf{f}}_N\| \in \widehat{\mathcal{F}}$. Then $\mathbf{u}_N^T \widehat{\mathbf{f}} = \|\widehat{\mathbf{f}}_N\|$ for every $\widehat{\mathbf{f}} \in \widehat{\mathcal{P}}$, and, reassigning

$$U := \left[U_d, \mathbf{u}_N\right] = \left[\mathbf{u}_1, \ldots, \mathbf{u}_d, \mathbf{u}_N\right] \in \mathbf{R}^{(d+1)\times n},$$

the columns of U form an orthonormal basis for $\widehat{\mathcal{F}}$.

The reduced data

Now the analysis continues just as in the nonsingular case. The mapping $\mathbf{x} \mapsto \widehat{\mathbf{f}} = U\mathbf{x}$, $\mathbf{x} \in \mathbf{R}^{d+1}$ is an isometric parametrization of $\widehat{\mathcal{F}}$ by $\mathbf{R}^{d+1}$. We denote the $\mathbf{x}$-coordinate representation of $\widehat{F}$ by $\widehat{F}_x = U^T\widehat{F} \in \mathbf{R}^{(d+1)\times n}$.

Put $R = [\bar{r}_1, \ldots, \bar{r}_n] \in \mathbf{R}^{1\times n}$. Let $\widehat{\mathbf{r}}_{Sx} = \widehat{F}_x^{+T} R^T \in \mathbf{R}^{d+1}$ and $\widehat{\mathbf{r}}_S = U\widehat{\mathbf{r}}_{Sx} \in \mathcal{F}$, where $\widehat{F}_x^{+T}$ is the pseudo-inverse of $\widehat{F}_x^T$, and define the (approximate) expected reward matrix by

$$\widehat{R} = [\widehat{r}_1, \ldots, \widehat{r}_n] = \widehat{\mathbf{r}}_{Sx}^T \widehat{F}_x = \widehat{\mathbf{r}}_S^T \widehat{F}.$$

Then $\widehat{R} = R$ in the nonsingular case $(d+1 = n)$. Otherwise there is a certain amount of expected reward error $\|\widehat{R} - R\|$. One can show that

$$\|\widehat{R} - R\| = \|R[\mathbf{v}_{d+1}, \ldots, \mathbf{v}_{n-1}]\|,$$

just as

$$\|\widehat{F} - F\| = \|F[\mathbf{v}_{d+1}, \ldots, \mathbf{v}_{n-1}]\| \quad (= \sigma_{d+1}).$$

The vectors $\mathbf{v}_{d+1}, \ldots, \mathbf{v}_{n-1}$ represent directions of change in fund proportions that produce very little change in the risk of a portfolio. We tacitly assume that changes of proportions in these directions produce little change in reward as well.

(At the other extreme consider a universe of two funds, Fund A and Fund B, that invest in exactly the same securities in the same proportions but charge different management fees. Assume that Fund B charges a constant rate more per period than Fund A. Then any increase in the proportion of Fund A in a portfolio of A and B increases the expected reward of the portfolio with no change in risk whatsoever.)

The (approximate) expected reward and risk matrices, $\widehat{R}$ and $\widehat{F}$ are the expected rewards and risks of the (approximate) reward vectors

$$\widehat{\mathbf{r}}_i = \widehat{\mathbf{f}}_i + \mathbf{e}_R \left(\frac{\widehat{r}_i}{\sqrt{n_P}} \right), \quad i = 1, \ldots, n.$$

Just as a reward vector has a risk and an expected reward component, the reward vector error of approximation has risk and expected reward parts. This can be best expressed with the Frobenius norm ($\|A\|_F = \sqrt{\mathrm{trace}(A^T A)}$) in the form

$$\begin{aligned} \| [\widehat{\mathbf{r}}_1 - \mathbf{r}_1, \ldots, \widehat{\mathbf{r}}_n - \mathbf{r}_n] \|_F^2 &= \|\widehat{F} - F\|_F^2 + \tfrac{1}{n_P} \|\widehat{R} - R\|_F^2 \\ &= \sigma_{d+1}^2 + \ldots + \sigma_p^2 + \tfrac{1}{n_P} \left\{ (R\mathbf{v}_{d+1})^2 + \ldots + (R\mathbf{v}_{n-1})^2 \right\}. \end{aligned}$$

The (approximate) periodic returns corresponding to the $\widehat{\mathbf{r}}_i$ can be obtained as

$$\begin{bmatrix} \widehat{X}_i(q_1) \\ \vdots \\ \widehat{X}_i(q_m) \end{bmatrix} = \frac{1}{\sqrt{n_P}} \begin{bmatrix} \frac{1}{\sqrt{\mu_1}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{\mu_m}} \end{bmatrix} \widehat{\mathbf{r}}_i + \begin{bmatrix} X_0(q_1) \\ \vdots \\ X_0(q_m) \end{bmatrix}.$$

Typically, these (approximate) returns do not differ in any meaningful way from the actual returns.

Just as in the nonsingular case we see that an (approximate) risk vector $\widehat{\mathbf{f}} \in \mathcal{P}$ contains all the (approximate) reward information about the portfolio. In particular the (approximate) expected reward is a linear function of the (approximate) risk, $\widehat{r} = \widehat{\mathbf{r}}_S^T \widehat{\mathbf{f}}$. However, one thing does not carry over from the nonsingular case. The specific proportions of the funds in the portfolio cannot necessarily be recovered from $\widehat{\mathbf{f}}$.

Let us examine this situation in more detail. Suppose $\widehat{\mathbf{f}} = \widehat{F}\mathbf{p}$ where $\mathbf{1}_n^T\mathbf{p} = 1$. ($\mathbf{1}_n = [1, \ldots, 1]^T \in \mathbf{R}^n$) Set $\mathbf{q} = \widehat{F}_x^+ \widehat{\mathbf{f}}_x \in \mathbf{R}^n$ ($\widehat{\mathbf{f}}_x = U^T \widehat{\mathbf{f}}$). In the nonsingular case, when $d + 1 = n$, the proportions are determined by the risk vector and $\mathbf{q} = \mathbf{p}$. When $d + 1 < n$ we can only be sure that $\widehat{\mathbf{f}} = \widehat{F}\mathbf{q}$ and $\mathbf{1}_n^T\mathbf{q} = 1$. Indeed, $\mathbf{q}$ is the vector of least 2-norm in $\mathbf{R}^n$ that produces this result.

To see that $\widehat{\mathbf{f}} = \widehat{F}\mathbf{q}$ we rely on the fact that $\widehat{F}_x \widehat{F}_x^+ = I_{d+1}$ (since $\widehat{F}_x$ is a $(d + 1) \times n$ matrix of rank $d + 1$). Then

$$\widehat{\mathbf{f}}_x = \widehat{F}_x \widehat{F}_x^+ \widehat{\mathbf{f}}_x = \widehat{F}_x \mathbf{q},$$

and this lifts to $\widehat{\mathbf{f}} = \widehat{F}\mathbf{q}$ via U. The relation $\mathbf{1}_n^T\mathbf{q} = 1$ follows from the fact that $\mathbf{u}_N^T \widehat{F} = \|\widehat{\mathbf{f}}_N\| \, \mathbf{1}_n^T$:

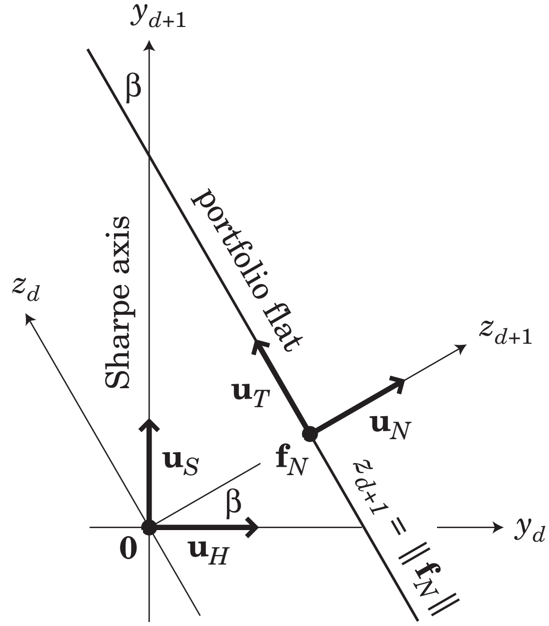
$$\mathbf{1}_n^T\mathbf{q} = \|\widehat{\mathbf{f}}_N\|^{-1} \mathbf{u}_N^T \widehat{F}\mathbf{q} = \|\widehat{\mathbf{f}}_N\|^{-1} \mathbf{u}_N^T \widehat{F}\mathbf{p} = \mathbf{1}_n^T\mathbf{p} = 1.$$

Continuation

Now forget about the original data. Drop the circumflexes over the $\mathcal{P}, \mathcal{F}, \mathbf{f}, \ldots$ and replace $\widehat{r}$'s with $\bar{r}$'s. We will work with the svd-reduced data set, where Axiom 1 ($\mathbf{f}_N \neq \mathbf{0}$) and Axiom 2 ($\bar{r} = \mathbf{r}_S^T \mathbf{f}$) do hold.

Everything about the components of risk from the last section holds verbatim now, except that n must be replaced by $d + 1$ in certain places. The portfolio hyperplane $\mathcal{P}$ makes an angle β with the Sharpe axis in risk space $\mathcal{F}$. The E_y and E_z coordinate systems for $\mathbf{R}^{d+1}$ are defined the same way, and the five properties of these systems need only be rewritten as

1. $\mathbf{f} \in \mathcal{P}$ iff $z_{d+1} = \|\mathbf{f}_N\|$.
2. $\bar{r} = \mathbf{r}_S^T \mathbf{f} = \|\mathbf{r}_S\| \, y_n = \|\mathbf{r}_S\|(z_d \cos\beta + z_{d+1} \sin\beta)$.
3. $\|\mathbf{f}\| = \sqrt{y_1^2 + \cdots + y_{d+1}^2} = \sqrt{z_1^2 + \cdots + z_{d+1}^2}$.
4. The Sharpe ratio of $\mathbf{f}$ is given by
$$s = \frac{\bar{r}}{\|\mathbf{f}\|} = \|\mathbf{r}_S\| \left(\frac{y_{d+1}}{\sqrt{y_1^2 + \cdots + y_{d+1}^2}} \right) = \|\mathbf{r}_S\| \left(\frac{z_d \cos\beta + z_{d+1} \sin\beta}{\sqrt{z_1^2 + \cdots + z_{d+1}^2}} \right).$$
5. $[y_1, \ldots, y_{d-1}, 0, 0]^T = [z_1, \ldots, z_{d-1}, 0, 0]^T$ represents the incidental component of $\mathbf{f}$ in either coordinate system; whereas $\mathbf{e}_d z_d$ and $\mathbf{e}_{d+1} z_{d+1}$ are the E_z-coordinate vectors of the productive and systemic components of $\mathbf{f}$.



7. A look at some data

Let us apply our techniques to some quarterly returns from the funds mentioned in section 1.

Risky funds:

`VFINX:` Vanguard 500 Index
`VIGRX:` Vanguard Growth Index
`VIVAX:` Vanguard Value Index
`VTSMX:` Vanguard Total Stock Market Index
`VEXMX:` Vanguard Extended Market Index
`VBMFX:` Vanguard Total Bond Market Index

Risk-free fund:

`VMMXX:` Vanguard Prime Money Market

We will work with three years of data from the second quarter of 1997 through the first quarter of 2000, comparing results for weight systems $w = 0$ and $w = 2$:

weights	$w = 0$	$w = 2$
$\mu_{1:4}$	0.0833	0.1557
$\mu_{5:8}$	0.0833	0.0658
$\mu_{9:12}$	0.0833	0.0286

The relative singular values of the respective portfolio flats in $\mathbf{R}^{12}$ are

$$w = 0: \qquad \sigma_i/\sigma_1 = 1.0000,\ 0.4573,\ 0.3160,\ 0.0216,\ 0.0033.$$
$$w = 2: \qquad \sigma_i/\sigma_1 = 1.0000,\ 0.4173,\ 0.3606,\ 0.0152,\ 0.0022.$$

These values show that, while either weight system produces a five-dimensional portfolio flat, the data have almost no width in the 4th and 5th singular directions. An svd-cutoff of $\varepsilon = 5\%$ will reduce either portfolio flat dimension to three. Table 2 (Appendix A) shows how the corresponding svd-reduced returns compare with the actual returns. The relative 2-norm errors of the svd-reduced return matrices are about 1%. From a practical standpoint there is really no difference between the svd-reduced returns and the real data.

The effect of weights

The choice of weight system may have a substantial effect on computed expected rewards and risks. The following table illustrates this quite graphically. Besides statistics for the six risky funds it includes statistics for the "Sharpe-optimal long portfolio," dubbed `SOLNG`. This is the long portfolio of the risky funds with maximal Sharpe ratio. The table shows only nonsingular ($\varepsilon = 0$) results. We will examine the effects of svd-reduction later.

Some Statistics 12 Quarters 1997q2–2000q1							
statistic	VFINX	VIGRX	VIVAX	VTSMX	VEXMX	VBMFX	SOLNG
weight system: $w = 0.0$ svd-cutoff: $\varepsilon = 0$							
expected reward	21.18	28.23	13.34	21.38	23.85	1.35	7.99
scalar risk	17.77	18.88	17.66	19.17	25.67	3.20	4.63
Sharpe ratio	1.192	1.496	0.755	1.115	0.929	0.422	1.728
SOLNG percentage	0.0%	24.7%	0.0%	0.0%	0.0%	75.3%	100.0%
weight system: $w = 2.0$ svd-cutoff: $\varepsilon = 0$							
expected reward	15.89	23.22	7.72	18.11	28.13	−0.95	15.44
scalar risk	17.51	18.86	17.60	19.34	27.10	3.02	12.51
Sharpe ratio	0.908	1.231	0.439	0.936	1.038	−0.314	1.235
SOLNG percentage	0.0%	67.8%	0.0%	0.0%	0.0%	32.2%	100.0%

Look at the statistics for the first three (large cap) funds. The risks don't change much from one weight system to the other, but the expected rewards are substantially less under the $w = 2.0$ system. The reason for this is apparent from a look at Table 1 (or 2). On average the highest returns were produced in the earliest of the three years of quarters, but, under the $w = 2.0$ system, these returns carry much less weight than the most recent year's returns.

A similar situation occurs with the bond index fund, which had strong returns during the first year under consideration but which actually lost 3% to money market over the 1999q2–2000q1 period. Only "Extended Market Index", `VEXMX`, reverses the trend toward higher $w = 0.0$ expected rewards. Its highest returns (on average) occurred in the 1999q2–2000q1 year, and its $w = 2.0$ expected reward is significantly higher than the $w = 0.0$ one.

Under either weight system the optimal portfolio, `SOLNG`, is a mix of "Growth Index", `VIGRX`, and "Total Bond Market Index", `VBMFX`, but the $w = 0.0$ mix is mostly bond, and the $w = 2.0$ mix is mostly stock. That VBMFX should appear in the $w = 2.0$ SOLNG portfolio at all seems counterintuitive in view of the fact that the expected reward of VBMFX is negative.

Correlation is what counts here. The stock funds are highly correlated. One stock fund turns out to be sufficient for each SOLNG portfolio. On the other hand the bond fund is included in each portfolio because it is negatively correlated with all the stock funds.

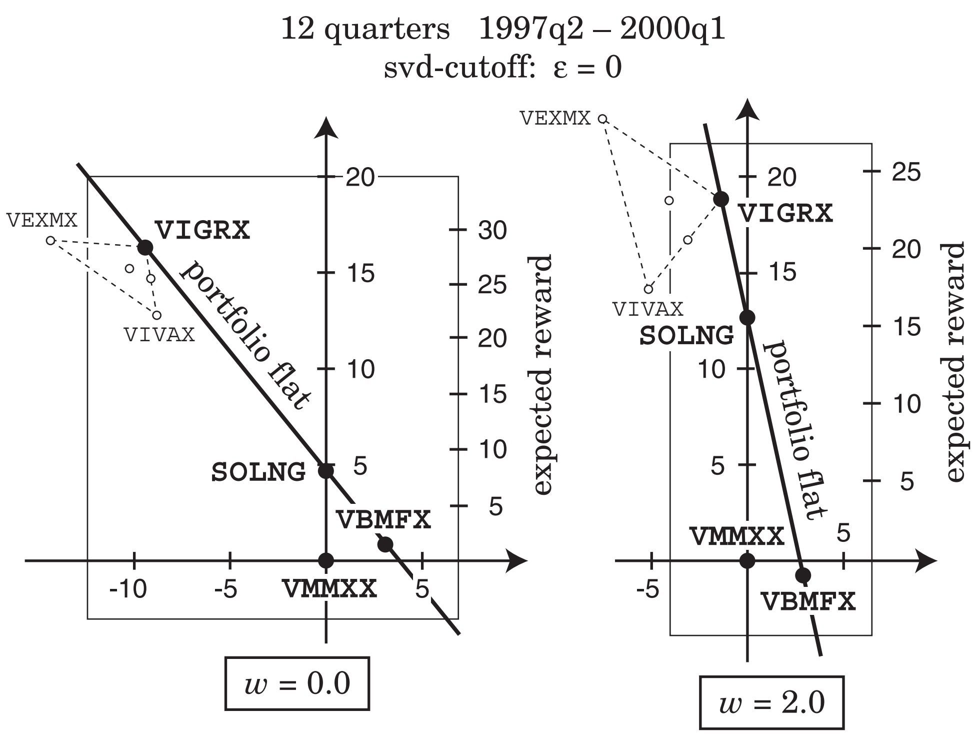
To see why negative correlation counts so much, consider the extreme situation of perfect negative correlation, when one risk vector is a negative multiple of another. Assume that the expected rewards do not exhibit exactly the same relationship. Then one can construct a portfolio in the two funds with positive expected reward and no risk at all. Its Sharpe ratio must, of course, be infinite (well, actually, undefined).

For example, suppose that the risk vectors and expected rewards of funds A and B satisfy

$$\mathbf{f}_A + k\,\mathbf{f}_B = \mathbf{0}$$
$$\bar{r}_A + k\,\bar{r}_B > 0$$

with $k > 0$. Put $p_A = \frac{1}{k+1}$ and $p_B = \frac{k}{k+1}$. Then the portfolio consisting of p_A of fund A and p_B of fund B has zero risk but positive reward $(\bar{r}_A + k\bar{r}_B)/(k+1)$.

Pictures of the VIGRX-VBMFX-VMMXX plane in the $w = 0.0$ and $w = 2.0$ risk spaces show what is happening with our data. The vertical axes are the perpendicular projections of the respective Sharpe axes. The lines labeled "portfolio flat" are the intersections of the full portfolio flats with the respective VIGRX-VBMFX-VMMXX planes.



The effect of weights: 2D sections of two 6D risk spaces

These pictures contain all risk-reward information for the VIGRX and VBMFX funds and the SOLNG portfolio. The risk vectors of the other stock funds take up two-plus dimensions of space off of the VIGRX-VBMFX-VMMXX planes. The small circles show the perpendicular projections of the funds onto the planes, and the dashed triangles bound the projections of all long portfolios in the five stock funds.

Here are the coordinates of the points on the planes to two decimal places.

	$w = 0.0$		$w = 2.0$	
fund	x	y	x	y
VIGRX	−9.46	16.33	−1.39	18.81
VBMFX	3.10	0.78	2.92	−0.77
SOLNG	0.00	4.62	0.00	12.51
VMMXX	0.00	0.00	0.00	0.00

The expected reward scales on the right of the pictures correspond to the Sharpe ratios of the SOLNG portfolio: 1.728 units of expected reward per unit of vertical risk under the $w = 0.0$ system and 1.235 units of expected reward per unit of vertical risk under the $w = 2.0$ system. These scales and the coordinates of the points determine all of the previously presented statistics for VIGRX, VBMFX, and SOLNG. For example, under the $w = 0.0$ system, the scalar risk of VIGRX is $\sqrt{(-9.46)^2 + 16.33^2} = 18.87$, and the expected return is $16.33 \times 1.728 = 28.22$. The one digit differences in the last decimal places from the previous values are due to rounding.

We have emphasized the significance of the negative correlations of the stock funds with VBMFX. The correlations of VIGRX with VBMFX can be computed from the above table as the cosines of the angles between the corresponding vectors:

$$\begin{aligned} \mathrm{cor}(\texttt{VIGRX}, \texttt{VBMFX}) &= -0.275 \qquad \text{under weight system } w = 0.0, \\ \mathrm{cor}(\texttt{VIGRX}, \texttt{VBMFX}) &= -0.326 \qquad \text{under weight system } w = 2.0. \end{aligned}$$

The angles are 106° and 109° respectively.

In this analysis we have restricted our attention to the 1997q2–2000q1 period. Clearly weight systems can have a significant effect on various statistics. The long-term effect of the weight system chosen will be discussed in more detail later; however it does appear that an investment strategy based on the $w = 2.0$ system is far more effective than one based on the $w = 0.0$ system.

Svd-reduction

Now fix the weight system at $w = 2.0$. Let us examine the effects of the singular value dimension reduction produced by $\varepsilon = 5.0\%$. As noted earlier this svd-cutoff will reduce the portfolio flat dimension of our data by 2 (from 5 to 3), but the reduction has very little effect on the statistics we have just considered. The following table is almost the same as the corresponding portion of the previous ($\varepsilon = 0$) one.

Some Statistics 12 Quarters 1997q2–2000q1 weight system: $w = 2.0$ svd-cutoff: $\varepsilon = 5.0\%$							
statistic	VFINX	VIGRX	VIVAX	VTSMX	VEXMX	VBMFX	SOLNG
expected reward	15.84	22.93	7.48	18.85	27.96	−0.95	15.57
scalar risk	17.51	18.86	17.61	19.32	27.10	3.02	12.78
Sharpe ratio	0.905	1.216	0.425	0.976	1.032	−0.313	1.218
SOLNG percentage	0.0%	69.2%	0.0%	0.0%	0.0%	30.8%	100.0%

The reason these statistics change so little from the previous ones is that they are interpolative. For example the SOLNG portfolio is between (in the convex hull of) the other data points. Interpolative statistics are not significantly affected by singular value reduction.

Before looking at the effect of svd-reduction on certain extrapolative statistics, let us examine how dependencies among the data may be exposed by singular value decomposition. As in the previous section let $\widetilde{F}$ be the matrix of deviations of the fund reward vectors from their arithmetic mean and $\widetilde{F} = U\Sigma V^T$ its singular value decomposition. The upper box of the following matrix shows the relative singular values and the corresponding right singular vectors that are effectively discarded by an svd-cutoff of 5.0%.

12 quarters 1997q2–2000q1 $w = 2.0$ Gauss-Jordan reduction of the singular rows of V^T						
σ_i/σ_1	VFINX	VIGRX	VIVAX	VTSMX	VEXMX	VBMFX
1.52%	−0.0581	−0.3413	−0.2746	0.8738	−0.2028	0.0030
0.22%	0.8462	−0.3763	−0.3270	−0.1840	0.0407	0.0004
0	0	−0.43	−0.34	1	−0.23	0.00
0	1	−0.54	−0.46	0	0.00	0.00

First a comment about the rows of V^T. By construction the last (n^{th}) row of V^T is $\mathbf{1}_n^T = [1,\ldots,1]/\sqrt{n}$ (with $n = 6$ in our case). This row must be orthogonal to all the other rows; consequently the coefficients of the other rows, in particular the top two rows above, must sum to zero. These rows represent unit vectors tangent to the portfolio flat coefficient space.

Let us be more specific. A portfolio coefficient vector is an n-vector that sums to one. The coefficients are the proportions of each fund in the portfolio. The difference of two portfolio coefficient vectors is thus an n-vector that sums to zero. All such zero-sum n-vectors make up the tangent space of the portfolio flat coefficient space.

Look at the $\sigma_4/\sigma_1 = 1.52\%$ row above. This row displays the coefficient direction $\mathbf{v}_4$. (For us "direction" is synonymous with "unit vector.") A change in portfolio coefficients by a certain amount in the $\mathbf{v}_4$ direction will produce only 1.52% as much change in (vector) risk as a change in coefficients of the same magnitude in the $\mathbf{v}_1$ direction.

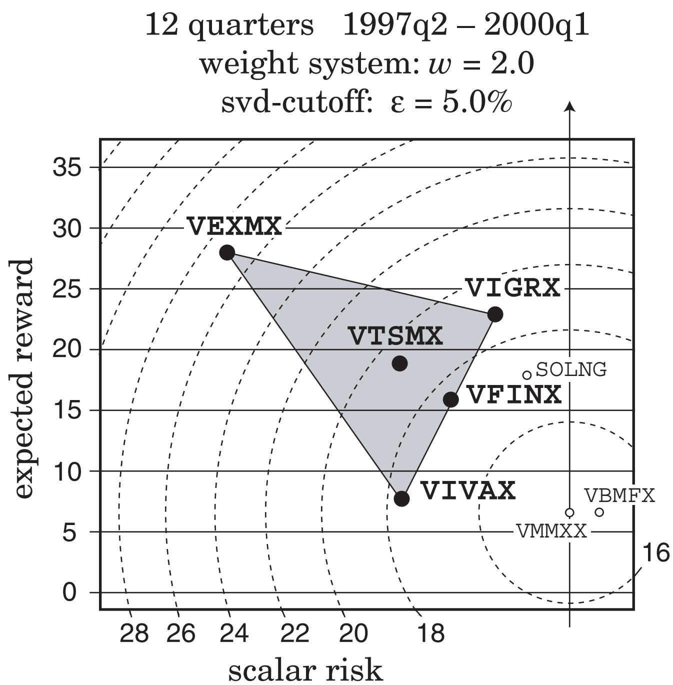
An $\varepsilon = 5.0\%$ svd-reduction essentially zeros the relative singular values shown. This means that coefficient changes in the $\mathbf{v}_4$ ($\sigma_4/\sigma_1 = 1.52\%$) and $\mathbf{v}_5$ ($\sigma_4/\sigma_1 = 0.22\%$) directions no longer produce any change in risk or reward at all. The last two rows of the above matrix reflect the effects of this reduction. They were obtained from the previous two (V^T) rows by elementary row operations after the relative singular values had been set to zero. The final results were rounded to two decimal places.

How should we interpret these rows? Look at the last row for example. A change from 100% in the 500 Index Fund to 54% in Growth Index and 46% in Value Index now produces no change in risk or reward. The 500 Index Fund is now this mix of Growth and Value. Thus the last two rows of the matrix display the (approximate) dependencies among the stock funds:

$$\texttt{VTSMX} = 43\%\ \texttt{VIGRX} + 34\%\ \texttt{VIVAX} + 23\%\ \texttt{VEXMX}$$
$$\texttt{VFINX} = 54\%\ \texttt{VIGRX} + 46\%\ \texttt{VIVAX}$$

The 500 Index, `VFINX`, and the Total Stock Market Index, `VTSMX`, have become redundant. They are now, in essence, portfolios in the other three stock funds.

After svd-reduction by $\varepsilon = 5.0\%$, the 5 stock funds essentially lie on a plane in risk space. The following picture of the plane graphically displays the convex dependencies of `VTSMX` and `VFINX` on the other three stock funds. The shaded triangle represents all long portfolios in the five stock funds.



Svd-reduction: stock fund plane

The vertical axis here is the perpendicular projection of the Sharpe axis onto the plane. The actual Sharpe axis does not meet the plane at any point.

The small circles represent projections of funds that are not on the plane. They are the planar points that are closest to the respective funds. The indicated risk and reward levels do not apply to these points.

The large dashed circles are curves of constant scalar risk. They are the intersections of 3-spheres about the origin of risk space with the stock fund plane. The planar point labelled `VMMXX` is 15.23 risk units from the origin, `VMMXX`, of risk space, and the central 3-sphere of radius 15.23 is tangent to the stock fund plane at this point. The constant risk circles (of radius necessarily greater than 15.23) are centered at the `VMMXX` point.

Finally, the points that are not on the stock fund plane do not lie on one side or the other of the plane. This plane lives in 4-dimensional space; it does not divide the space in half. In fact the stock fund plane is the intersection of two distinct 3-flats—the flat generated by the stock funds and `VBMFX` (this is the portfolio flat) and the flat generated by the stock funds and `VMMXX`.

Reduction and extrapolation

Singular value decomposition may indicate how certain funds are, effectively, portfolios of others. Even if such relationships are not apparent, svd-reduction can save one from making fanciful extrapolations on noise in the return data.

> To extrapolate: To extend a range of values on the assumption that the trend exhibited inside the given part is maintained outside of it.
>
> — The New Oxford Shorter Dictionary

Here are some extrapolative statistics for the data under consideration. These statistics are affected very substantially by svd-reduction. For comparative purposes we have included a "four fund" column where the redundant stock funds, `VFINX` and `VTSMX`, have not been considered at all. Note that the four fund statistics are essentially the same as the six fund statistics after svd-reduction by $\varepsilon = 5.0\%$. The slight difference is due to the fact that reduction does not exactly squash the stock funds onto a plane.

Some Extrapolative Statistics 1997q2–2000q1 $w = 2.0$			
statistic	$\varepsilon = 0$	$\varepsilon = 5.0\%$	4 funds
β	17.2°	−6.3°	−6.1°
Sharpe ratio supremum	3.302	1.648	1.655
systemic square risk (variance)	1.2%	1.5%	1.5%
productive square risk (variance)	7.8%	36.1%	37.3%
incidental square risk (variance)	91.0%	62.3%	61.2%

It is rather hard to see what is going on when you are in at least four dimensions. Let us give a three-dimensional example of the extrapolative nature of systemic risk.

Imagine several data points lying on a line in 3-dimensional space. Say that they spread out 20 cm from first to last. And suppose the line is 10 cm from the origin, though there is no reason to assume that the point on the line closest to the origin is in the midst of the data. This 10 cm would correspond to what we have called systemic (scalar) risk.

Now suppose that the points don't actually lie on a line at all, but rather on a very thin, flat ribbon that is 20 cm long. The ribbon just looks like a line, it is so thin. This ribbon determines a plane. If the ribbon tilts a bit, so does the plane. The plane cannot be more than 10 cm from the origin, but it could be arbitrarily close, depending on the tilt of the ribbon. The distance from the plane to the origin corresponds to our systemic (scalar) risk before singular value reduction. It is mostly noise (or tilt, if you will). It has no particular meaning. Thin ribbon or not, the data essentially spreads out 20 cm along a line that is 10 cm from the origin.

In this paper we will be especially interested in Sharpe-optimal portfolios. These are portfolios that maximize the Sharpe ratio subject to certain constraints. The Sharpe-optimal long portfolio, `SOLNG`, is uniquely determined by the data. This portfolio lies in the midst of the funds. It is an interpolative object.

(We think of a portfolio as its risk vector. It is in this sense that the `SOLNG` portfolio is unique. The proportions of the individual funds are not uniquely determined by the risk vector when dimension has been reduced by svd-reduction. Generally we exhibit pseudo-inverse proportions—proportions with minimal sum of squares.)

On the other hand, the unconstrained Sharpe-optimal portfolio, `SOLSH` (for Sharpe-optimal long-short portfolio), allows short selling. This portfolio may be far removed from the fund points. It is an extrapolative object.

(Actually `SOLSH` exists only when $\beta > 0$. Otherwise the Sharpe ratio supremum is achieved as the limit

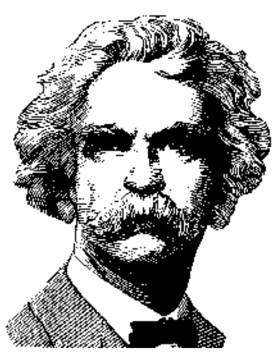
$$\text{Sharpe ratio supremum} = \lim_{t\to\infty} \frac{\mathbf{r}_S^T\,(\mathbf{f}_N + \mathbf{u}_T\,t)}{\|\mathbf{f}_N + \mathbf{u}_T\,t\|} = \mathbf{r}_S^T\mathbf{u}_T = \|\mathbf{r}_S\|\cos\beta.\ )$$

Here are some statistics for the `SOLNG` and `SOLSH` portfolios. We include the four fund case for comparative purposes. The asterisked `SOLSH` portfolios, which correspond to the $\varepsilon = 5.0\%$ and four fund cases, represent 95% of the respective Sharpe ratio suprema.

The statistics for the interpolative, `SOLNG`, portfolios are pretty much the same. Svd-reduction has little effect on these. On the other hand look at the `SOLSH` portfolio corresponding to $\varepsilon = 0$. It has substantially higher reward and less risk than any of the other portfolios, but its (uniquely determined) fund proportions are quite bizarre. This extrapolative portfolio is essentially constructed out of noise, noise that is eliminated by singular value reduction. Its risk and reward figures may be very inviting, but don't bet the farm!

	Sharpe-optimal Portfolios 1997q2–2000q1 $w = 2.0$					
	SOLNG (interpolative)			SOLSH* (extrapolative)		
statistic	$\varepsilon = 0$	$\varepsilon = 5.0\%$	4 funds	$\varepsilon = 0$	$\varepsilon = 5.0\%$	4 funds
VFINX portion	0	0	—	643.9%	0	—
VIGRX portion	67.8%	69.2%	67.8%	501.3%	66.1%	66.1%
VIVAX portion	0	0	0	357.0%	−65.3%	−63.7%
VTSMX portion	0	0	—	−1944.3%	0	—
VEXMX portion	0	0	0	461.6%	23.9%	22.8%
VBMFX portion	32.2%	30.8%	32.2%	80.5%	75.3%	74.8%
expected reward	15.44	15.57	15.44	23.28	16.23	16.13
scalar risk	12.51	12.78	12.51	7.05	10.37	10.26
Sharpe ratio	1.235	1.218	1.235	3.302	1.566*	1.573*

We can't resist closing this section with Mark Twain's description of the wonders of extrapolation.



In the space of one hundred and seventy-six years the Lower Mississippi has shortened itself two hundred and forty-two miles. This is an average of a trifle over one mile and a third per year. Therefore, any calm person, who is not blind or idiotic, can see that in the Old Oolitic Silurian Period, just a million years ago next November, the Lower Mississippi River was upward of one million three hundred thousand miles long, and stuck out over the Gulf of Mexico like a fishing-rod. And by the same token any person can see that seven hundred and forty-two years from now the Lower Mississippi will be only a mile and three-quarters long, and Cairo and New Orleans will have joined their streets together, and be plodding comfortably along under a single mayor and a mutual board of aldermen. There is something fascinating about science. One gets such wholesale returns of conjecture out of such a trifling investment of fact. — Twain [1917, p.156]

8. Sharpe conics

Let us return to the general framework. We have a portfolio flat $\mathcal{P}$ of dimension d lying in risk space $\mathcal{F}$ of dimension $d+1$. We assume that Axiom 1 and Axiom 2 hold:

Axiom 1. $\mathbf{0} \notin \mathcal{P}$.

Axiom 2. There exists an $\mathbf{r}_S \in \mathcal{F}$ such that $\bar{r} = \mathbf{r}_S^T \mathbf{f}$ for any $\mathbf{f} \in \mathcal{P}$.

To eliminate a trivial and extremely unlikely case, we will add

Axiom 3. Not all $\mathbf{f} \in \mathcal{P}$ have the same expected reward.

We are primarily interested in the portfolio flat. Before continuing let us list some salient features of that space:

1. The portfolio flat is spanned by the points that represent the n risky funds. Said another way, every point in the portfolio flat is a linear combination of the risky fund points, where the coefficients of the combination sum to 1.

2. There is a unique point in the portfolio flat that is closest to the origin of risk space. This is the systemic portfolio, $\mathbf{f}_N$. We will also use the symbol `SYSTM` to designate the systemic portfolio.

3. There is a distinguished line in the portfolio flat that we call the *productive risk axis*. It is the line through $\mathbf{f}_N$ in the direction of increasing expected reward—in the $\mathbf{u}_T$ direction. This directed line is the perpendicular projection of the Sharpe axis onto the portfolio flat. Under this projection $\mathbf{0}$ maps to $\mathbf{f}_N$.

4. There is a distinguished hyperplane (subflat of codimension 1) in the portfolio flat that consists of all portfolios with zero expected reward. This zero reward hyperplane is perpendicular to the productive risk axis. It divides the portfolio flat in half. We think of the positive expected reward half as the upper half and the negative expected reward half as the lower half of the portfolio flat.

5. The angle $\beta = \arcsin(\mathbf{u}_S^T \mathbf{u}_N)$, of the portfolio flat with the Sharpe axis, is positive (resp. negative) if and only if the systemic portfolio, $\mathbf{f}_N$, is in the upper half (resp. lower half) of the portfolio flat. This is an immediate consequence of the relation $\mathbf{r}_S^T \mathbf{f}_N = \|\mathbf{r}_S\| \|\mathbf{f}_N\| \mathbf{u}_S^T \mathbf{u}_N$. (If $\beta = 0$, the Sharpe axis is parallel to and does not meet the portfolio flat, and $\mathbf{f}_N$ is on the zero expected reward hyperplane.)

6. The hypersurfaces of constant expected reward in the portfolio flat are hyperplanes parallel to the zero expected reward hyperplane.

7. The hypersurfaces of constant scalar risk in the portfolio flat are hyperspheres centered at the systemic portfolio, $\mathbf{f}_N$.

(We use the prefix "hyper" above and in the sequel to mean codimension 1. Thus a hypersurface in the d-dimensional portfolio flat is submanifold of dimension $d-1$.)

Now let us see what the hypersurfaces of constant Sharpe ratio in the portfolio flat look like. In view of Axiom 2 the Sharpe ratio of $\mathbf{f} \in \mathcal{P}$ can be written as

$$s = \frac{\mathbf{r}_S^T \mathbf{f}}{\|\mathbf{f}\|} = \|\mathbf{r}_S\| \cos \alpha,$$

where α is the angle between $\mathbf{f}$ and $\mathbf{r}_S$. If s remains constant, so does α. It follows that the set of all $\mathbf{f} \in \mathcal{P}$ with Sharpe ratio s is the intersection of the α-cone about the Sharpe axis in $\mathcal{F}$ with the portfolio flat $\mathcal{P}$. This intersection is (a connected component of) a conic section of $\mathcal{P}$.

(Cone above means positive cone: only nonnegative multiples of nonzero vectors on the cone are again on the cone [unless $\alpha = \pi/2$ when the α-cone is the zero reward hyperplane of risk space].)

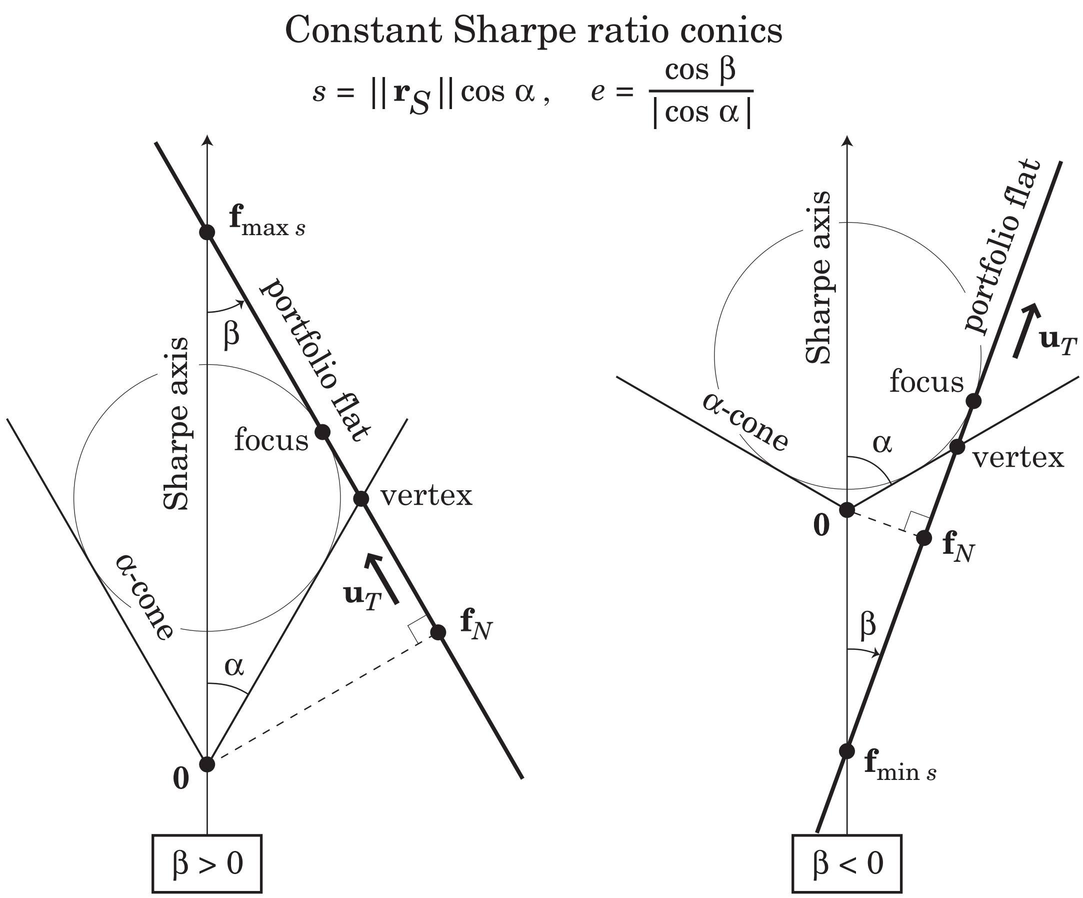
Suppose for the moment that $d = 2$, that the portfolio flat is a plane. The pertinent vertex of the conic section is the intersection of the α-cone with the productive risk axis on the portfolio flat. If there are two points of intersection, the conic is an ellipse and the pertinent vertex is the point of intersection closest to the systemic portfolio, $\mathbf{f}_N$.

Now we follow the elegant elaboration of Hilbert and Cohen-Vossen [1952, Ch.1]. Imagine that the Sharpe axis points up and α is less than or equal to $\pi/2$. Drop a small sphere into the bottom of the α-cone (keeping the center on the Sharpe axis if $\alpha = \pi/2$) and blow the sphere up until it just touches the portfolio plane. (If α is greater than $\pi/2$, the vertex, $\mathbf{0}$, of the α-cone is at the top. In this case simply reverse gravity: float a small sphere up to the top of the cone, etc.) The point where the sphere touches the portfolio plane is the focus of the conic section. The eccentricity is $e = \cos\beta/|\cos\alpha|$.

These parameters—vertex, focus, and eccentricity—completely determine a connected component of a conic section on the portfolio plane via the polar equation

$$\rho = \frac{|FV|(1+e)}{1+e\cos\theta}.$$

Here $|FV|$ is the distance from the focus to the vertex, ρ is the distance from the focus to a variable point on the conic, and θ is the angle between the vector from the focus to the vertex and the vector from the focus to the variable point.



The angle β between the Sharpe axis and the portfolio flat ranges over $-\pi/2 < \beta < \pi/2$. The range of α for which the α-cone intersects the portfolio flat is limited by β. This table describes the situation in the planar ($d = 2$) case.

Constant Sharpe ratio conic sections – planar case			
range of α	$\beta > 0$	$\beta = 0$	$\beta < 0$
$\alpha = 0$	point (max s)	empty	empty
$0 < \alpha < \|\beta\|$	ellipse	empty	empty
$\alpha = \|\beta\|$	up parabola	empty	empty
$\|\beta\| < \alpha < \frac{\pi}{2}$	up hyperbola	up hyperbola	up hyberbola
$\alpha = \frac{\pi}{2}$	line ($\bar{r} = 0$)	line ($\bar{r} = 0$)	line ($\bar{r} = 0$)
$\frac{\pi}{2} < \alpha < \pi - \|\beta\|$	down hyperbola	down hyperbola	down hyberbola
$\alpha = \pi - \|\beta\|$	empty	empty	down parabola
$\pi - \|\beta\| < \alpha < \pi$	empty	empty	ellipse
$\alpha = \pi$	empty	empty	point (min s)

The axis of productive risk is the principal axis of all of these conic sections since the focus and vertex of any section clearly lie on this axis. Sections corresponding to $\alpha < \pi/2$ are in the upper half of the portfolio flat. Those corresponding to $\alpha > \pi/2$ are in the lower half. And $\alpha = \pi/2$ produces the zero reward hyperplane of the portfolio flat. (In the planar case, "hyperplane" translates to line.)

The higher-dimensional situation is much the same as the 2-dimensional case. Each constant Sharpe ratio conic section of the portfolio flat that is not empty, a point, or the zero expected reward hyperplane is a connected hypersurface of revolution about the productive risk axis of the portfolio flat. Any such α-section has a vertex and focus on the productive risk axis and an eccentricity e given by the previous formula, and the intersection of this hypersurface with any plane in the portfolio flat containing the productive risk axis is a planar conic section with just this focus, vertex, and eccentricity. Thus the classifications of the higher-dimensional sections can be read directly from the above table. "Ellipse" should be replaced with "prolate (hyper)spheriod", "parabola" with "(hyper)paraboloid of revolution", "hyperbola" with "sheet of a (hyper)hyperboloid of revolution of two sheets", and "line" with "(hyper)plane". "Point" and "empty" have the same meaning in any dimension, and the parenthesized "hyper" prefixes need be used only when the portfolio flat dimension is greater than 3.

Sharpe-optimal long-short portfolios

The Sharpe-optimal long-short portfolio (symbol SOLSH) is, by definition, the portfolio of maximal Sharpe ratio. The relationship $s = \|\mathbf{r}_S\| \cos\alpha$ shows that the Sharpe ratio cannot, under any circumstances, exceed $\|\mathbf{r}_S\|$. When $\beta > 0$ this value is achieved by the single portfolio corresponding to $\alpha = 0$, the point of intersection of the Sharpe axis with the portfolio flat. We indicate this portfolio by $\mathbf{f}_{\max \mathbf{s}}$ in the picture above.

If $\beta \le 0$ then there is no maximum Sharpe ratio. The supremum is the limit of Sharpe ratios of portfolios on the productive risk axis:

$$\sup s = \lim_{t\to\infty} \frac{\mathbf{r}_S^T\,(\mathbf{f}_N + \mathbf{u}_T\,t)}{\|\mathbf{f}_N + \mathbf{u}_T\,t\|} = \|\mathbf{r}_S\|\cos\beta\,.$$

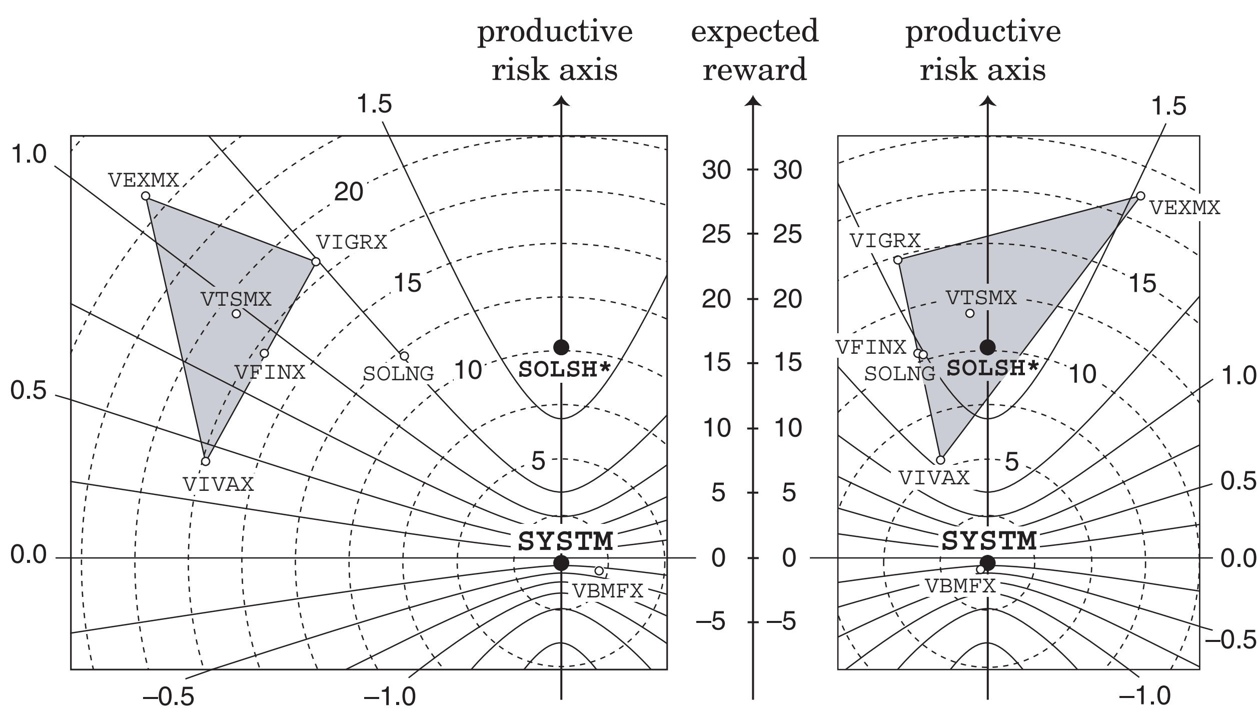
This can be read from the picture.

In the $w = 2.0$, $\varepsilon = 5.0\%$ case of the last section we had $\beta = -6.3°$ and $\|\mathbf{r}_S\| = 1.658$ (not shown). These numbers produced the listed Sharpe ratio supremum, $1.658\cos(-6.3°) = 1.648$. While this supremum is never achieved, we got 95% of it by proceeding up the axis of productive risk to the SOLSH* portfolio.

Pictures of data

Here is a picture of the 3-dimensional portfolio flat for data considered in the last section. The planar slices correspond to the principal components of incidental (horizontal) risk—maximal component on the left, minimal on the right. The planes are mutually perpendicular and each contains the productive risk axis.

The dashed circles are intersections of spheres of constant scalar risk (centered at the systemic portfolio, SYSTM) with the planes. The solid curves show how the hyperboloids of revolution of constant Sharpe ratio cut the planar sections. The solid points (SYSTM and SOLSH*) are on the planes, the hollow fund points are not.



The Portfolio Flat

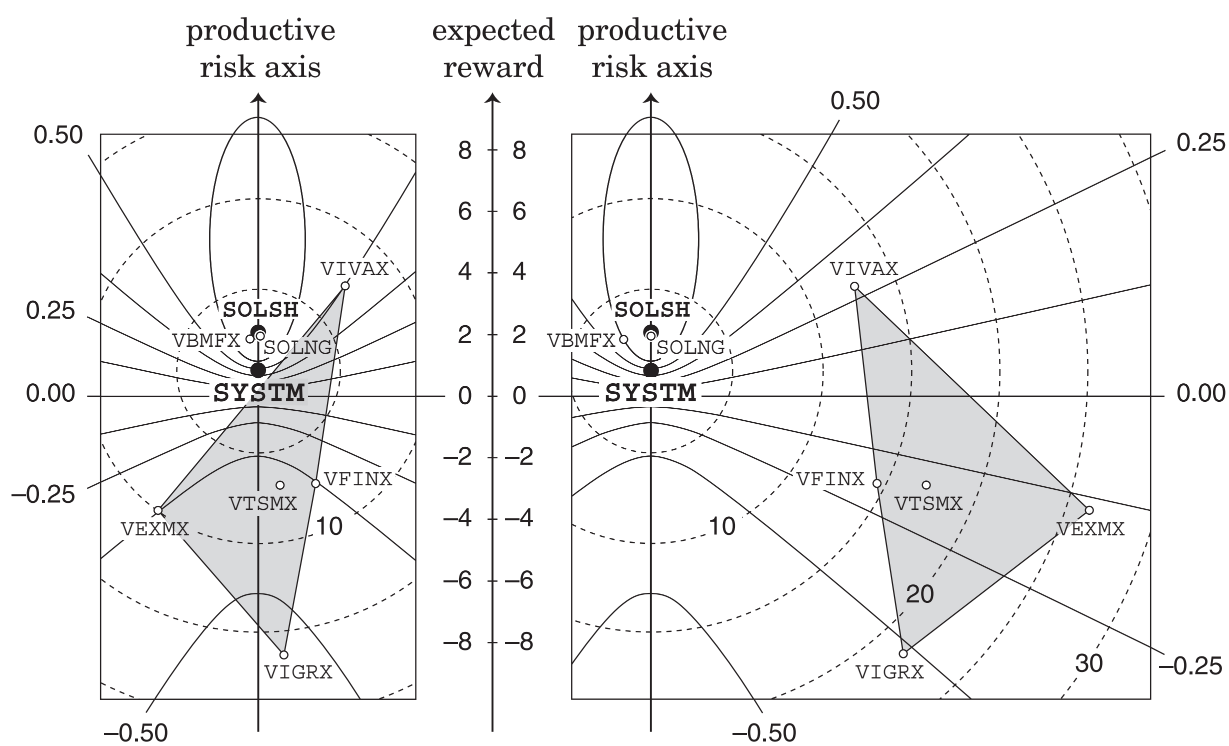
12 quarters 1997q2 – 2000q1

weight system: $w = 2.0$, svd-cutoff: $\varepsilon = 5.0\%$

The first quarter of 2000 marked the end of America's "irrational exuberance" over stocks, particularly information technology stocks. The Nasdaq index reached an all-time high of 5048.62 on March 10, 2000. As we write this paragraph, in the spring of 2002, the Nasdaq is just under 1800. The following table, covering the 1998–2000 period, shows how our index funds were affected by the shift in sentiment. Among the stock funds only the Value Index now has positive expected reward.

Some Statistics 12 Quarters 1998q1–2000q4 weight system: $w = 2.0$ svd-cutoff: $\varepsilon = 5.0\%$							
statistic	VFINX	VIGRX	VIVAX	VTSMX	VEXMX	VBMFX	SOLNG
expected reward	−2.84	−8.39	3.57	−2.90	−3.71	1.85	1.99
scalar risk	14.90	21.79	13.67	17.18	26.97	2.99	2.69
Sharpe ratio	−0.191	−0.385	0.261	−0.169	−0.138	0.620	0.737
SOLNG percentage	0.0%	0.0%	10.1%	0.0%	0.7%	89.2%	100.0%

Here is a picture of the corresponding portfolio flat. Risk levels are about the same as before, but the expected reward scale and the conics of constant Sharpe ratio have changed substantially in three quarters.



The Portfolio Flat

12 quarters 1998q1 – 2000q4

weight system: $w = 2.0$, svd-cutoff: $\varepsilon = 5.0\%$

This picture reads like the previous one. The two planar slices are mutually perpendicular, but now the one on the right shows the incidental risk component of maximal variation. The solid points are on the planes, the hollow points are not. (`SOLSH` is supposed to be solid, `SOLNG` hollow.) The dashed circles centered at `SYSTM` are the intersections of spheres of constant risk with the respective planes. The solid conic sections are the intersections of constant Sharpe ratio surfaces of revolution with the planes.

Now β is positive; consequently the systemic portfolio is in the upper half of the portfolio flat. The Sharpe-optimal long and long-short portfolios are almost the same and 90% bond. They are immediately surrounded by prolate spheroids of constant Sharpe ratio.

The stock funds still span a plane (approximately), but the characteristic, long triangle has done a 180° flip-flop. Now "value" is in and "growth" is out. The stock funds spread out more than before. They are not so highly correlated with one another.

9. The Sharpe-optimal long portfolio

If any risky fund has positive expected reward, then there is a unique long portfolio of risky funds of maximal Sharpe ratio. (Unique as a point on the portfolio flat. The proportions of the individual funds need not be unique.) This is what we are calling the Sharpe-optimal long portfolio and labelling `SOLNG`. Here is the proof of its uniqueness.

Theorem. *Let L be a line segment in the upper half of the portfolio flat, $\mathcal{P}$. Then the Sharpe ratio is either strictly monotone on L or it achieves its maximum on L at a unique interior point.*

Corollary 1. *Let C be any compact, convex subset of $\mathcal{P}$ that has nonempty intersection with the upper half of $\mathcal{P}$. Then there is a unique point in C at which the Sharpe ratio, restricted to C, is maximal.*

Corollary 2. *If any risky fund has positive expected reward, then there is a unique Sharpe-optimal long portfolio of risky funds.*

Proof (of Corollary 1 and Corollary 2): Corollary 2 is just a specialization of Corollary 1. As for Corollary 1, the Sharpe ratio, $s = \bar{r}/\|\mathbf{f}\|$, is defined and continuous on all of $\mathcal{P}$ since $\mathbf{0} \notin \mathcal{P}$. Consequently it has a maximum value on C. If the Sharpe ratio is maximal at two distinct points of C, then the line segment connecting these points lies in C and is also in the upper half of $\mathcal{P}$. But, according to the Theorem, the Sharpe ratio cannot possibly be maximal at two different points on such a line segment. □

Proof (of Theorem): For the moment extend L to an entire line.

If $\bar{r}$ is (a necessarily positive) constant on L, then the Sharpe ratio, $s = \bar{r}/\|\mathbf{f}\|$, is maximal on L where its denominator is smallest, at the unique point of L that is closest to $\mathbf{0}$. ($\mathbf{0}$ is not on L by Axiom 1.) As we move away from this point along L, the denominator,

$\|\mathbf{f}\|$, increases without bound while the numerator, $\bar{r}$, remains a fixed positive number. Thus the Sharpe ratio decreases monotonically toward zero.

On the other hand, if $\bar{r}$ is not constant on L, then L cuts through the $\bar{r} = 0$ hyperplane of $\mathcal{P}$ at some nonzero point $\mathbf{f}_0$. Let $\mathbf{u}$ be the unit vector parallel to L in the increasing expected reward direction. Then $\mathbf{r}_S^T\mathbf{u} > 0$ and

$$L = \{\mathbf{f}_0 + \mathbf{u}\,t \;:\; -\infty < t < \infty\},$$

with $t > 0$ describing the part of L in the upper half of $\mathcal{P}$. Moreover the Sharpe ratio on L is given by

$$s = S(\mathbf{f}_0 + \mathbf{u}\,t) = \frac{(\mathbf{r}_S^T\mathbf{u})\,t}{\|\mathbf{f}_0 + \mathbf{u}\,t\|}$$

with derivative

$$\frac{ds}{dt} = \frac{(\mathbf{r}_S^T\mathbf{u})(\mathbf{f}_0 + \mathbf{u}\,t)^T\mathbf{f}_0}{\|\mathbf{f}_0 + \mathbf{u}\,t\|^3}\;.$$

Restricting our attention to the part of L in the upper half of $\mathcal{P}$, where $t > 0$, there are two possible situations. If $\mathbf{u}^T\mathbf{f}_0 \geq 0$, then $\frac{ds}{dt} > 0$ and s is strictly increasing with t. If $\mathbf{u}^T\mathbf{f}_0 < 0$, then s is strictly increasing ($\frac{ds}{dt} > 0$) up to the critical $t = -\mathbf{u}^T\mathbf{f}_0/\|\mathbf{f}_0\|^2$ and strictly decreasing ($\frac{ds}{dt} < 0$) thereafter. Apparently s achieves a maximum value at the critical t. (In either case, $s \to \mathbf{r}_S^T\mathbf{u}$ as $t \to \infty$.)

Putting together the constant and varying $\bar{r}$ cases we see that the Sharpe ratio, S, has at most one critical point on any part of a line in the upper half of $\mathcal{P}$, and, if it does have a critical point, S achieves its maximum value at this point. The theorem, as stated, follows immediately. □

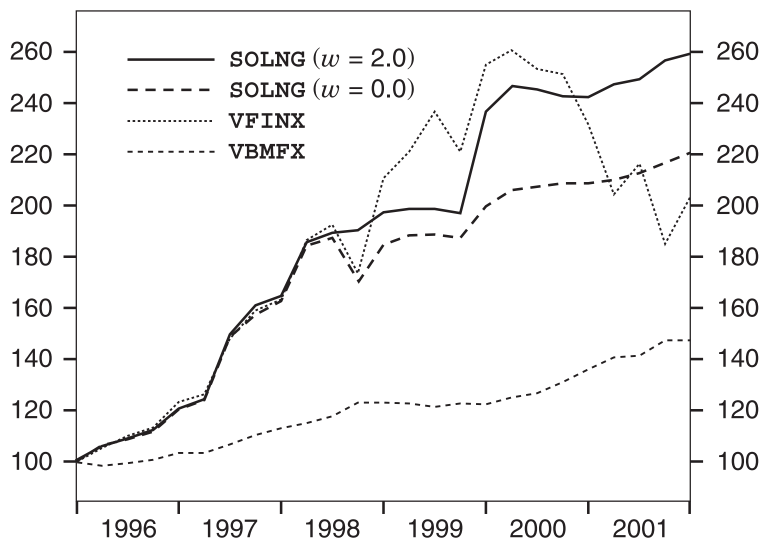
10. An investment strategy

The Sharpe-optimal long portfolio is, by definition, the long portfolio of highest expected reward per unit of risk. It is an *ex post* portfolio of risky funds that is determined by a range of successive periods, and a choice of weight system and svd-cutoff. However, if no risky fund has positive expected reward over the time span under consideration, then no long portfolio of risky funds has positive expected reward either. In this case we will say that the underlying money market fund is the Sharpe-optimal long portfolio (though the money market fund is neither a portfolio of risky funds, nor does it optimize the Sharpe ratio in any way [since it doesn't even have a Sharpe ratio]).

A natural investment strategy immediately suggests itself. Fix on a set of risky funds. At the end of each period, use the total return data for the past m periods to compute the Sharpe-optimal long portfolio for that time span. Then reinvest all your funds in this portfolio for the next period. When that period is over, repeat the process.

For example, fix on the four Vanguard index funds `VIGRX`, `VIVAX`, `VEXMX`, and `VBMFX`. (There is no reason to include the 500 Index, `VFINX`, or the Total Stock Market Index, VTSMX, since these funds are effectively long portfolios of the other stock funds.) Here is a picture of how our investment strategy would have fared over the six-year period 1996–2001. We show the total returns of the *ex ante* SOLNG portfolios corresponding to the $w = 0.0$ and $w = 2.0$ weight systems. (The svd-cutoff, ε, is irrelevant here. Our four risky funds are substantially independent.) The total returns of VFINX and VBMFX are also shown, for comparison.

Total Return Comparison



4 fund Sharpe-optimal long portfolios
vs
500 Index and Total Bond Market Index

Some observations that are immediate from the picture:

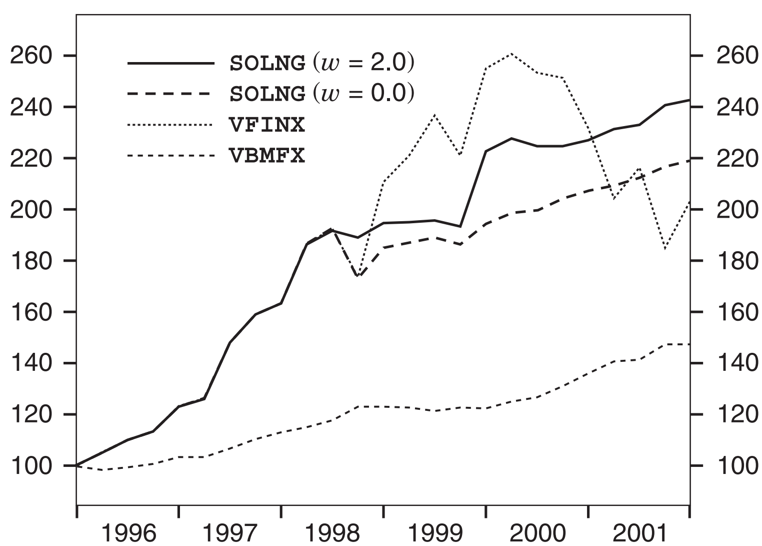
1. In terms of total return, the SOLNG portfolio corresponding to $w = 2.0$ was the best performer for the six-year period. The $w = 0.0$ SOLNG portfolio was also able to outperform the 500 Index Fund, VFINX.
2. Either SOLNG portfolio exhibited substantially less risk (volatility) than the 500 Index over the six-year period. The bond fund, VBMFX, was the least risky—of course.
3. The SOLNG portfolios were able to keep up with VFINX in good times, but, when things got shaky, these portfolios could hang onto their gains.

Table 3 gives the proportions of the four funds in the two *ex post*, 12-quarter SOLNG portfolios. The *ex ante* SOLNG returns can be computed from these proportions and the returns in Table 1. The following table shows the components of the 1998q3 *ex ante* SOLNG returns. In particular it shows why SOLNG($w = 2.0$) suddenly diverged from SOLNG($w = 0.0$) and VFINX in this quarter. Because SOLNG($w = 2.0$) was $\frac{2}{3}$ bond at that time, that's why.

Computing quarterly returns of *ex ante* SOLNG portfolios						
	VIGRX	VIVAX	VEXMX	VBMFX	*ex ante* SOLNG $w = 0.0$	*ex ante* SOLNG $w = 2.0$
ex post SOLNG, 1998q2 $w = 0.0$	63.4%	36.6%	0.0%	0.0%		
ex post SOLNG, 1998q2 $w = 2.0$	33.3%	0.0%	0.0%	66.7%		
1998q3 returns	−7.21	−12.96	−18.7	4.13	−9.31	0.35

You don't need to go to SOLNG portfolios of four risky funds to beat the 500 Index. Over the 1996–2001 period both the $w = 0.0$ and the $w = 2.0$ *ex ante* SOLNG portfolios in just two funds, VFINX and VBMFX, would have done the trick, as the following picture shows.

Total Return Comparison



2 fund Sharpe-optimal long portfolios
vs
500 Index and Total Bond Market Index

The effect of weights

The fact that the $w = 2.0$ SOLNG portfolio outperformed the $w = 0.0$ SOLNG portfolio in both of the above pictures is not accidental. If one runs through the weight system range $0 \leq w \leq 4$, $w = 0.0$ generally produces the worst results, and $w = 2.0$ seems to be near optimal. This is shown in the following table of statistics for variously weighted, 12-quarter *ex ante* SOLNG portfolios of the four funds, VIGRX, VIVAX, VEXMX, VBMFX, over the 1996–2001 period.

Performance Statistics			*ex ante* SOLNG portfolios					1996–2001
weight system	0.0	0.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
mean return	13.91	13.97	15.00	16.84	16.82	16.63	16.43	16.25	16.07
mean reward	8.75	8.81	9.84	11.68	11.66	11.47	11.27	11.09	10.91
stdv of reward	10.91	10.39	10.10	11.73	11.74	11.84	11.97	12.10	12.26
Sharpe ratio	0.802	0.848	0.975	0.995	0.993	0.969	0.942	0.917	0.891

Here are the same statistics for the six risky funds in our test group. Look especially at the five stock funds. When compared with the $w = 2.0$ *ex ante* SOLNG portfolio over the period under consideration, for example, their rewards were substantially lower and their risks substantially higher.

Performance Statistics		6 index funds			1996–2001
fund	VFINX	VIGRX	VIVAX	VTSMX	VEXMX	VBMFX
mean return	13.61	14.80	12.26	12.90	12.30	6.60
mean reward	8.45	9.64	7.10	7.74	7.14	1.44
stdv of reward	18.22	22.13	16.67	19.72	27.07	3.53
Sharpe ratio	0.463	0.436	0.426	0.392	0.264	0.406

11. Conclusion

The purpose of this paper has been to present a way—a geometrical and numerical way—of approaching the problem of portfolio choice. Though we have tested our ideas on a limited amount of data, the results have generally been impressive.

Ours is a short-term approach. "Expected rewards" are only expected over the next period. Our method would probably benefit from shorter periods than we have experimented with. We have simply used quarterly data that was freely available on The Web.

The *ex ante* Sharpe-optimal long portfolio shows promise, though it is definitely not tax efficient. Front-end weighting of data is essential here.

We feel that singular value decomposition, with its ability to determine dependencies among data, can play a significant role in portfolio choice. For example, if expected reward is not, effectively, a linear function of risk, this is due to a redundancy in the funds (securities) under consideration and some degree of mispricing. One should try to determine the redundant and mispriced funds, eliminate them, and look for possible alternatives.

Appendix A. Tables

The remainder of this page is purposely empty.

Table 1.	Quarterly Returns 1993–2001						
quarter	VFINX	VIGRX	VIVAX	VTSMX	VEXMX	VBMFX	VMMXX
2001q4	10.65	12.97	7.89	12.32	19.51	−0.08	0.63
2001q3	−14.72	−13.26	−16.26	−15.93	−21.05	4.29	0.90
2001q2	5.82	7.70	4.41	7.47	14.30	0.79	1.14
2001q1	−11.90	−17.50	−6.58	−12.27	−15.74	3.24	1.43
2000q4	−7.81	−16.88	1.70	−10.17	−18.76	3.98	1.41
2000q3	−0.93	−8.75	8.80	0.27	3.79	3.07	1.52
2000q2	−2.61	−1.41	−4.28	−4.39	−8.68	1.48	1.61
2000q1	2.24	4.03	0.17	3.84	9.68	2.42	1.61
1999q4	14.96	20.15	8.96	18.26	29.54	−0.13	1.18
1999q3	−6.25	−3.52	−9.24	−6.42	−5.84	0.79	1.16
1999q2	7.00	3.90	10.75	7.88	12.45	−0.98	1.23
1999q1	4.98	6.90	2.78	3.71	−0.68	−0.43	1.35
1998q4	21.39	24.64	17.50	21.51	22.14	0.31	1.32
1998q3	−9.95	−7.21	−12.96	−12.07	−18.70	4.13	1.33
1998q2	3.29	5.82	0.53	1.84	−2.18	2.36	1.34
1998q1	13.91	16.20	11.52	13.28	11.51	1.56	1.28
1997q4	2.84	3.49	2.18	1.52	−1.42	2.83	1.29
1997q3	7.48	5.83	9.12	9.75	15.08	3.40	1.33
1997q2	17.41	20.23	14.47	16.81	15.73	3.56	1.35
1997q1	2.63	3.54	1.68	0.65	−3.48	−0.62	1.36
1996q4	8.36	7.29	9.47	6.96	3.82	3.12	1.30
1996q3	3.05	3.51	2.55	2.81	2.59	1.79	1.27
1996q2	4.45	6.85	2.08	4.25	4.04	0.60	1.31
1996q1	5.34	4.28	6.34	5.52	6.16	−1.91	1.31
1995q4	6.01	5.65	6.41	4.54	2.48	4.41	1.43
1995q3	7.94	7.94	8.02	8.94	11.53	1.88	1.46
1995q2	9.49	10.33	8.66	9.23	9.00	6.01	1.42
1995q1	9.71	9.73	9.64	9.15	7.40	4.81	1.39
1994q4	−0.05	0.58	−0.79	−0.95	−2.25	0.55	0.76
1994q3	4.86	7.16	2.50	5.62	6.60	0.52	0.90
1994q2	0.40	−0.11	0.97	−0.90	−2.68	−1.01	1.09
1994q1	−3.84	−4.43	−3.32	−3.71	−3.13	−2.71	1.27
1993q4	2.30	4.37	0.40	1.90	1.50	−0.19	0.77
1993q3	2.52	−0.10	5.09	3.66	7.00	2.72	0.72
1993q2	0.43	−2.06	2.79	0.73	1.33	2.70	0.74
1993q1	4.33	−0.58	9.13	3.98	4.04	4.17	0.75

Table 2.	Quarterly Returns 1997q2–2000q1					
quarter	VFINX	VIGRX	VIVAX	VTSMX	VEXMX	VBMFX
	Actual Returns					
2000q1	2.24	4.03	0.17	3.84	9.68	2.42
1999q4	14.96	20.15	8.96	18.26	29.54	−0.13
1999q3	−6.25	−3.52	−9.24	−6.42	−5.84	0.79
1999q2	7.00	3.90	10.75	7.88	12.45	−0.98
1999q1	4.98	6.90	2.78	3.71	−0.68	−0.43
1998q4	21.39	24.64	17.50	21.51	22.14	0.31
1998q3	−9.95	−7.21	−12.96	−12.07	−18.70	4.13
1998q2	3.29	5.82	0.53	1.84	−2.18	2.36
1998q1	13.91	16.20	11.52	13.28	11.51	1.56
1997q4	2.84	3.49	2.18	1.52	−1.42	2.83
1997q3	7.48	5.83	9.12	9.75	15.08	3.40
1997q2	17.41	20.23	14.47	16.81	15.73	3.56
	SVD-Reduced Returns: $w = 0.0$ $\varepsilon = 5.0\%$					
2000q1	2.22	3.98	0.13	3.99	9.64	2.42
1999q4	14.96	20.09	8.91	18.40	29.50	−0.13
1999q3	−6.26	−3.60	−9.32	−6.19	−5.89	0.79
1999q2	6.97	3.75	10.62	8.29	12.35	−0.98
1999q1	4.98	6.93	2.81	3.62	−0.66	−0.43
1998q4	21.34	24.67	17.52	21.51	22.14	0.31
1998q3	−9.94	−7.22	−12.97	−12.06	−18.70	4.13
1998q2	3.29	5.78	0.49	1.96	−2.21	2.36
1998q1	13.96	16.15	11.47	13.35	11.49	1.56
1997q4	2.79	3.42	2.12	1.76	−1.48	2.83
1997q3	7.52	5.96	9.24	9.37	15.17	3.40
1997q2	17.44	20.13	14.38	17.01	15.68	3.56
	SVD-Reduced Returns: $w = 2.0$ $\varepsilon = 5.0\%$					
2000q1	2.22	3.98	0.13	3.98	9.65	2.42
1999q4	14.95	20.06	8.89	18.47	29.49	−0.13
1999q3	−6.27	−3.63	−9.33	−6.14	−5.91	0.79
1999q2	6.98	3.79	10.66	8.17	12.38	−0.98
1999q1	4.98	6.89	2.77	3.74	−0.69	−0.43
1998q4	21.34	24.61	17.47	21.65	22.11	0.31
1998q3	−9.95	−7.26	−13.00	−11.95	−18.73	4.13
1998q2	3.28	5.73	0.46	2.07	−2.23	2.36
1998q1	13.95	16.10	11.44	13.47	11.47	1.56
1997q4	2.79	3.39	2.10	1.82	−1.49	2.83
1997q3	7.53	6.01	9.26	9.27	15.19	3.40
1997q2	17.44	20.08	14.34	17.13	15.66	3.56

Table 3.	Fund proportions for *ex post* Sharpe-optimal long portfolios							
	weight system: $w = 0.0$				weight system: $w = 2.0$			
quarter	VIGRX	VIVAX	VEXMX	VBMFX	VIGRX	VIVAX	VEXMX	VBMFX
2001q4	0.0%	0.0%	7.6%	92.4%	0.0%	0.0%	7.1%	92.9%
2001q3	0.0%	1.7%	7.1%	91.2%	0.0%	0.0%	6.0%	94.0%
2001q2	0.0%	3.5%	6.3%	90.2%	0.0%	0.0%	5.4%	94.6%
2001q1	4.5%	5.8%	0.0%	89.7%	0.0%	3.3%	0.0%	96.7%
2000q4	9.6%	5.7%	0.0%	84.7%	0.0%	10.3%	0.7%	89.0%
2000q3	18.3%	0.0%	0.0%	81.7%	6.1%	0.0%	12.8%	81.1%
2000q2	22.7%	0.0%	0.0%	77.3%	35.6%	0.0%	0.0%	64.4%
2000q1	24.7%	0.0%	0.0%	75.3%	67.8%	0.0%	0.0%	32.2%
1999q4	40.6%	0.0%	0.0%	59.4%	100.0%	0.0%	0.0%	0.0%
1999q3	32.9%	0.0%	0.0%	67.1%	100.0%	0.0%	0.0%	0.0%
1999q2	31.0%	0.0%	0.0%	69.0%	30.3%	0.0%	0.0%	69.7%
1999q1	23.7%	0.0%	0.0%	76.3%	17.6%	0.0%	0.0%	82.4%
1998q4	32.9%	0.0%	0.0%	67.1%	15.7%	0.0%	0.0%	84.3%
1998q3	33.9%	0.0%	0.0%	66.1%	14.1%	0.0%	0.0%	85.9%
1998q2	63.4%	36.6%	0.0%	0.0%	33.3%	0.0%	0.0%	66.7%
1998q1	23.2%	76.8%	0.0%	0.0%	14.4%	52.6%	0.0%	33.0%
1997q4	34.2%	65.8%	0.0%	0.0%	28.5%	71.5%	0.0%	0.0%
1997q3	42.1%	57.9%	0.0%	0.0%	9.0%	91.0%	0.0%	0.0%
1997q2	85.2%	14.8%	0.0%	0.0%	47.6%	52.4%	0.0%	0.0%
1997q1	100.0%	0.0%	0.0%	0.0%	100.0%	0.0%	0.0%	0.0%
1996q4	88.4%	11.6%	0.0%	0.0%	87.6%	12.4%	0.0%	0.0%
1996q3	100.0%	0.0%	0.0%	0.0%	95.6%	4.4%	0.0%	0.0%
1996q2	61.3%	38.7%	0.0%	0.0%	74.9%	25.1%	0.0%	0.0%
1996q1	0.0%	100.0%	0.0%	0.0%	8.2%	91.8%	0.0%	0.0%
1995q4	2.6%	97.4%	0.0%	0.0%	12.5%	87.5%	0.0%	0.0%

Appendix B. Mathematical tools

We list some mathematical tools used in the development of this paper. More complete information can be found in Golub and Van Loan [1989].

Orthogonal (Perpendicular) Projection. *Suppose* $U \in \mathbf{R}^{m \times n}$, $n \leq m$, *has orthonormal columns* ($U^T U = I_n$). *Then*

$$UU^T \quad \textit{and} \quad I_m - UU^T$$

represent the orthogonal projections onto range(U) *and its orthogonal complement, respectively. More generally, if* $A \in \mathbf{R}^{m \times n}$ *has rank* n, *then*

$$A(A^T A)^{-1} A^T \quad \textit{and} \quad I_m - A(A^T A)^{-1} A^T$$

represent the orthogonal projections onto range(A) *and its orthogonal complement, respectively.*

Householder Reflection. An $m \times m$ matrix of form

$$H = I_m - 2\frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}}, \qquad \mathbf{v} \in \mathbf{R}^m, \quad \mathbf{v} \neq \mathbf{0}$$

is called a *Householder reflection.* It represents the reflection of $\mathbf{R}^m$ though the hyperplane $\mathbf{v}^T\mathbf{x} = 0$. Note in passing that the "half-way" reflection

$$H = I_m - \frac{\mathbf{v}\mathbf{v}^T}{\mathbf{v}^T\mathbf{v}}$$

is just the perpendicular projection of $\mathbf{R}^m$ onto the hyperplane $\mathbf{v}^T\mathbf{x} = 0$.

Singular Value Decomposition (SVD). *If* A *is a real* $m \times n$ *matrix then there exist orthogonal matrices*

$$U = [\mathbf{u}_1, \ldots, \mathbf{u}_m] \in \mathbf{R}^{m \times m} \quad \textit{and} \quad V = [\mathbf{v}_1, \ldots, \mathbf{v}_n] \in \mathbf{R}^{n \times n}$$

such that

$$A = U\Sigma V^T, \qquad \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbf{R}^{m \times n}, \qquad p = \min(m, n),$$

with $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_p \geq 0$.

The σ_i are the *singular values* of A, and the vectors $\mathbf{u}_i$ and $\mathbf{v}_i$ are the corresponding left and right *singular vectors.* The first singular value, σ_1, is the 2-norm of A, $\sigma_1 =$

$\sup\{\|A\mathbf{x}\| : \|\mathbf{x}\| = 1\}$. More generally the singular values σ_i are the lengths of the semi-axes of the hyperellipsoid $E = \{A\mathbf{x} : \|\mathbf{x}\| = 1\}$, and the left singular vectors $\mathbf{u}_i$ point down the corresponding axes.

If A has rank r, then all but the first r singular values are zero. In this case the first r left singular vectors, $\mathbf{u}_i$ $(i = 1, \ldots, r)$, form an orthonormal basis for the range of A, and the last $n - r$ right singular vectors, $\mathbf{v}_i$ $(i = r + 1, \ldots, n)$, form an orthonormal basis for the null space of A.

Many matrix computation packages have an SVD routine. In MATLAB, the package we have used for this paper, one simply writes `[U, S, V] = svd(A)` to get the matrices U, Σ (= `S`), and V.

We use singular value decomposition for dimension reduction (among other things). Typically $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$ will be a matrix of "deviation vectors" with "total variance"

$$v = \|\mathbf{a}_1\|^2 + \cdots + \|\mathbf{a}_n\|^2 = \sigma_1^2 + \cdots + \sigma_p^2.$$

To eliminate noise and to enhance numerical stability we throw away singular values σ_i that do not exceed some relative threshold value $\varepsilon \times \sigma_1$. That is to say, we replace A with

$$A := [\mathbf{u}_1, \ldots, \mathbf{u}_d][\mathbf{v}_1\sigma_1, \ldots, \mathbf{v}_d\sigma_d]^T,$$

where $\sigma_d > \varepsilon \times \sigma_1 \geq \sigma_{d+1}$. (We have often used $\varepsilon = 0.05 = 5.0\%$ for this paper.)

The components $\mathbf{u}_i^T A = \sigma_i \mathbf{v}_i^T$ of A, relative to the orthonormal set $\{\mathbf{u}_i : i = 1, \ldots, d\}$, are often called the "principal components" of A. These components "explain" $100 \times \frac{1}{v}\sum_{i=1}^{d} \sigma_i^2$ percent of the total variance v of the original A.

Pseudo-Inverse. Given a real $m \times n$ matrix A with singular value decomposition $A = U\Sigma V^T$ as above, the *pseudo-inverse* A^+ of A is the $n \times m$ matrix

$$A^+ = V\Sigma^+ U^T,$$

with

$$\Sigma^+ = \mathrm{diag}(\sigma_1^+, \ldots, \sigma_p^+) \in \mathbf{R}^{n\times m}, \qquad \sigma_i^+ = \begin{cases} \sigma_i^{-1} & if\ \sigma_i > 0 \\ 0 & otherwise \end{cases}$$

The pseudo-inverse A^+ can be characterized as the unique matrix $X \in \mathbf{R}^{n\times m}$ that satisfies the *Moore-Penrose conditions:*

$$\begin{array}{llll} \text{(i)} & AXA = A & \text{(iii)} & (AX)^T = AX \\ \text{(ii)} & XAX = X & \text{(iv)} & (XA)^T = XA \end{array}$$

If $m = n$ and A is nonsingular, then $A^+ = A^{-1}$. More generally, $A^+ = (A^T A)^{-1} A^T$ when A has rank n.

For $\mathbf{y} \in \mathbf{R}^m$, $\mathbf{x} = A^+\mathbf{y} \in \mathbf{R}^n$ is the unique $\mathbf{x}$ of minimal norm that minimizes $\|A\mathbf{x} - \mathbf{y}\|$. In MATLAB one simply writes `x = A\y`.

Appendix C. Mathematical results

We are interested in affine dependencies among data points. Six points in general position span a 5-flat, but when do they almost lie in a three-dimensional sub-flat—and which sub-flat? The following propositions answer this kind of question.

Proposition 1. *Suppose* $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n] \in \mathbf{R}^{m\times n}$ *has singular value decomposition* $A = U\Sigma V^T$ *with*

$$U = [\mathbf{u}_1, \ldots, \mathbf{u}_m] \in \mathbf{R}^{m\times m}, \quad \Sigma = \mathrm{diag}(\sigma_1, \ldots, \sigma_p) \in \mathbf{R}^{m\times n}, \quad p = \min(m, n).$$

For $d = 1, \ldots, p$, *set* $U_d = [\mathbf{u}_1, \ldots, \mathbf{u}_d]$. *Then, for each* $d = 1, \ldots, p$,

$$\sum_{i=1}^{n} \|U_d^T \mathbf{a}_i\|^2 = \sigma_1^2 + \cdots + \sigma_d^2 = \max\left\{\sum_{i=1}^{n} \|W_d^T \mathbf{a}_i\|^2 : W_d \in \mathbf{R}^{m\times d}, \quad W_d^T W_d = I_d\right\}.$$

Proof. Set $\Sigma_d = diag(\sigma_1, \ldots, \sigma_d) \in \mathbf{R}^{d\times n}$. Then

$$\begin{aligned}
\textstyle\sum_{i=1}^{n} \|U_d^T \mathbf{a}_i\|^2 &= \mathrm{trace}(A^T U_d U_d^T A) \\
&= \mathrm{trace}(V\Sigma^T U^T U_d U_d^T U \Sigma V^T) \\
&= \mathrm{trace}(\Sigma^T U^T U_d U_d^T U \Sigma) \\
&= \mathrm{trace}(\Sigma_d^T \Sigma_d) \\
&= \sigma_1^2 + \cdots + \sigma_d^2 \ .
\end{aligned}$$

This proves the first part. As for the second, assume $W_d \in \mathbf{R}^{m\times d}$ with $W_d^T W_d = I_d$. Then

$$\begin{aligned}
\textstyle\sum_{i=1}^{n} \|W_d^T \mathbf{a}_i\|^2 &= \mathrm{trace}(A^T W_d W_d^T A) \\
&= \mathrm{trace}(V\Sigma^T U^T W_d W_d^T U \Sigma V^T) \\
&= \mathrm{trace}(\Sigma^T U^T W_d W_d^T U \Sigma) \\
&= \textstyle\sum_{i=1}^{n} \|W_d^T \mathbf{u}_i\|^2 \sigma_i^2 \\
&= \textstyle\sum_{i=1}^{d} \sigma_i^2 - \left\{\sum_{i=1}^{d} \left(1 - \|W_d^T \mathbf{u}_i\|^2\right) \sigma_i^2 - \sum_{i=d+1}^{n} \|W_d^T \mathbf{u}_i\|^2 \sigma_i^2\right\} \\
&\leq \textstyle\sum_{i=1}^{d} \sigma_i^2 - \left\{\sum_{i=1}^{d} \left(1 - \|W_d^T \mathbf{u}_i\|^2\right) \sigma_d^2 - \sum_{i=d+1}^{n} \|W_d^T \mathbf{u}_i\|^2 \sigma_d^2\right\} \\
&= \textstyle\sum_{i=1}^{d} \sigma_i^2 - \left(d - \sum_{i=1}^{n} \|W_d^T \mathbf{u}_i\|^2\right) \sigma_d^2 \\
&= \textstyle\sum_{i=1}^{d} \sigma_i^2 \ ,
\end{aligned}$$

the last equation since

$$\begin{aligned}\sum_{i=1}^{n} \|W_d^T \mathbf{u}_i\|^2 &= \operatorname{trace}(U^T W_d W_d^T U) \\ &= \operatorname{trace}(W_d W_d^T) \\ &= \operatorname{trace}(W_d^T W_d) \\ &= \operatorname{trace}(I_d) \\ &= d\ . \quad \square\end{aligned}$$

In this proof we have used the properties

$$\operatorname{trace}(B^{-1}AB) = \operatorname{trace}(A) \qquad \operatorname{trace}(AA^T) = \operatorname{trace}(A^T A)$$

of the trace function and the fact that $\|W_d^T \mathbf{u}_i\| \le 1$ when $\|W_d\| = \|\mathbf{u}_i\| = 1$.

Flats

By a *flat* in $\mathbf{R}^m$ we simply mean an affine subspace of $\mathbf{R}^m$. When the dimension d is specified, we speak of a *d-flat*. Thus a 0-flat is a point, a 1-flat a line, a 2-flat a plane, and so forth.

Problem. Given a collection of points $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in $\mathbf{R}^m$ and a dimension d, find the d-flat that is closest to the points in the sense that the sum of the square distances from the points to the flat is minimal.

The statement of this problem suggests that there is a unique solution. This need not be the case. For example, if all the points lie on a line and $d \ge 2$, then every d-flat containing the line is a solution. Be that as it may, here is one solution to the general problem.

Proposition 2. *Given $\mathbf{x}_1, \ldots, \mathbf{x}_n$ in $\mathbf{R}^m$, set $A = [\mathbf{a}_1, \ldots, \mathbf{a}_n]$ with $\mathbf{a}_i = \mathbf{x}_i - \bar{\mathbf{x}}$ $(i = 1, \ldots, n)$, $\bar{\mathbf{x}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i$. Let $A = U\Sigma V^T$ (as in Proposition 1) be the singular value decomposition of A, and, for $d = 1, \ldots, m$, put $U_d = [\mathbf{u}_1, \ldots, \mathbf{u}_d]$. Then $\bar{\mathbf{x}} + \operatorname{range}(U_d)$ is the d-flat closest to $\mathbf{x}_1, \ldots, \mathbf{x}_n$. If $d < p = \min(m, n)$, the sum of square distances of the $\mathbf{x}_i$ from this flat is $\sigma_{d+1}^2 + \cdots + \sigma_p^2$; otherwise the flat contains all the $\mathbf{x}_i$ and the sum of square distances is zero.*

Proof. We can assume that $d < \operatorname{rank}(A) \le p$ because, if $d \ge \operatorname{rank}(A)$, then $\operatorname{range}(A) \subseteq \operatorname{range}(U_d)$, all the $\mathbf{x}_i$ are contained in $\bar{\mathbf{x}} + \operatorname{range}(U_d)$, and the sum of square distances of the $\mathbf{x}_i$ from this flat is zero.

Choose $\mathbf{x}_0 \in \mathbf{R}^m$ and $W_d \in \mathbf{R}^{m \times d}$ satisfying $W_d^T W_d = I_d$. Then $\mathbf{x}_0 + \operatorname{range}(W_d)$ represents an arbitrary d-flat in $\mathbf{R}^m$, and the sum of square distances of the $\mathbf{x}_i$ from $\mathbf{x}_0 + \operatorname{range}(W_d)$ is

$$v = \sum_{i=1}^{n} \left\|(\mathbf{x}_i - \mathbf{x}_0) - W_d W_d^T (\mathbf{x}_i - \mathbf{x}_0)\right\|^2.$$

This is what we want to minimize.

Substituting $\mathbf{x}_0 = \bar{\mathbf{x}} + \mathbf{y}$ into the expression for v and simplifying,

$$v = \sum_{i=1}^{n} \left\| \mathbf{a}_i - W_d W_d^T \mathbf{a}_i - \mathbf{w} \right\|^2,$$

where $\mathbf{w} = \mathbf{y} - W_d W_d^T \mathbf{y}$. Using the relations $W_d^T W_d = I_d$ and $W_d^T \mathbf{w} = \mathbf{0}$,

$$v = \sum_{i=1}^{n} \left\{ \|\mathbf{a}_i\|^2 - \|W_d^T \mathbf{a}_i\|^2 + \|\mathbf{w}\|^2 - 2\mathbf{w}^T \mathbf{a}_i \right\} .$$

But $\sum_{i=1}^{n} \mathbf{a}_i = \mathbf{0}$. Therefor

$$v = \sum_{i=1}^{n} \|\mathbf{a}_i\|^2 - \sum_{i=1}^{n} \|W_d^T \mathbf{a}_i\|^2 + n\|\mathbf{w}\|^2.$$

Looking at the last expression for v it is apparent that v will be minimized by choosing a W_d that maximizes $\sum_{i=1}^{n} \|W_d^T \mathbf{a}_i\|^2$ and a point $\mathbf{x}_0$ that makes $\mathbf{w} = \mathbf{0}$. By Proposition 1, $W_d = U_d$ accomplishes the first requirement, and $\mathbf{x}_0 = \bar{\mathbf{x}}$ accomplishes the second. Thus the minimal v is

$$v = \sum_{i=1}^{n} \|\mathbf{a}_i\|^2 - \sum_{i=1}^{n} \|U_d^T \mathbf{a}_i\|^2 = \sum_{i=1}^{p} \sigma_i^2 - \sum_{i=1}^{d} \sigma_i^2 = \sum_{i=d+1}^{p} \sigma_i^2. \quad \square$$

References

GOLUB, GENE H., AND VAN LOAN, CHARLES F., *Matrix Computations, 2d ed.,* Johns Hopkins University Press, Baltimore, 1989.

HILBERT, DAVID, AND COHN-VOSSEN, S., *Geometry and the Imagination,* Chelsea, New York, 1952.

MARKOWITZ, HARRY M., *Mean–Variance Analysis in Portfolio Choice and Capital Markets,* Blackwell, Oxford and Cambridge, Mass., 1987.

SHARPE, WILLIAM F., The Sharpe Ratio, *Journal of Portfolio Management,* **21** (Fall 1994), 49–58.

SIGMON, KERMIT, *MATLAB Primer, 5th ed.,* CRC Press, Boca Raton, Florida, 1998.

TWAIN, MARK, *Life on the Mississippi,* Harper, New York and London, 1917.