MISCELLANY

BRAD BAXTER

Contents

1. Introduction

This is my collection of miscellanies, most of which are for teaching.

Version: 202401251231

2. Roots of Unity

Example 2.1. *Let* $\omega = e^{2\pi i/3}$. *Thus* $1, \omega, \omega^2$ *are the three cube roots of unity. Then it is easily checked that* $|1-\omega| = |1-\omega^2| = \sqrt{3}$, *so that*

$$|1-\omega||1-\omega^2| = 3.$$

Example 2.2. *Suppose we take the* 4 *points* ± 1 *and* $\pm i$. *Then* $|1-i| = \sqrt{2}$, *so that*

$$|1-i||1-(-1)||1-(-i)| = 4.$$

These examples lead to a conjecture:

$$\prod_{k=1}^{n-1} |1-\omega^k| = n, \tag{1}$$

where $\omega = e^{2\pi i/n}$ and $n \ge 2$, and here is the Matlab code to check this.

```
I=sqrt(-1);
n=5;omega=exp(2*pi*I/n); P=1; for k=1:n-1, P=P*abs(1-omega^k); end; P
```

In fact, we shall see that a stronger statement is true:

Let $n > 1$ be an integer and let $\omega = e^{2\pi i/n}$. Thus the complex numbers $\{\omega^k : k = 0, 1, \ldots, n-1\}$ are the nth roots of unity. Thus

$$z^n - 1 = (z-1)\prod_{k=1}^{n-1}\left(z - \omega^k\right).$$

Hence

$$\prod_{k=1}^{n-1}\left(1-\omega^k\right) = \lim_{z\to 1}\frac{z^n-1}{z-1} = n,$$

by de L'Hôpital's rule.

3. Stereographic Projection

3.1. **Stereographic Projection as a Möbius mapping.** We let $S\colon \mathbb{R} \to \partial\Delta$ denote stereographic projection from the real line to the unit circle, with i as the pole. Thus given $t \in \mathbb{R}$, the correponding point $(x, y) \in \partial\Delta$ satisfies

$$0 = \det \begin{pmatrix} 1 & 0 & 1 \\ 1 & x & y \\ 1 & t & 0 \end{pmatrix} = -yt + t - x, \tag{2}$$

or

$$x = t(1-y). \tag{3}$$

Theorem 3.1. *Stereographic projection $S\colon \mathbb{R} \to \partial\Delta$ is given by the Möbius map*

$$S(t) = \frac{it+1}{t+i} = \frac{t-i}{-it+1}, \qquad \text{for } t \in \mathbb{R}. \tag{4}$$

Proof. Squaring (3) and using $x^2 + y^2 = 1$, we obtain the quadratic

$$1 - y^2 = x^2 = t^2 \left(1-y\right)^2,$$

that is,

$$0 = \left(1+t^2\right) y^2 - 2t^2 y^2 + t^2 - 1. \tag{5}$$

The roots of this quadratic satisfy

$$y = \frac{t^2 \pm [t^4 - (t^2+1)(t^2-1)]^{1/2}}{t^2+1} = 1 \quad \text{or} \quad \frac{t^2-1}{t^2+1}.$$

Substituting $y = (t^2-1)/(t^2+1)$ in (3) yields

$$x = \frac{2t}{t^2+1}. \tag{6}$$

Thus stereographic projection is given by

$$S(t) = \frac{2t}{t^2+1} + i\left(\frac{t^2-1}{t^2+1}\right) = \frac{it^2+2t-i}{t^2+1}.$$

Numerator and denominator vanish when $t = i$, and dividing by the common factor $t - i$ yields

$$S(t) = \frac{it+1}{t+i}, \qquad t \in \mathbb{R}.$$

□

By Theorem 13.9.1 of [1], S corresponds to the 90^o rotation of the Riemann sphere, with axis ±1, sending i to 0.

3.1.1. *Inversion in the Origin.* Let $T(z) = -z$. We want to compute the induced map $\tilde{T} = S^{-1}TS$, to see its action on $\mathbb{R}$. The corresponding 2×2 matrices are

$$S \sim \begin{pmatrix} i & 1 \\ 1 & i \end{pmatrix}, \quad S^{-1} \sim \begin{pmatrix} i & -1 \\ -1 & i \end{pmatrix} \quad \text{and} \quad T \sim \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix}.$$

Thus

$$\tilde{T} = S^{-1}TS \sim \begin{pmatrix} i & -1 \\ -1 & i \end{pmatrix} \begin{pmatrix} -1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} i & 1 \\ 1 & i \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

In other words, $z\tilde{T}(z) = -1$.

3.2. **Quasi-inversion in** ia**.** Let $a \in (-1, 1)$. Given any point $u_1 = S(t_1) \in \partial\Delta$, we want the corresponding point $u_2 = S(t_2) \in \partial\Delta$ for which u_1, u_2 and ia are collinear. Thus we have the equation

$$0 = \det \begin{pmatrix} 1 & 0 & a \\ 1 & \frac{2t_1}{1+t_1^2} & \frac{t_1^2-1}{t_1^2+1} \\ 1 & \frac{2t_2}{1+t_2^2} & \frac{t_2^2-1}{t_2^2+1}. \end{pmatrix},$$

i.e.

$$0 = \det \begin{pmatrix} 1 & 0 & a \\ 1+t_1^2 & 2t_1 & t_1^2-1 \\ 1+t_2^2 & 2t_2 & t_2^2-1. \end{pmatrix}.$$

Subtracting the second row from the third row and factorizing, we obtain

$$0 = 2(t_2 - t_1) \det \begin{pmatrix} 1 & 0 & a \\ 1+t_1^2 & t_1 & t_1^2-1 \\ t_1+t_2 & t_2 & t_1+t_2. \end{pmatrix} = 2(t_2-t_1)(t_1t_2(1-a)+1+a).$$

Thus the corresponding map is

$$t_1t_2 = \frac{a+1}{a-1}.$$

4. The Length of the Day

4.1. **The Length of the Day at the Solstices.** We shall compute the length of the day at the Summer solstice in the northern hemisphere. The origin of our coordinate system will be at the centre of the Earth, the x-axis will point directly towards the Sun, and the z-axis will be perpendicular to the Earth's orbital plane and will be directed into the northern hemisphere. We need the following orthonormal vectors to describe the motion of a point on the Earth's surface:

$$\mathbf{u}_1 = \begin{pmatrix} \cos\alpha \\ 0 \\ -\sin\alpha \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \text{ and } \mathbf{u}_3 = \begin{pmatrix} \sin\alpha \\ 0 \\ \cos\alpha \end{pmatrix}, \tag{7}$$

where $\alpha = 23.5$ degrees approximately for the Earth.

The motion of a point at latitude θ is then

$$\begin{pmatrix} x(t) \\ y(t) \\ z(t) \end{pmatrix} = (\mathbf{u}_1 \cos t + \mathbf{u}_2 \sin t)\cos\theta + \sin\theta \mathbf{u}_3. \tag{8}$$

In particular, we have

$$\begin{aligned} x(t) &= \cos\theta\cos\alpha\cos t + \sin\theta\sin\alpha \\ y(t) &= \cos\theta\sin t. \end{aligned} \tag{9}$$

At the Summer solstice, day corresponds to $x(t) > 0$. Thus, solving $x(t) = 0$, we obtain the length of day as a function of the latitude θ:

$$L(\theta) = 2\cos^{-1}(-\tan\theta\tan\alpha), \quad |\theta| \le 90 - \alpha, \tag{10}$$

and this gives the length of the day in *degrees*. Thus the length of the day in hours is given by$L_h(\theta) = (24/360)L(\theta) = (1/15)L(\theta)$, i.e.

$$L_h(\theta) = \frac{2}{15}\cos^{-1}(-\tan\theta\tan\alpha), \quad |\theta| \le 90 - \alpha, \tag{11}$$

For $\theta \in (90 - \alpha, 90)$, $L_h(\theta) = 24$; similarly $L(\theta) = 0$ for $\theta \in (-90, -90 + \alpha)$.

The following Matlab code generates the ratio of the longest day to the shortest day.

```
alpha= 23.5*pi/180;
theta=0:pi/100:(pi/2) - alpha;
y = acos(-tan(alpha)*tan(theta));
R = y ./ (pi - y);
plot(theta,R)
plot((180/pi)*theta,R)
grid
```

4.2. **The variation in the length of the day during the year.** We solve the equation

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix}^T \begin{pmatrix} \cos u \\ \sin u \end{pmatrix} = 0, \tag{12}$$

where $0 \le u \le 360$ measures orbital time in degrees, i.e. one year corresponds to 360 degrees. Expanding (12), we obtain

$$\cos\theta\cos\alpha\cos u\cos t + \cos\theta\sin u\sin t = -\sin\theta\sin\alpha\cos u, \tag{13}$$

or

$$\cos(t - \beta) = \frac{\sin\theta\sin\alpha\cos u}{\gamma}, \tag{14}$$

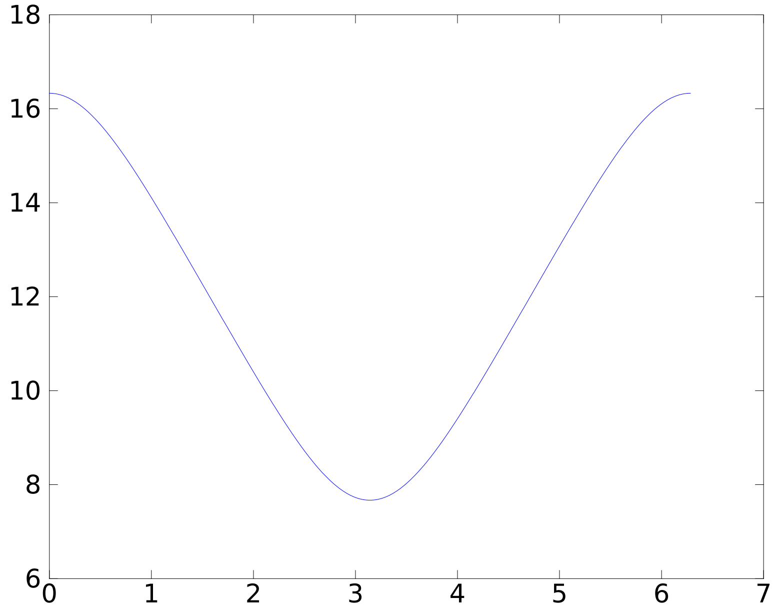

FIGURE 1. Annual variation in day-length at 51 degrees North

where

$$\gamma^2 = \cos^2\theta\cos^2\alpha\cos^2 u + \cos^2\theta\sin^2 u. \tag{15}$$

Hence the sunrise and sunset times are given by

$$t_\pm - \beta = \pm\cos^{-1}\left(\frac{-\sin\theta\sin\alpha\cos u}{\gamma}\right), \tag{16}$$

and the length of the day is then

$$t_+ - t_- = 2\cos^{-1}\left(\frac{-\sin\theta\sin\alpha\cos u}{\gamma}\right), \tag{17}$$

```
%
% Displays the yearly variation in the length of the day
% (in hours) at latitude theta, where |theta| < pi/2 - alpha.
%
alpha= 23.5*pi/180;
theta=51*pi/180;
u=0:pi/1000:2*pi;
A = -sin(theta)*sin(alpha)*cos(u);
B = cos(theta)*sqrt( (cos(alpha)^2)*(cos(u).^2) + (sin(u).^2) );
D = 2*acos(A ./ B)*12/pi;
plot(u,D)
%
% D is quite close to sinusoidal
%
%hold on
%plot(u, 12+(max(D)-12)*cos(u),'r')
%hold off
```

5. Regular Pentagons and the Golden Ratio

Let $\omega = \exp(2\pi i/5)$, so that $1, \omega, \omega^2, \omega^3, \omega^4$ are the fifth roots of unity. The aim here is to prove that the Golden Ratio $\phi = (1+\sqrt{5})/2$ satisfies

$$\phi = 2\cos\pi/5 = \frac{\left|1-\omega^2\right|}{|1-\omega|}. \tag{18}$$

Indeed,

$$\left|1-\omega^k\right|^2 = \left(1-\omega^k\right)\left(1-\omega^{-k}\right) = 2 - 2\cos(2k\pi/5) = 4\sin^2 k\pi/5, \tag{19}$$

for $0 \le k \le 4$. Hence

$$\frac{\left|1-\omega^2\right|}{|1-\omega|} = \frac{\sin 2\pi/5}{\sin\pi/5} = 2\cos\pi/5. \tag{20}$$

Now, setting $\alpha = e^{\pi i/5}$, i.e. the principal tenth root of unity, and

$$\beta = 2\cos\pi/5 = \alpha + \alpha^{-1},$$

we obtain

$$1 + \alpha^2 + \alpha^4 + \alpha^6 + \alpha^8 = \frac{1-\alpha^{10}}{1-\alpha^2} = 0. \tag{21}$$

However

$$\begin{aligned}
1 + \alpha^2 + \alpha^4 + \alpha^6 + \alpha^8 \\
&= 1 + \alpha^2 + \alpha^8 + \alpha^4 + \alpha^6 \\
&= 1 + \alpha^2 + \alpha^{-2} + \alpha^5\left(\alpha^{-1} + \alpha\right) \\
&= 1 + \alpha^2 + \alpha^{-2} - \left(\alpha^{-1} + \alpha\right) \\
&= 1 + \left(\alpha + \alpha^{-1}\right)^2 - 2 - \left(\alpha^{-1} + \alpha\right) \\
&= \beta^2 - \beta - 1.
\end{aligned}$$

Thus $\beta = \phi$.

6. Distance seen and Height

If we take the Earth to be a perfect sphere of radius R, then the distance seen D at height H is given by

$$D = R\theta \tag{22}$$

where

$$(R + H)\cos\theta = R. \tag{23}$$

It's useful to define

$$h = \frac{H}{R}. \tag{24}$$

Thus (22) and (23) become

$$(1 + h)\cos\theta = 1 \tag{25}$$

and

$$\cos\theta = \left[\frac{1}{1+h}\right]. \tag{26}$$

If h is small, then θ must also be small, so we have

$$1 - \theta^2/2 + \cdots = 1 - h + \cdots , \tag{27}$$

or

$$\theta^2 \approx 2h. \tag{28}$$

Returning to our original variables, we find

$$D^2 \approx 2HR. \tag{29}$$

Example 6.1. *Taking $R = 6.4 \times 10^6$ m and $H = 100$ m. Then $\sqrt{2HR} \approx 36$ km.*

Example 6.2. *Here is some MATLAB code to illustrate the approximation's worth.*

```
R=6.4e6; h = 0:100:100000;
dtrue = R*acos((1 + h/R).^(-1));
dapprox = (2*R*h).^(1/2);
```

7. The Railway problem

This is a very old chestnut indeed. We imagine a straight piece of rail of unit length which, under thermal expansion, becomes a circular arc of length $1+\delta$, where $1 \gg \delta > 0$. The rail will bow upwards, attaining a maximum height h at its centre, and the problem is to determine h, which is surprisingly large.

If we let R denote the radius of the circular arc after expansion, and θ denote the half-angle subtended at the centre of the circle, then we have the equations

$$2R\theta = 1+\delta, \tag{30}$$

$$R - h = R\cos\theta, \tag{31}$$

$$\frac{1}{2} = R\sin\theta. \tag{32}$$

Of course

$$h = R\left(1-\cos\theta\right). \tag{33}$$

Eliminating R from (30) and (32), we obtain

$$\frac{\sin\theta}{2\theta} = \frac{1/2}{1+\delta}$$

or

$$\frac{\sin\theta}{\theta} = \frac{1}{1+\delta}. \tag{34}$$

Now $0 < \delta \ll 1$ implies that θ is also small, so that

$$\frac{\sin\theta}{\theta} = 1 - \frac{1}{6}\theta^2 + \cdots = 1 - \delta + \cdots , \tag{35}$$

or

$$\theta^2 \approx 6\delta. \tag{36}$$

Substituting this approximation in (30) yields

$$R = \frac{1+\delta}{2\theta} \approx \frac{1+\delta}{2\sqrt{6\delta}}. \tag{37}$$

Substituing (37) in (33) then provides

$$h \approx R\theta^2/2 \approx \left(\frac{1+\delta}{2\sqrt{6\delta}}\right)\frac{6\delta}{2} = \frac{\sqrt{6\delta}\,(1+\delta)}{4}$$

i.e.

$$h \approx \sqrt{3\delta/8}. \tag{38}$$

This is at the root of the surprising size of h: $\sqrt{\delta}$ dominates δ for small δ.

Example 7.1. *Suppose $\delta = 10^{-4}$, which corresponds to expansion of 10 cm for a rail of length one kilometre. In this case $h = \sqrt{3\times 10^{-4}/8} = 6.1m$.*

8. A derivation of the FFT

We choose $n = 2^M$ and illustrate the Fast Fourier Transform algorithm, which computes the DFT in $O(M2^M)$ operations.

Our primary data are the values $\{f(2\pi j/2^M) : j = 01, 2, \ldots, 2^M - 1\}$ of our function evaluated at the 2^M-th roots of one. For each $m \in \{0, 1, \ldots, M\}$, we define

$$F_{jk}^{(m)} = \sum_{p=0}^{2^m-1} f\left(e^{2\pi i\left(\frac{p}{2^m}+\frac{k}{2^M}\right)}\right) e^{-2\pi ijp/2^m}. \tag{39}$$

for $j = 0, 1, \ldots, 2^m - 1$ and $k = 0, 1, \ldots, 2^{M-m} - 1$. Thus $F^{(m)} \in \mathbb{C}^{2^m \times 2^{M-m}}$. In other words, each $F^{(m)}$ contains 2^M numbers, but their sizes are as follows:

$$\begin{aligned}
F^{(0)} &\text{ is } 1 \times 2^M; \\
F^{(1)} &\text{ is } 2 \times 2^{M-1}; \\
F^{(2)} &\text{ is } 2^2 \times 2^{M-2}; \\
&\vdots \\
F^{(M-1)} &\text{ is } 2^{M-1} \times 2; \\
F^{(M)} &\text{ is } 2^M \times 1.
\end{aligned}$$

In other words, $F^{(0)}$ is a row vector, $F^{(m)}$ has twice the number of rows as $F^{(m-1)}$, but half the number of columns, and $F^{(M)}$ is a column vector.

Example 8.1. *When $M = 3$ and $m = 2$, there are 2 4-transforms.*

Example 8.2. *When $M = 3$ and $m = 1$, there are 4 2-transforms.*

We now define a mapping constructing $F^{(m)}$ from $F^{(m-1)}$. Specifically, we divide the sum over p in (39) into even p and odd p, as follows

$$F_{jk}^{(m)} = E_m + O_m, \tag{40}$$

where

$$E_m = \sum_{q=0}^{2^{m-1}-1} f(\exp(2\pi i\left(\frac{q}{2^{m-1}} + \frac{k}{2^M}\right) \exp(-2\pi ijq/2^{m-1}) \tag{41}$$

and

$$O_m = \sum_{r=0}^{2^{m-1}-1} \left(f(\exp(2\pi i\left(\frac{r}{2^{m-1}} + \frac{k+2^{M-m}}{2^M}\right) \exp(-2\pi ijr/2^{m-1})\right) \exp(-\pi ij/2^{m-1}). \tag{42}$$

Now $F^{(m-1)}$ is a $2^{m-1} \times 2^{M-m+1}$ matrix, but it is useful to slightly abuse notation noting that $\mathbb{Z} \ni j \mapsto F_{jk}^{(m-1)}$ is a 2^{m-1}-periodic sequence. With this abuse of notation in mind, we obtain

$$F_{jk}^{(m)} = F_{jk}^{(m-1)} + e^{-\pi ij/2^{m-1}} F_{j,k+2^{M-m}}^{(m-1)}, \tag{43}$$

for $j = 0, 1, \ldots, 2^m - 1$, $k = 0, 1, \ldots, 2^{M-m} - 1$.

9. Schur Products

In this note we provide a (possibly) new slant on a theorem of I. Schur.

It is well known that a self-adjoint matrix $A \in \mathbb{C}^{n\times n}$ is non-negative definite if and only if there are elements $(f_k)_{k=1}^n$ in a Hilbert space H for which

$$A_{jk} = (f_j, f_k)_H, \qquad j, k = 1, \ldots, n,$$

where $(\cdot,\cdot)_H$ denotes the inner product of the Hilbert space. Of course the usual choice is $H = \mathbb{C}^n$, but we shall take $H = L^2(\mathbb{T})$, the vector space of 2π-periodic square-integrable functions. Thus the elements of any self-adjoint, non-negative definite matrix $A \in \mathbb{C}^{n\times n}$ can be expressed by the equations

$$A_{jk} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_j(x)\overline{f_k(x)}\,dx, \qquad j, k = 1, \ldots, n.$$

Let us now recall that the *Schur product* $A * B$ of any two complex $n \times n$ matrices A and B is defined by $(A * B)_{jk} := A_{jk}B_{jk}$. We now state our main result.

Theorem 9.1. *The Schur product of non-negative definite self-adjoint matrices is also non-negative definite.*

Proof. Following the remarks above, we have

$$A_{jk} = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_j(x)\overline{f_k(x)}\,dx \quad \text{and} \quad B_{jk} = \frac{1}{2\pi}\int_{-\pi}^{\pi} g_j(y)\overline{g_k(y)}\,dy, \qquad j, k = 1, \ldots, n,$$

whenever A and B are non-negative definite and self-adjoint. Thus the elements of the Schur product $A * B$ are inner products in $L^2([-\pi,\pi]^2)$ of tensor products of the functions $(f_j)_1^n$ and $(g_k)_1^n$. Specifically, we have

$$(A * B)_{jk} = (2\pi)^{-2}\int_{\mathbb{T}^2} f_j \otimes g_j(x,y)\overline{f_k \otimes g_k(x,y)}\,dx\,dy, \qquad j, k = 1, \ldots, n,$$

where $f_j \otimes g_j(x,y) := f_j(x)g_k(y)$. Hence $A * B$ is non-negative definite. $\square$

10. Constrained Optimization

Suppose we are considering investing money in two assets whose returns are independent random variables X_1 and X_2. Their distribution is unknown, but we do know the mean $\mu_k = \mathbb{E}X_k$ and the variance $\sigma_k^2 = \operatorname{var} X_k$, for $k = 1, 2$, and we shall assume that these variances are strictly positive.

Being risk-averse, we want to divide our investment between the two assets to minimize our risk. More formally, we have

$$Y = s_1 X_1 + s_2 X_2, \qquad \text{where } s_1 + s_2 = 1. \tag{44}$$

Now, by the independence of X_1 and X_2, we have

$$\operatorname{var} Y = f(\mathbf{s}) = s_1^2\sigma_1^2 + s_2^2\sigma_2^2, \qquad \mathbf{s} = (s_1, s_2)^T \in \mathbb{R}^2. \tag{45}$$

Thus our problem is as follows:

$$\begin{array}{ll} \text{minimize} & f(\mathbf{s}) \\ \text{subject to} & g(\mathbf{s}) = 1, \end{array} \tag{46}$$

where

$$g(\mathbf{s}) = s_1 + s_2, \quad \mathbf{s} = (s_1, s_2)^T \in \mathbb{R}^2. \tag{47}$$

Now the function $f(\mathbf{s})$ satisfies

$$\nabla f(\mathbf{s}) = \begin{pmatrix} 2\sigma_1^2 s_1 \\ 2\sigma_2^2 s_2 \end{pmatrix}$$

and

$$D^2 f(\mathbf{s}) = \begin{pmatrix} 2\sigma_1^2 & 0 \\ 0 & 2\sigma_2^2 \end{pmatrix},$$

and all higher derivatives vanish. In other words, $f(\mathbf{s})$ is a quadratic and satisfies

$$f(\mathbf{s} + \mathbf{h}) = f(\mathbf{s}) + \mathbf{h}^T \nabla f(\mathbf{s}) + \frac{1}{2}\mathbf{h}^T D^2 f(\mathbf{s})\mathbf{h}. \tag{48}$$

Further, the constraint function $g(\mathbf{s})$ is linear and satisfies

$$g(\mathbf{s} + \mathbf{h}) = g(\mathbf{s}) + \mathbf{h}^T \nabla g(\mathbf{s}) = g(\mathbf{s}) + \mathbf{h}^T \mathbf{e}, \tag{49}$$

where

$$\nabla g(\mathbf{s}) \equiv \mathbf{e} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{50}$$

One way to understand such problems is via *line search*: we choose a point $\mathbf{s} \in \mathbb{R}^2$ and a search direction $\mathbf{d} \in \mathbb{R}^2$ and consider the univariate function

$$\phi(t) = f(\mathbf{s} + t\mathbf{d}), \qquad t \in \mathbb{R}. \tag{51}$$

Thus

$$\phi(t) = f(\mathbf{s}) + t\mathbf{d}^T \nabla f(\mathbf{s}) + \frac{1}{2}t^2 \mathbf{d}^T D^2 f(\mathbf{s})\mathbf{d}, \tag{52}$$

but we also require the search direction to satisfy the linear constraint:

$$1 = g(\mathbf{s} + t\mathbf{d}) = g(\mathbf{s}) + t\mathbf{d}^T \nabla g(\mathbf{s}) = 1 + t\mathbf{d}^T \nabla g(\mathbf{s}), \tag{53}$$

or

$$\mathbf{d}^T \nabla g(\mathbf{s}) = 0 \tag{54}$$

When do we know we are at a minimum? In this case, we must have $\phi'(0) = 0$ for any $\mathbf{d}$ satisfying (54). Hence

$$\mathbf{d}^T \nabla f(\mathbf{s}) = \mathbf{d}^T \nabla g(\mathbf{s}) = 0, \tag{55}$$

which implies that

$$\nabla f(\mathbf{s}) = \lambda \nabla g(\mathbf{s}), \tag{56}$$

for some $\lambda \in \mathbb{R}$. In other words, we have

$$\begin{pmatrix} 2\sigma_1^2 s_1 \\ 2\sigma_2^2 s_2 \end{pmatrix} = \lambda \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \tag{57}$$

which imply that

$$s_k = \frac{1}{2}\lambda\sigma_k^{-2}, \quad k = 1, 2, \quad \text{and } s_1 + s_2 = 1. \tag{58}$$

Thus

$$\lambda = \frac{2}{\sigma_1^{-2} + \sigma_2^{-2}}$$

and

$$s_k = \frac{\sigma_k^{-2}}{\sigma_1^{-2} + \sigma_2^{-2}}, \quad k = 1, 2. \tag{59}$$

The resulting minimal variance is then given by

$$\begin{aligned} \sigma^2 &\equiv f(\mathbf{s}) \\ &= \sigma_1^2 \frac{\sigma_1^{-4}}{\left(\sigma_1^{-2} + \sigma_2^{-2}\right)^2} + \sigma_2^2 \frac{\sigma_2^{-4}}{\left(\sigma_1^{-2} + \sigma_2^{-2}\right)^2} \\ &= \frac{1}{\sigma_1^{-2} + \sigma_2^{-2}}, \end{aligned}$$

or

$$\sigma^{-2} = \sigma_1^{-2} + \sigma_2^{-2}. \tag{60}$$

Example 10.1. *When $\sigma_1^2 = 1/10$ and $\sigma_2^2 = 1/5$, the minimal variance is given by $\sigma^{-2} = 10 + 5 = 15$, or $\sigma^2 = 1/15$.*

In general, we have

$$Y = s_1 X_1 + s_2 X_2 = \frac{\sigma_1^{-2} X_1 + \sigma_2^{-2} X_2}{\sigma_1^{-2} + \sigma_2^{-2}}$$

and

$$\mathbb{E}Y = s_1 \mu_1 + s_2 \mu_2 = \frac{\sigma_1^{-2}\mu_1 + \sigma_2^{-2}\mu_2}{\sigma_1^{-2} + \sigma_2^{-2}}$$

Thus, if $\sigma_1 \gg \sigma_2$, then $\mathbb{E}Y \approx \mu_2$, which is to be expected, whilst $\sigma_1 = \sigma_2$ implies $\mathbb{E}Y = (\mu_1 + \mu_2)/2$.

10.1. **Lagrange Multipliers.** The above technique is much more general. Suppose we have a risk-metric for investments in n assets which is given by

$$f(\mathbf{s}) = \mathbf{s}^T A \mathbf{s}, \qquad \mathbf{s} \in \mathbb{R}^n, \tag{61}$$

where $A \in \mathbb{R}^{n\times n}$ is a symmetric, positive definite matrix. We want to solve the constrained optimization problem

$$\begin{array}{ll} \text{minimize} & f(\mathbf{s}) \\ \text{subject to} & g(\mathbf{s}) = 1, \end{array} \tag{62}$$

where

$$g(\mathbf{s}) = \mathbf{w}^T \mathbf{s}, \qquad \mathbf{s} \in \mathbb{R}^n, \tag{63}$$

where $\mathbf{w} \in \mathbb{R}^n$ is some fixed vector. Then a similar argument implies that

$$\nabla f(\mathbf{s}) = \lambda \nabla g(\mathbf{s}), \tag{64}$$

where

$$\nabla f(\mathbf{s}) = 2A\mathbf{s} \quad \text{and} \quad \nabla g(\mathbf{s}) = \mathbf{w}. \tag{65}$$

Hence

$$\mathbf{s} = \frac{1}{2}\lambda A^{-1}\mathbf{w} \quad \text{and} \quad 1 = \mathbf{w}^T\mathbf{s}, \tag{66}$$

which implies

$$\lambda = \frac{2}{\mathbf{w}^T A^{-1}\mathbf{w}} \tag{67}$$

and

$$\mathbf{s} = \frac{A^{-1}\mathbf{w}}{\mathbf{w}^T A^{-1}\mathbf{w}}. \tag{68}$$

Exercise 10.1. *Prove that* (68) *implies that the corresponding minimal risk-metric is given by*

$$f(\mathbf{s}) = \mathbf{w}^T A^{-1}\mathbf{w}. \tag{69}$$

11. The Cholesky Factorization

Let $\mathbb{P}_n$ denote the set of all non-negative definite symmetric matrices in $\mathbb{R}^{n\times n}$. Given any $A_n \in \mathbb{P}_n$, there is a unique lower triangular matrix $L_n \in \mathbb{R}^{n\times n}$, with positive diagonal elements, for which $A_n = L_n L^T$, and this is called the Cholesky factorization. This section provides a constructive proof of this result, the factorization being obvious when $n = 1$.

Let us now consider the problem of computing the Cholesky factorization $A_{n+1} = L_{n+1}L_{n+1}^T$, where

$$A_{n+1} = \begin{pmatrix} A_n & \mathbf{a} \\ \mathbf{a}^T & b \end{pmatrix} \in \mathbb{R}^{(n+1)\times(n+1)}, \tag{70}$$

where $\mathbf{a} \in \mathbb{R}^n$, $b \geq 0$ and we assume that we have already computed the Cholesky factorization $A_n = L_n L_n^T$. We define

$$L_{n+1} = \begin{pmatrix} L_n & \mathbf{0} \\ \mathbf{p}^T & q \end{pmatrix}, \tag{71}$$

where $\mathbf{p} \in \mathbb{R}^n$ and $q \geq 0$ are to be determined. Then

$$A_{n+1} = \begin{pmatrix} A_n & \mathbf{a} \\ \mathbf{a}^T & b \end{pmatrix} = \begin{pmatrix} L_n & \mathbf{0} \\ \mathbf{p}^T & q \end{pmatrix}\begin{pmatrix} L_n^T & \mathbf{p} \\ \mathbf{0}^T & q \end{pmatrix}, \tag{72}$$

and $\mathbf{p}$ and q must therefore satisfy the equations

$$L_n\mathbf{p} = \mathbf{a} \tag{73}$$

and

$$\|\mathbf{p}\|^2 + q^2 = b. \tag{74}$$

It is (74) that presents the difficulty: we must prove that

$$b \geq \|\mathbf{p}\|^2 = \|L_n^{-1}\mathbf{a}\|^2 \tag{75}$$

to ensure that $q^2 \geq 0$. To this end, we shall first deal with the simpler case when $A_n = I_n$.

Lemma 11.1. *Let $A_n = I_n$. Then $b \geq \|\mathbf{a}\|^2$.*

Proof. For any $\mathbf{v} \in \mathbb{R}^n$ and $w \in \mathbb{R}$ we have

$$\begin{aligned} 0 &\leq \begin{pmatrix} \mathbf{v} \\ w \end{pmatrix}^T \begin{pmatrix} I_n & \mathbf{a} \\ \mathbf{a}^T & b \end{pmatrix}\begin{pmatrix} \mathbf{v} \\ w \end{pmatrix} \\ &= \mathbf{v}^T\mathbf{v} + 2w\mathbf{v}^T\mathbf{a} + bw^2 \\ &= \|\mathbf{v} + w\mathbf{a}\|^2 + \left(b - \|\mathbf{a}\|^2\right) w^2. \end{aligned}$$

Setting $w = 1$ and $v = -a$, we obtain $0 \leq b - \|\mathbf{a}\|^2$, as desired. □

To extend this result to the original case, we use the following trick to relate the general A_{n+1} to the case where $A_n = I_n$.

$$\begin{aligned} &\begin{pmatrix} L_n^{-1} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix}\begin{pmatrix} A_n & \mathbf{a} \\ \mathbf{a}^T & b \end{pmatrix}\begin{pmatrix} L_n^{-T} & \mathbf{0} \\ \mathbf{0}^T & 1 \end{pmatrix} \\ &= \begin{pmatrix} L_n^{-1}A_nL_n^{-T} & L_n^{-1}\mathbf{a} \\ \left(L_n^{-1}\mathbf{a}\right)^T & b \end{pmatrix} \\ &= \begin{pmatrix} I_n & L_n^{-1}\mathbf{a} \\ \left(L_n^{-1}\mathbf{a}\right)^T & b \end{pmatrix}. \end{aligned}$$

Hence Lemma 11.1 implies that

$$b \geq \|L_n^{-1}\mathbf{a}\|^2,$$

which is (75), as required.

Students are often be more familiar with the square-root defined by $A^{1/2} = QD^{1/2}A^TQ$, where $A = QDQ^T$ is the spectral factorization of A, rather than the Cholesky factorization $A = LL^T$. Thus $(A^{1/2})^2 = LL^T$, and it can be shown that $L = A^{1/2}W$, where W is an orthogonal matrix. [Essentially the argument is as follows. If we compute the SVD $L = USV^T$, where U and V are orthogonal matrices and S is the diagonal matrix of singular values of L, then $LL^T = (USV^T)(VSU^T) = US^2U^T = A = QDQ^T$. Hence $U = Q$ and $S = D^{1/2}$. Thus $L = QD^{1/2}V^T = A^{1/2}W$, where $W = QV^T$.]

With this in mind, we see that (75) becomes

$$b \geq \|L_n^{-1}\mathbf{a}\|^2 = \|A_n^{-1/2}\mathbf{a}\|^2 = \mathbf{a}^T A_n^{-1}\mathbf{a}. \tag{76}$$

Once we know condition (76), it's possible to remove all of the scaffolding used above, although I believe most readers will find the more circuitous route described above useful: it's often good to leave some scaffolding in place!

Lemma 11.2. *Let $A_n \in \mathbb{R}^{n\times n}$ be any symmetric non-negative definite matrix and define $A_{n+1} \in \mathbb{R}^{(n+1)\times(n+1)}$ by* (70)*. Then A_{n+1} is non-negative definite if and only if*

$$b \geq \mathbf{a}^T A_n^{-1}\mathbf{a}. \tag{77}$$

Proof. For any $\mathbf{v} \in \mathbb{R}^n$ and $w \in \mathbb{R}$ we have

$$\begin{aligned} 0 &\leq \begin{pmatrix} \mathbf{v} \\ w \end{pmatrix}^T \begin{pmatrix} A_n & \mathbf{a} \\ \mathbf{a}^T & b \end{pmatrix} \begin{pmatrix} \mathbf{v} \\ w \end{pmatrix} \\ &= \mathbf{v}^T A_n \mathbf{v} + 2w\mathbf{v}^T\mathbf{a} + bw^2 \\ &= \|A_n^{1/2}\mathbf{v} + wA_n^{-1/2}\mathbf{a}\|^2 + \left(b - \mathbf{a}^T A_n^{-1}\mathbf{a}\right) w^2. \end{aligned}$$

[How did I complete the square here? The key point is that $\mathbf{v}^T A_n \mathbf{v} = (A_n^{1/2}\mathbf{v})^T(A_n^{1/2}\mathbf{v})$, which implies that we must then write the second term as $\mathbf{v}^T\mathbf{a} = (A_n^{1/2}\mathbf{v})^T(A_n^{-1/2}\mathbf{a})$.] If A_{n+1} is non-negative definite, then setting $w = 1$ and $v = -a$ we obtain $0 \leq b - \mathbf{a}^T A_n^{-1}\mathbf{a}$. Conversely, if $b - \mathbf{a}^T A_n^{-1}\mathbf{a} \geq 0$, then A_{n+1} is non-negative definite. □

12. The Multivariate Beta Function

The classical Beta function is defined by

$$B(\alpha_1, \alpha_2) = \int_0^1 t^{\alpha_1 - 1}(1-t)^{\alpha_2 - 1}\, dt, \tag{78}$$

for positive α_1 and α_2, and satisfies the well-known relation

$$\Gamma(\alpha_1)\Gamma(\alpha_2) = \Gamma(\alpha_1 + \alpha_2)B(\alpha_1, \alpha_2). \tag{79}$$

This note provides a multivariate generalization of (78) and (79), together with a short derivation of Dirichlet's integral as a bonus.

I shall define the n-variate Beta function by

$$B(\alpha_1, \dots, \alpha_n) = n^{-1/2} \int_{\overline{\text{co}}(e_1,\dots,e_n)} t_1^{\alpha_1 - 1} t_2^{\alpha_2 - 1} \cdots t_n^{\alpha_n - 1} d^{n-1}(t), \tag{80}$$

where $\alpha_1, \dots, \alpha_n$ can be any positive numbers and $d^{n-1}(t)$ denotes $(n-1)$-dimensional Lebesgue measure on the closed convex hull $\overline{\text{co}}(e_1, \dots, e_n)$ of the coordinate vectors $e_1, \dots, e_n \in \mathbb{R}^n$. The $n^{-1/2}$ factor ensures that this definition agrees with the classical Beta function when $n = 2$.

There is a natural generalization of (79):

Lemma 12.1. *If $\alpha_1, \dots, \alpha_n > 0$, then*

$$\Gamma(\alpha_1) \cdots \Gamma(\alpha_n) = \Gamma(\alpha_1 + \cdots + \alpha_n) B(\alpha_1, \dots, \alpha_n). \tag{81}$$

Proof. We have

$$\begin{aligned}
&\Gamma(\alpha_1) \cdots \Gamma(\alpha_n) \\
&= \int_{[0,\infty)^n} e^{-(t_1 + \cdots + t_n)} t_1^{\alpha_1 - 1} \cdots t_n^{\alpha_n - 1}\, dt \\
&= \int_0^\infty e^{-u} \left(n^{-1/2} \int_{\overline{\text{co}}(ue_1,\dots,ue_n)} v_1^{\alpha_1 - 1} \cdots v_n^{\alpha_n - 1} d^{n-1}(v) \right) du \\
&= \int_0^\infty e^{-u} u^{n-1+\alpha_1 - 1 + \cdots + \alpha_n - 1} \left(n^{-1/2} \int_{\overline{\text{co}}(e_1,\dots,e_n)} w_1^{\alpha_1 - 1} \cdots w_n^{\alpha_n - 1} d^{n-1}(w) \right) du \\
&= \int_0^\infty e^{-u} u^{\alpha_1 + \cdots + \alpha_n - 1} \left(n^{-1/2} \int_{\overline{\text{co}}(e_1,\dots,e_n)} w_1^{\alpha_1 - 1} \cdots w_n^{\alpha_n - 1} d^{n-1}(w) \right) du \\
&= \Gamma(\alpha_1 + \cdots + \alpha_n) B(\alpha_1, \dots, \alpha_n).
\end{aligned}$$

□

Theorem 12.2. *Let $f \colon [0, \infty) \to \mathbb{R}$ be a continuous function. Then*

$$\begin{aligned}
\int_{S_n} f(t_1 + \cdots + t_n) t_1^{\alpha_1 - 1} \cdots t_n^{\alpha_n - 1}\, dt &= B(\alpha_1, \dots, \alpha_n) \int_0^1 f(u) u^{\alpha_1 + \cdots + \alpha_n - 1}\, du \\
&= \frac{\Gamma(\alpha_1) \cdots \Gamma(\alpha_n)}{\Gamma(\alpha_1 + \cdots + \alpha_n)} \int_0^1 f(u) u^{\alpha_1 + \cdots + \alpha_n - 1}\, du. \qquad (82)
\end{aligned}$$

Proof. Immediate. □

13. Conics

13.1. **The Ellipse and Hyperbola.** Let's begin with the ellipse and the hyperbola, which we shall define as contours of the functions

$$f_{\pm}(\mathbf{x}) = \|\mathbf{x}+\mathbf{s}\| \pm \|\mathbf{x}-\mathbf{s}\|, \qquad \mathbf{x} \in \mathbb{R}^n. \tag{83}$$

The key trick is to observe that

$$4\mathbf{x}^T\mathbf{s} = \|\mathbf{x}+\mathbf{s}\|^2 - \|\mathbf{x}-\mathbf{s}\|^2 = f_+(\mathbf{x})f_-(\mathbf{x}). \tag{84}$$

If $f_+(\mathbf{x}) = \alpha$, then $f_-(\mathbf{x}) = 4\mathbf{x}^T\mathbf{s}/\alpha$. Adding these equations gives

$$f_+(\mathbf{x}) + f_-(\mathbf{x}) = 2\|\mathbf{x}+\mathbf{s}\| = \alpha + \frac{4\mathbf{x}^T\mathbf{s}}{\alpha} = \frac{\alpha^2 + 4\mathbf{x}^T\mathbf{s}}{\alpha}, \tag{85}$$

and squaring both sides yields the quadratic form

$$4\|\mathbf{x}+\mathbf{s}\|^2 = \left(\frac{\alpha^2 + 4\mathbf{x}^T\mathbf{s}}{\alpha}\right)^2. \tag{86}$$

The matrix occurring in this quadratic form is

$$M = 4\left(I_n - \frac{4}{\alpha^2}\mathbf{s}\mathbf{s}^T\right). \tag{87}$$

Similarly, if $f_-(\mathbf{x}) = \alpha$, then $f_+(\mathbf{x}) = 4\mathbf{x}^T\mathbf{s}/\alpha$ and adding them yields (86) and (87). The distinction between the contours of $f_{\pm}$ lies in the eigenvalues of M, which are 1 (with multiplicity $n-1$) and

$$\lambda = 4\left(1 - \frac{4\|\mathbf{s}\|^2}{\alpha^2}\right).$$

If $f_+(\mathbf{x}) = \alpha$, then the triangle inequality implies that

$$\|\mathbf{s}\| \le \frac{1}{2}\left(\|\mathbf{s}+\mathbf{x}\| + \|\mathbf{s}-\mathbf{x}\|\right) = \frac{\alpha}{2},$$

i.e.

$$4\|\mathbf{s}\|^2 \le \alpha^2,$$

which implies $\lambda \ge 0$, with inequality if and only if $\mathbf{x}$ and $\pm\mathbf{s}$ are collinear. Thus M is non-negative definite on contours of f_+ and M is positive definite when the contour is not the line segment joining $\pm\mathbf{s}$.

In contrast, the triangle inequality also implies that

$$\alpha = \|\mathbf{x}+\mathbf{s}\| - \|\mathbf{x}-\mathbf{s}\| \le \|\mathbf{x}+\mathbf{s} - (\mathbf{x}-\mathbf{s})\| = 2\|\mathbf{s}\|,$$

or $4\|\mathbf{s}\|^2/\alpha^2 \ge 1$ on contours of f_-, i.e. $\lambda \le 0$.

13.2. **The Reflector Property.** We have

$$\nabla f_{\pm}(\mathbf{x}) = \frac{\mathbf{x}+\mathbf{s}}{\|\mathbf{x}+\mathbf{s}\|} \pm \frac{\mathbf{x}-\mathbf{s}}{\|\mathbf{x}-\mathbf{s}\|}. \tag{88}$$

Then

$$\left(\frac{\mathbf{x}+\mathbf{s}}{\|\mathbf{x}+\mathbf{s}\|}\right)^T \nabla f_{\pm}(\mathbf{x}) = 1 \pm \frac{(\mathbf{x}+\mathbf{s})^T(\mathbf{x}-\mathbf{s})}{\|\mathbf{x}+\mathbf{s}\|\|\mathbf{x}-\mathbf{s}\|}. \tag{89}$$

and

$$\left(\frac{\mathbf{x}-\mathbf{s}}{\|\mathbf{x}-\mathbf{s}\|}\right)^T \nabla f_{\pm}(\mathbf{x}) = \frac{(\mathbf{x}+\mathbf{s})^T(\mathbf{x}-\mathbf{s})}{\|\mathbf{x}+\mathbf{s}\|\|\mathbf{x}-\mathbf{s}\|} \pm 1. \tag{90}$$

Thus

$$\left(\frac{\mathbf{x}+\mathbf{s}}{\|\mathbf{x}+\mathbf{s}\|}\right)^T \nabla f_{\pm}(\mathbf{x}) = \pm\left(\frac{\mathbf{x}-\mathbf{s}}{\|\mathbf{x}-\mathbf{s}\|}\right)^T \nabla f_{\pm}(\mathbf{x}), \tag{91}$$

which is the reflector property for the ellipse and hyperbola.

13.3. **Conic sections really are conic sections.** Let us take a cone of semi-angle θ in $\mathbb{R}^n$ whose axis is the line generated by a unit vector $\mathbf{u} \in \mathbb{R}^n$, i.e.

$$C = \{\mathbf{x} \in \mathbb{R}^n : \frac{\mathbf{u}^T\mathbf{x}}{\|\mathbf{x}\|} = \pm\cos\theta\}. \tag{92}$$

In other words, the equation for the cone is

$$\left(\mathbf{u}^T\mathbf{x}\right)^2 = \|\mathbf{x}\|^2\cos^2\theta, \tag{93}$$

or

$$\mathbf{x}^T\left(I_n\cos^2\theta - \mathbf{u}\mathbf{u}^T\right)\mathbf{x} = 0. \tag{94}$$

Now let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ be any orthonormal basis for $\mathbb{R}^n$ and consider the hyperplane P with normal vector $\mathbf{v}_n$ at signed distance z_n from the origin. In other words,

$$P = \{\mathbf{x} = \sum_{k=1}^{n} z_k\mathbf{v}_k : z_1, z_2, \ldots, z_{n-1} \in \mathbb{R}\}. \tag{95}$$

If we let $V \in \mathbb{R}^{n\times n}$ be the orthogonal matrix with columns $\mathbf{v}_1, \ldots, \mathbf{v}_n$ and substitute $\mathbf{x} = V\mathbf{z}$ in (93), then we obtain

$$\left(\mathbf{z}^TV^T\mathbf{u}\right)^2 = \|\mathbf{z}\|^2\cos^2\theta. \tag{96}$$

Setting

$$\mathbf{U} = V^T\mathbf{u}, \tag{97}$$

we see that (96) becomes

$$\left(\sum_{k=1}^{n} z_kU_k\right)^2 = \left(\sum_{k=1}^{n} z_k^2\right)\cos^2\theta, \tag{98}$$

or

$$\sum_{k,\ell=1}^{n-1} z_kz_\ell U_kU_\ell + 2z_nU_n\sum_{k=1}^{n-1} z_kU_k + z_n^2U_n^2 = \left(\sum_{k=1}^{n-1} z_k^2\right)\cos^2\theta + z_n^2\cos^2\theta. \tag{99}$$

Hence, writing

$$\widehat{\mathbf{z}} = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_{n-1} \end{pmatrix} \quad \text{and} \quad \widehat{\mathbf{U}} = \begin{pmatrix} U_1 \\ U_2 \\ \vdots \\ U_{n-1} \end{pmatrix}$$

(99) becomes the quadratic form

$$\widehat{\mathbf{z}}^TM\widehat{\mathbf{z}} - 2z_nU_n\widehat{\mathbf{z}}^T\widehat{\mathbf{U}} + z_n^2\left(\cos^2\theta - U_n^2\right) = 0, \tag{100}$$

where the matrix $M \in \mathbb{R}^{(n-1)\times(n-1)}$ is given by

$$M = I_{n-1}\cos^2\theta - \widehat{\mathbf{U}}\widehat{\mathbf{U}}^T. \tag{101}$$

Example 13.1. *Let us choose* $\mathbf{v}_n = \mathbf{u}$, *so that* $\widehat{\mathbf{U}} = 0$ *and* $U_n = 1$. *Then* (100) *becomes*

$$\left(\sum_{k=1}^{n-1} z_k^2\right)\cos^2\theta - z_n^2\sin^2\theta = 0,$$

or

$$\sum_{k=1}^{n-1} z_k^2 = z_n^2\tan^2\theta.$$

The eigenvalues of M are $\cos^2\theta$ (with multiplicity $n-2$) and

$$\mu := \cos^2\theta - \|\widehat{\mathbf{U}}\|^2.$$

Now

$$1 = \|\mathbf{u}\|^2 = \sum_{k=1}^{n} \left(\mathbf{u}^T\mathbf{v}_k\right)^2 = \|\widehat{\mathbf{U}}\|^2 + U_n^2,$$

which implies

$$\mu = \cos^2\theta - \left(1 - U_n^2\right) = U_n^2 - \sin^2\theta. \tag{102}$$

Example 13.2. *Let $n = 3$ and suppose $\mu = 0$. Then*

$$\|\widehat{\mathbf{U}}\|^2 = \cos^2\theta$$

and

$$U_n = \pm\sin\theta.$$

If $\mathbf{q}_1 = \widehat{\mathbf{U}}/\|\widehat{\mathbf{U}}\|$ and $\mathbf{q}_2 \in \mathbb{R}^2$ is orthogonal to $\widehat{\mathbf{U}}$, then the matrix

$$Q = \begin{pmatrix} \mathbf{q}_1 & \mathbf{q}_2 \end{pmatrix} \in \mathbb{R}^{2\times 2}$$

is orthogonal and

$$D := Q^T M Q = \begin{pmatrix} \mu & 0 \\ 0 & \cos^2\theta \end{pmatrix}.$$

If we let $\widehat{\mathbf{z}} = Q\mathbf{x}$, then

$$\cos^2\theta\ x_2^2 \pm 2z_3\sin\theta\cos\theta x_1 + z_3^2\left(\cos^2\theta - \sin^2\theta\right) = 0. \tag{103}$$

13.4. **The Parabola.** We now contour surfaces of the function

$$f(\mathbf{x}) = \mathbf{u}^t\mathbf{x} - \|\mathbf{x} - \mathbf{s}\|, \qquad \text{for } \mathbf{x} \in \mathbb{R}^n. \tag{104}$$

Thus, if $f(\mathbf{x}) = c$, then

$$\mathbf{u}^T\mathbf{x} - c = \|\mathbf{x} - \mathbf{s}\|, \tag{105}$$

i.e. points equidistant from the plane $\mathbf{u}^T\mathbf{z} = c$ and the point $\mathbf{s}$. Squaring (105) we obtain

$$\left(\mathbf{u}^T\mathbf{x} - c\right)^2 = \|\mathbf{x} - \mathbf{s}\|^2. \tag{106}$$

Example 13.3. *If $\mathbf{u} = \mathbf{e}_n$, the n^{th} coordinate vector, and $\mathbf{s} = a\mathbf{e}_n$, then* (106) *becomes*

$$(x_n - c)^2 = \|\mathbf{x} - a\mathbf{e}_n\|^2 = x_1^2 + \cdots + x_{n-1}^2 + (x_n - a)^2.$$

Thus

$$-2cx_n + c^2 = x_1^2 + \cdots + x_{n-1}^2 - 2ax_n + a^2$$

or

$$2(a-c)x_n = x_1^2 + \cdots + x_{n-1}^2 + a^2 - c^2.$$

We can also prove the Reflector Property for the parabola, since

$$\nabla f(\mathbf{x}) = \mathbf{u} - \left(\frac{\mathbf{x} - \mathbf{s}}{\|\mathbf{x} - \mathbf{s}\|}\right) \tag{107}$$

whence

$$\mathbf{u}^T\nabla f(\mathbf{x}) = 1 - \frac{\mathbf{u}^T\left(\mathbf{x} - \mathbf{s}\right)}{\|\mathbf{x} - \mathbf{s}\|} \tag{108}$$

and

$$\left(\frac{\mathbf{x} - \mathbf{s}}{\|\mathbf{x} - \mathbf{s}\|}\right)^T \nabla f(\mathbf{x}) = \frac{\mathbf{u}^T\left(\mathbf{x} - \mathbf{s}\right)}{\|\mathbf{x} - \mathbf{s}\|} - 1. \tag{109}$$

Hence

$$\mathbf{u}^T \nabla f(\mathbf{x}) = -\left(\frac{\mathbf{x}-\mathbf{s}}{\|\mathbf{x}-\mathbf{s}\|}\right)^T \nabla f(\mathbf{x}), \tag{110}$$

or

$$\mathbf{u}^T \nabla f(\mathbf{x}) = \left(\frac{\mathbf{s}-\mathbf{x}}{\|\mathbf{s}-\mathbf{x}\|}\right)^T \nabla f(\mathbf{x}), \tag{111}$$

which is the reflector property.

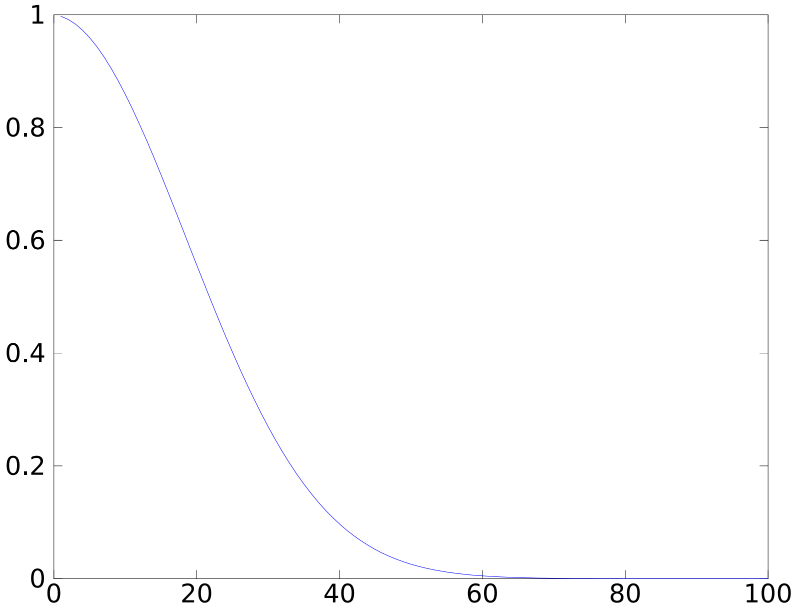

FIGURE 2. The probability that all n birthdays are different

14. The Birthday Problem

This is a traditional probabilistic problem: given n people, whose birthdays are assumed to be uniformly distributed over the $N = 365$ days of the year (ignoring leap years), find the probability that at least two of them share a birthday. Now

$$\mathbb{P}(\text{at least two share a birthday}) = 1 - \mathbb{P}(\text{all birthdays distinct}) =: 1 - p_n. \tag{112}$$

Now

$$p_n = \frac{N(N-1)(N-2)\cdots(N-n+1)}{N^n} \tag{113}$$

and, dividing numerator and denominator by N^n, we obtain

$$p_n = \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right)\cdots\left(1 - \frac{n-1}{N}\right). \tag{114}$$

In one sense the problem is now solved. The surprise is that p_n tends to zero rather quickly. Indeed, $p_{23} = 0.5073$, by direct calculation. However, plotting p_n reveals a suspiciously Gaussian curve, as we see in Figure 2. Why does p_n decay so quickly and can we understand the seemingly Gaussian behaviour?

```
N=365; n=100; p=1; prob=zeros(1,n);
%
% prob(k) = (1 - 1/N)(1 - 2/N)...(1-k/N)
%         = prob(all k+1 bdays different)
%
for k=1:n
  p=p*(1-k/N);
  prob(k)=p;
end
```

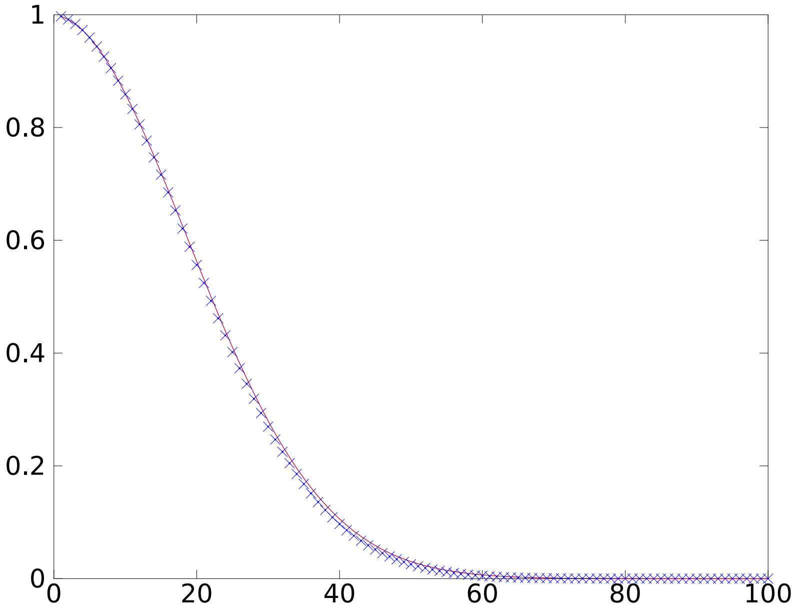

FIGURE 3. x: p_n; line: $\exp(-n(n-1)/(2N))$

First take logarithms:

$$\log p_n = \sum_{k=1}^{n-1} \log\left(1 - \frac{k}{N}\right). \tag{115}$$

Now

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \frac{x^4}{4} + \cdots,$$

and the series is convergent for $|x| < 1$. Thus

$$\log(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \frac{x^4}{4} - \cdots \leq -x, \tag{116}$$

for $0 \leq x < 1$. If $|x| \ll 1$, then we also have the approximation $\log(1+x) \approx -x$. Thus

$$\log p_n \leq -\sum_{k=1}^{n-1} \frac{k}{N} = -\frac{n(n-1)}{2N}, \tag{117}$$

which implies

$$p_n \leq e^{-n(n-1)/(2N)}. \tag{118}$$

This explains the rapid decay and the Gaussian resemblance, as we see in Figure 3.

15. The Bike Problem via the Inclusion–Exclusion Formula

Suppose n cyclists randomly permute their bikes. What is the probability that at least one cyclist has the correct bike?

More formally, the sample space X consits of all positive permutations of the integers $1, 2, \ldots, n$, i.e.

$$X = \{(i_1, i_2, \ldots, i_n) : i_1, \ldots, i_n \text{ a permutation of } 1, \ldots, n\}. \tag{119}$$

We shall assign each of these permutations the same probability $1/n!$.

Further, define

$$A_k = \{x \in X : i_k = k\}, \qquad \text{for } k = 1, 2, \ldots, n. \tag{120}$$

Thus A_k is the set of outcomes for which cyclist k gets bike k. We want to calculate the probability

$$\mathbb{P}\left(A_1 \cup A_2 \cup \cdots \cup A_n\right).$$

Example 15.1. *If $n = 3$, then the sample space is*

$$X = \{(123), (132), (213), (231), (312), (321)\}.$$

Then $A_1 = \{(123), (132)\}$, $A_2 = \{(123), (321)\}$ and $A_3 = \{(123), (213)\}$, whilst $\mathbb{P}(A_1 \cup A_2 \cup A_3) = 4/6 = 2/3$.

The solution requires the inclusion–exclusion formula:

$$\mathbb{P}(A_1 \cup A_2 \cup \cdots \cup A_k) = \sum_{\ell=1}^{n} (-1)^{\ell-1} \sum_{1 \le k_1 < k_2 < \cdots < k_\ell \le n} \mathbb{P}(A_{k_1} \cap A_{k_2} \cap \cdots \cap A_{k_\ell}). \tag{121}$$

Exercise 15.1.

$$\mathbb{P}(A_{k_1} \cap \cdots \cap A_{k_m} = \frac{(n-m)!}{n!}.$$

Thus

$$\sum_{1 \le k_1 < k_2 < \cdots < k_\ell \le n} \mathbb{P}(A_{k_1} \cap A_{k_2} \cap \cdots \cap A_{k_\ell}) = \binom{n}{m} \frac{(n-m)!}{n!} = \frac{1}{m!}.$$

Hence

$$\mathbb{P}(A_1 \cup \cdots \cup A_n) = 1 - \frac{1}{2!} + \frac{1}{3!} + \cdots + \frac{(-1)^{n-1}}{n!} \to 1 - e^{-1}. \tag{122}$$

15.1. The Inclusion–Exclusion Formula.

For each subset A of the sample space X, the *indicator function* $I_A : X \to \{0, 1\}$ is defined by $I_A(x) = 1$ if and only if $x \in A$.

Example 15.2. *For Example 15.1 we have*

$$I_{A_1}(123) = I_{A_1}(132) = 1$$

but

$$I_{A_1}(213) = I_{A_1}(231) = I_{A_1}(312) = I_{A_1}(321) = 0.$$

The indicator function has some crucial properties. Firstly

$$1 - I_A(x) = I_{A^c}(x)$$

where A_c is the complement of A, i.e. $X \setminus A$. Further,

$$I_{A \cap B}(x) = I_A(x) I_B(x).$$

Further, we use de Morgan's Law:

$$(A_1 \cup \cdots \cup A_n)^c = A_1^c \cap \cdots \cap A_n^c.$$

Thus

$$\begin{aligned} I_{A_1\cup\cdots\cup A_n}(x) &= 1 - I_{(A_1\cup\cdots\cup A_n)^c}(x) \\ &= 1 - \prod_{k=1}^{n} \left(1 - I_{A_k}(x)\right). \end{aligned}$$

16. Binomial Coefficients and the Central Limit Theorem

16.1. **The Aymptotic Behaviour of $\binom{2n}{n}$.** Let's begin with the integral

$$\begin{aligned} I_n &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \cos^{2n}\theta\, d\theta \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} \left(\frac{e^{i\theta}+e^{-i\theta}}{2}\right)^{2n} d\theta \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} 2^{-2n}\sum_{k=0}^{2n}\binom{2n}{k} e^{ik\theta}e^{-i(2n-k)\theta}\, d\theta \\ &= 2^{-2n}\binom{2n}{n}. \qquad (123) \end{aligned}$$

Now, using the substitution $\theta = t/\sqrt{n}$,

$$\begin{aligned} \sqrt{n}I_n &= \frac{\sqrt{n}}{\pi}\int_{-\pi/2}^{\pi/2} \cos^{2n}\theta\, d\theta \\ &= \frac{1}{\pi}\int_{-(\pi/2)\sqrt{n}}^{(\pi/2)\sqrt{n}} \cos^{2n}\left(t/\sqrt{n}\right)\, dt \\ &\to \frac{1}{\pi}\int_{-\infty}^{\infty} e^{-t^2}\, dt = \frac{1}{\sqrt{\pi}}, \qquad (124) \end{aligned}$$

by the Dominated Convergence Theorem.

Comparing (123) and (124) we obtain

$$\lim_{n\to\infty} \sqrt{n}\binom{2n}{n}4^{-n} = \frac{1}{\sqrt{\pi}}, \qquad (125)$$

or

$$\binom{2n}{n}4^{-n} \sim \frac{1}{\sqrt{\pi n}}. \qquad (126)$$

Example 16.1. *We can also check our calculation using Stirling's asymptotic formula for $n!$:*

$$\lim_{n\to\infty} \frac{n!}{\sqrt{2\pi n}n^n e^{-n}} = 1,$$

or

$$n! \sim \sqrt{2\pi n}n^n e^{-n}.$$

Then

$$\binom{2n}{n} \sim \frac{\sqrt{2\pi\cdot 2n}\,(2n)^{2n}\,e^{-2n}}{\left(\sqrt{2\pi n}n^n e^{-n}\right)^2} = \frac{\sqrt{4\pi n}4^n}{2\pi n} = \frac{4^n}{\sqrt{\pi n}}.$$

16.2. **Ratios of central binomial coefficients.** The ratios of the central binomial coefficients

$$\{\binom{2n}{n+k} : -L \le k \le L\},$$

where n is very large and L is small compared with n, look Gaussian, and this is reminiscent of our earlier description of the Birthday Problem. We have the ratio

$$\begin{aligned}
R(n,k) &:= \frac{\binom{2n}{n+k}}{\binom{2n}{n}} \\
&= \left(\frac{(2n)!}{(n-k)!(n+k)!}\right)\left(\frac{(n!)^2}{(2n)!}\right) \\
&= \frac{(n!)^2}{(n-k)!(n+k)!} \\
&= \frac{n!n(n-1)\cdots(n-k+1)(n-k)!}{(n-k)!(n+k)(n+k-1)\cdots(n+1)n!} \\
&= \frac{n(n-1)\cdots(n-k+1)}{(n+k)(n+k-1)\cdots(n+1)} \\
&= \frac{(1-\frac{1}{n})(1-\frac{2}{n})\cdots(1-\frac{k-1}{n})}{(1+\frac{1}{n})(1+\frac{2}{n})\cdots(1+\frac{k}{n})}.
\end{aligned}$$

Taking logarithms, we find

$$\begin{aligned}
\log R(n,k) &= \sum_{j=1}^{k-1}\log\left(1-\frac{j}{n}\right) - \sum_{j=1}^{k}\log\left(1+\frac{j}{n}\right) \\
&\approx -\sum_{j=1}^{k-1}\frac{j}{n} - \sum_{j=1}^{k}\frac{j}{n} \\
&= \frac{-k(k-1)-(k+1)k}{2n} = -\frac{k^2}{n}.
\end{aligned}$$

Hence, fixing L, we obtain

$$\lim_{n\to\infty}\frac{R(n,k)}{e^{-k^2/n}} = 1, \qquad \text{for } -L \le k \le L.$$

Writing $k = t\sqrt{n}$, we obtain $R(n, t\sqrt{n}) \sim e^{-t^2}$.

16.3. **Extending the integral approach.** We use

$$\begin{aligned}
\binom{2n}{n+k}4^{-n} &= \frac{1}{2\pi}\int_{-\pi}^{\pi}\cos^{2n}\theta e^{-2ik\theta}\,d\theta \\
&= \frac{1}{\pi}\int_{-\pi/2}^{\pi/2}\cos^{2n}\theta e^{-2ik\theta}\,d\theta \\
&= \frac{1}{\pi}\int_{-(\pi/2)\sqrt{n}}^{(\pi/2)\sqrt{n}}\cos^{2n}\left(\frac{t}{\sqrt{n}}\right)e^{-2ikt/\sqrt{n}}n^{-1/2}\,dt, \qquad (127)
\end{aligned}$$

using the substitution $\theta = t/\sqrt{n}$ in the last line. Writing this in terms of Gamma functions, we obtain

$$\frac{\sqrt{n}\Gamma(2n+1)4^{-n}}{\Gamma(n+k+1)\Gamma(n-k+1)} = \frac{1}{\pi}\int_{-(\pi/2)\sqrt{n}}^{(\pi/2)\sqrt{n}}\cos^{2n}\left(\frac{t}{\sqrt{n}}\right)e^{-2ikt/\sqrt{n}}\,dt, \qquad (128)$$

which emphasizes that both sides analytically continue. Now writing $k = z\sqrt{n}$, for fixed $z \in \mathbb{C}$, we obtain

$$\begin{aligned} \frac{\sqrt{n}\Gamma(2n+1)4^{-n}}{\Gamma(n+z\sqrt{n}+1)\Gamma(n-z\sqrt{n}+1)} &= \frac{1}{\pi}\int_{-(\pi/2)\sqrt{n}}^{(\pi/2)\sqrt{n}} \cos^{2n}\left(\frac{t}{\sqrt{n}}\right) e^{-2izt}\,dt \\ &\to \frac{1}{\pi}\int_{\mathbb{R}} e^{-t^2} e^{-2izt}\,dt \\ &= \frac{1}{\sqrt{\pi}} e^{-z^2}, \end{aligned} \tag{129}$$

or

$$\lim_{n\to\infty} \sqrt{\pi n}\binom{2n}{n+z\sqrt{n}}4^{-n} = e^{-z^2}. \tag{130}$$

One way to check this for large n is to use the logarithm of the Gamma function. For example,

```
z=2;
n=10000;
a = 0.5*log(pi*n)+gammaln(2*n+1)
b = gammaln(n+z*sqrt(n)+1) + gammaln(n-z*sqrt(n)+1)+n*log(4);
a-b
ans =   -4.00007917595212e+00
```

which is in excellent agreement.

17. Monotone Matrix Functions

Let A and B be any pair of non-negative definite, symmetric, $n \times n$ matrices for which $A \preceq B$, in the sense that $x^T Ax \le x^T Bx$, for every $x \in \mathbb{R}^n$. In response to a colleague's question, this note demonstrates that $A^\beta \le B^\beta$, for every $\beta \in (0, 1)$. This result is almost certainly *not* new, for the field of so called *monotone matrix functions* is large and well-established; see, for example, [3]. However, I found the literature sufficiently forbidding that it seemed easier to proceed independently, for the integral relations used here are similar to those arising in the theory of radial basis functions. In fact, the only reason for my being aware of [3] is its description of W. Feller's elegant proof of the Bernstein theorem (which states that the completely monotonic functions are precisely the Laplace transforms of positive Borel measures on $[0, \infty)$).

Let $\mathbb{P}_n$ denote the cone of $n \times n$ symmetric positive definite matrices, and let $\overline{\mathbb{P}}_n$ be the set of symmetric non-negative definite matrices[1]. Give any $A, B, \in \overline{\mathbb{P}}_n$, we write $A \preceq B$ if $B - A \in \overline{\mathbb{P}}_n$; equivalently, $A \preceq B$ if and only if $x^T Ax \le x^T Bx$, for all $\in \mathbb{R}^n$. We shall say that a mapping $f : \overline{\mathbb{P}}_n \to \overline{\mathbb{P}}_n$ is a *monotone matrix function* if $A \preceq B$ implies $f(A) \preceq f(B)$, for all $A, B \in \overline{\mathbb{P}}_n$. The primary purpose of this note is to demonstrate that $A \mapsto A^\beta$ is a monotone matrix function when $0 < \beta < 1$. Further details on the partial order $\preceq$ may be found in Chapter 7 of [4].

Lemma 17.1. (i) *If $A, B \in \mathbb{P}_n$, then $A \preceq B$ if and only if $\rho(AB^{-1}) \le 1$, where ρ is the spectral radius function.*

(ii) *If $A, B \in \mathbb{P}_n$, then $A \preceq B$ if and only if $B^{-1} \preceq A^{-1}$.*

(iii) *If $A, B \in \mathbb{P}_n$ and $A \preceq B$, then $(I + tA^{-1})^{-1} \preceq (I + tB^{-1})^{-1}$, for all $t \ge 0$.*

Proof. (i) First note that $B - A \in \overline{\mathbb{P}}_n$ if and only if $I - B^{-1/2}AB^{-1/2} \in \overline{\mathbb{P}}_n$, and the latter condition holds if and only if $\rho(B^{-1/2}AB^{-1/2}) \le 1$. Finally, we use the similarity transformation $AB^{-1} = B^{1/2}\left(B^{-1/2}AB^{-1/2}\right)B^{-1/2}$.

(ii) We only need to notice that AB^{-1} and $B^{-1}A$ are similar to $A^{-1/2}B^{-1}A^{-1/2}$.

(iii) If $A \preceq B$, then $B^{-1} \preceq A^{-1}$, which implies $I + tA^{-1} \succeq I + tB^{-1}$, for all $t \ge 0$. Hence $(I + tA^{-1})^{-1} \preceq (I + tB^{-1})^{-1}$.

□

We see that Lemma 17.1 (iii) implies that

$$A \preceq \sum_{k=1}^{m} w_k (I + t_k B^{-1})^{-1}$$

if $A \preceq B$ in $\overline{\mathbb{P}}_n$ and the numbers $\{w_k\}$ and $\{t_k\}$ are positive. We deduce the following continuous limit.

Corollary 17.2. *Let $w : (0, \infty) \to (0, \infty)$ be any continuous function for which*

$$\int_0^1 w(t)\, dt < \infty \qquad \text{and} \qquad \int_1^\infty t^{-1} w(t)\, dt < \infty.$$

Then the function $f : (0, \infty) \to (0, \infty)$ defined by the integral relation

$$f(A) = \int_0^\infty (I + tA^{-1})^{-1} w(t)\, dt, \qquad A \in \mathbb{P}_n, \tag{131}$$

is a monotone matrix function.

Proof. This is an immediate consequence of Lemma 17.1 (iii). I've chosen simple conditions on the weight function that are sufficient for the integral to be well-defined. □

[1] $\mathbb{P}_n$ is an *open* subset of $\mathbb{R}^{n\times n}$ in any norm, so $\overline{\mathbb{P}}_n$ is its closure.

We shall be using a particular weight function, but the same argument shows that

$$f(A) = \int_0^\infty (I + tA^{-1})^{-1}\, d\mu(t), \qquad A \in \mathbb{P}_n,$$

defines a monotone matrix function for any positive Borel measure μ satisfying

$$\int_0^1 d\mu(t) < \infty \quad \text{and} \quad \int_1^\infty t^{-1}\, d\mu(t) < \infty.$$

Lemma 17.3. *Let $\alpha \in (0,1)$. Then*

$$I(\alpha) := \int_0^\infty \tau^{-\alpha}(1+\tau)^{-1}\, d\tau = \frac{\pi}{\sin \alpha\pi}. \tag{132}$$

Proof. This is an undergraduate exercise in contour integration. □

Theorem 17.4. *Let $\beta \in (0,1)$. Then*

$$A^\beta = \frac{\sin \beta\pi}{\pi} \int_0^\infty t^{\beta-1}(I + tA^{-1})^{-1}\, dt, \tag{133}$$

which implies that $A \mapsto A^\beta$ is a monotone matrix function.

Proof. If we set $\tau = t/a$, for $a > 0$, in (132), then we obtain

$$a^{1-\alpha} = \frac{\sin \alpha\pi}{\pi} \int_0^\infty t^{-\alpha}(1 + ta^{-1})^{-1}\, dt,$$

which implies the formula

$$A^{1-\alpha} = \frac{\sin \alpha\pi}{\pi} \int_0^\infty t^{-\alpha}(I + tA^{-1})^{-1}\, dt, \tag{134}$$

for $A \in \mathbb{P}_n$. Setting $\beta = 1 - \alpha$, we deduce (133), for any $A \in \mathbb{P}_n$ and $\beta \in (0,1)$. □

My friend and colleague M. J. D. Powell has also supplied an elementary argument demonstrating that $A \mapsto A^{1/2}$ is a monotone matrix function.

Corollary 17.5. *Let $A, B \in \overline{\mathbb{P}}_n$. If $A \preceq B$, then $A^{1/2} \preceq B^{1/2}$.*

Proof. By Lemma 17.1 (i), $A^{1/2} \preceq B^{1/2}$ if and only if $\rho(A^{1/2}B^{-1/2}) \le 1$. Now

$$\|A^{1/2}B^{-1/2}v\|^2 = v^T B^{-1/2}AB^{-1/2}v \le \|v\|^2,$$

because $\rho(B^{-1/2}AB^{-1/2}) \le 1$ if $A \preceq B$, again by Lemma 17.1 (i). Hence $\rho(A^{1/2}B^{-1/2}) \le 1$, as required. □

All of this has been couched in matrix language, but the original question was posed in a Sobolev space.

18. Umbral Calculus

In the Nineteenth Century, and for much of the Twentieth Century, umbral calculus was a useful tool of dubious repute with strong formal similarities to operator calculus. However, while operator calculus became respectable via Fourier–Laplace analysis and distribution theory, umbral calculus had to wait until the work of Rota and others in the 1970s. This note uses umbral calculus to obtain yet another derivation of the solution of a constant-coefficient linear recurrence relation, but I hope it's an attractive advertisement for umbral calculus. I don't know if this particular use of umbral calculus is new, but I strongly suspect it's a rediscovery. I hope that readers interested in learning more umbral calculus will consult [6], [5], or one of the many fascinating papers of Zeilberger, e.g. [7].

For any complex sequence $\{a_n\}_{n=0}^{\infty}$, an *umbra* is any linear functional L on the algebra $\mathbb{P} = \mathbb{C}[z]$ of polynomials, defined by $L(z^n) = a_n$, for $n \ge 0$. This simple definition, due to Rota, might seem obvious in retrospect, but this is often the way.

18.1. Recurrence Relations.

Lemma 18.1. *The set $\mathcal{A}$ of complex sequences $\{a_n\}_{n=0}^{\infty}$ satisfying the recurrence relation*

$$a_{n+1} + \lambda a_n + \mu a_{n-1} = 0, \qquad \textit{for } n \ge 1, \tag{135}$$

where $\lambda, \mu \in \mathbb{C}$ are constants, is in bijection with the set $\mathcal{A}^*$ of linear functionals annihilating the polynomial ideal generated by $z^2 + \lambda z + \mu$.

Proof. If an umbra $L \in \mathbb{P}^*$ generates a sequence satisfying (135), i.e. $L(z^n) = a_n$, for $n \ge 0$, then

$$0 = a_{n+1} + \lambda a_n + \mu a_{n-1} = L\left(z^{n-1}\left[z^2 + \lambda z + \mu\right]\right), \qquad n \ge 1,$$

that is,

$$L(P(z)\left[z^2 + \lambda z + \mu\right]) = 0,$$

for any polynomial $P(z) \in \mathbb{P}$. The converse is immediate. □

Let $z^2 + \lambda z + \mu = (z - \nu_1)(z - \nu_2)$. Given any polynomial $p(z) \in \mathbb{P}$, we have

$$p(z) = p(\nu_1) + p[\nu_1, \nu_2](z - \nu_1) + (z - \nu_1)(z - \nu_2)q(z),$$

for some polynomial $q(z) \in \mathbb{P}$, where the divided difference $p[\nu_1, \nu_2]$ is defined by

$$p[\nu_1, \nu_2] = \begin{cases} \frac{p(\nu_2) - p(\nu_1)}{\nu_2 - \nu_1} & \text{for } \nu_1 \ne \nu_2, \\ p'(\nu_1) & \nu_1 = \nu_2. \end{cases}$$

Hence

$$\begin{aligned} L(p) &= p(\nu_1)L(1) + p[\nu_1, \nu_2](L(z) - \nu_1 L(1)) \\ &= p(\nu_1)a_0 + p[\nu_1, \nu_2](a_1 - \nu_1 a_0), \end{aligned} \tag{136}$$

and, when $p(z) = z^n$, for $n \ge 0$, we have

$$a_n = L(z^n) = \begin{cases} a_0\nu_1^n + (a_1 - \nu_1 a_0)\left(\frac{\nu_2^n - \nu_1^n}{\nu_2 - \nu_1}\right) & \text{for } \nu_1 \ne \nu_2, \\ a_0\nu_1^n + (a_1 - \nu_1 a_0)n\nu_1^{n-1} & \nu_1 = \nu_2. \end{cases} \tag{137}$$

19. The Lanczos Algorithm

A Krylov space is a subspace of the form

$$K \equiv K(A, v, l) := \left\langle v, Av, A^2v, \ldots, A^l v \right\rangle . \tag{138}$$

Thus every element of K can be written as $p(A)v$ for some $p \in \mathbb{P}_l$. Indeed, we could define K as $\mathbb{P}_l(A)v$.

Now let A be a symmetric matrix whose distinct eigenvalues are $(\lambda_k)_{k=1}^m$. Every vector v can be expressed as $v = \sum_{k=1}^m v_k u_k$, where $Au_j = \lambda_j u_j$ and $u_j^T u_k = \delta_{jk}$. Hence

$$p(A)v = \sum_{k=1}^m p(\lambda_k) v_k u_k, \qquad p \in \mathbb{P}_l. \tag{139}$$

How do we generate orthogonal bases for K? If $(p(A)v)^T q(A)v = 0$, then

$$\sum_{k=1}^m p(\lambda_k) q(\lambda_k) v_k^2 = 0, \tag{140}$$

that is $(p, q) = 0$ where $(\cdot, \cdot)$ denotes the semi-inner product

$$(f, g) = \sum_{k=1}^m f(\lambda_k) g(\lambda_k) v_k^2. \tag{141}$$

Thus every orthogonal basis $(p_k(A)v)_{k=1}^m$ of K corresponds to a set of orthogonal polynomials $(p_k)_1^m$. Therefore one way to generate an orthogonal basis for K is to use the three term recurrence relation

$$\phi_{k+1}(t) = (t - \rho_k)\phi_k(t) - \sigma_k^2 \phi_{k-1}(t), \qquad k \geq 1, \tag{142}$$

the corresponding orthogonal basis being $\{\phi_k(A)v : k = 0, 1, \ldots, m-1\}$. Rewriting (5) in terms of $r_k = \phi_k(A)v$, we obtain

$$r_{k+1} = (A - \rho_k I) r_k - \sigma_k^2 r_{k-1}, \qquad k \geq 1. \tag{143}$$

This provides Algorithm 19.1.

Algorithm 19.1. . *Set $r_0 = v$, $\rho_0 = r_0^T A r_0 / r_0^T r_0$ and $r_1 = (A - \rho_0 I) r_0$.*

```
For $k = 1, 2, \ldots$ do begin
  $\rho_k = r_k^T A r_k / r_k^T r_k$
  $\sigma_k^2 = r_{k-1}^T A r_k / r_{k-1}^T r_{k-1}$
    $r_{k+1} = (A - \rho_k I) r_k - \sigma_k^2 r_{k-1}$
  Stop if $\|r_{k+1}\|_2$ is sufficiently small.
end.
```

In matrix terms we have

$$AR = RT, \tag{144}$$

where $R = (r_0, r_1, \ldots, r_{m-1})$ and T is the tridiagonal matrix

$$T = \begin{pmatrix} \rho_0 & -\sigma_1^2 & & & & \\ 1 & \rho_1 & -\sigma_2^2 & & & \\ & 1 & \rho_2 & & & \\ & & & \ddots & & \\ & & & & \rho_{m-2} & -\sigma_{m-1}^2 \\ & & & & 1 & \rho_{m-1} \end{pmatrix}. \tag{145}$$

The Lanczos algorithm is a slightly different form of Algorithm 1 which uses the observation that $R = QD$, where Q is an orthogonal matrix and $D := \operatorname{diag}(d_0, \ldots, d_{m-1})$. Equation (7) becomes

$$AQD = QDT \qquad \text{or} \qquad Q^T AQ = \tilde{T} := DTD^{-1}, \tag{146}$$

and we see that $\tilde{T}$ is a symmetric tridiagonal matrix. The columns $(q_K)_{k=0}^{m-1}$ obey the recurrence relation

$$d_{k+1}q_{k+1} = (A - \rho_k I)d_k q_k - \sigma_k^2 d_{k-1}q_{k-1} \tag{147}$$

and, because $\tilde{T}$ is symmetric, this becomes

$$Aq_k = \rho_k q_k + \delta_{k-1}q_{k-1} + \delta_k q_{k+1}, \qquad k \geq 1, \tag{148}$$

or

$$\delta_k q_{k+1} = (A - \rho_k I)q_k - \delta_{k-1}q_{k-1}. \tag{149}$$

Algorithm 19.2. *Set* $q_0 = v/\|v\|_2$, $\rho_0 = q_0^T Aq_0$ *and* $r = (A - \rho_0 I)q_0$. *Let* $\delta_0 = \|r\|_2$ *and define* $q_1 = r/\delta_0$.

```
For $k = 1, 2, \ldots$ do begin
  $\rho_k = q_k^T Aq_k$
    $r = (A - \rho_k I)q_k + \delta_{k-1}q_{k-1}$
    Let $\delta_k = \|r\|_2$ and $q_{k+1} = r/\delta_k$.
  Stop if $\delta_k$ is sufficiently small.
end.
```

This short note establishes a conjecture of Arieh's. I would be very surprised if this were not known — indeed, I thought it was in one of Halmos' books on linear algebra.

Theorem 19.1. *Let U be any $m \times m$ unitary matrix. Then the principal submatrix $U(1:n, 1:n)$ has at least* $\max\{2n-m, 0\}$ *singular values equal to unity.*

Proof. I shall prove a more general theorem. Let $2n > m$ and let W be any n-dimensional subspace of $\mathbb{C}^m$. Then $W_U := W \cap UW$ is a U-invariant subspace of dimension

$$\dim W_U = \dim W + \dim UW - \dim{(W + UW)} \geq 2n - m \geq 1.$$

Now let P_W denote orthogonal projection onto W and consider the matrix $\hat{U} := P_W U P_W$. Since $\hat{U} = U$ on W_U, and U is an isometry, we deduce that $\hat{U}$ has at least $\dim W_U \geq 2n-m$ singular values equal to unity. We obtain the stated special case by choosing

$$W = \operatorname{span}\{e_1, \ldots, e_n\}.$$

□

20. Mortgages – a once exotic instrument

You are presumably all too familiar with a repayment mortgage: we borrow a large sum M for a fairly large slice T of our lifespan, repaying capital and interest using N regular payments. The interest rate is assumed to be constant and it's a secured loan: our homes are forfeit on default. How do we calculate our repayments?

Let $h = T/N$ be the interval between payments, let $D_h : [0, T] \to \mathbb{R}$ be our debt as a function of time, and let $A(h)$ be our payment. We shall assume that our initial debt is $D_h(0) = 1$, because we can always multiply by the true initial cost M of our house after the calculation. Thus D must satisfy the equations

$$D_h(0) = 1, \quad D_h(T) = 0 \quad \text{and} \quad D_h(\ell h) = D_h((\ell - 1))e^{rh} - A(h). \tag{150}$$

We see that $D(h)_h = e^{rh} - A(h)$, while

$$D_h(2h) = D_h(h)e^{rh} - A(h) = e^{2rh} - A(h)\left(1 + e^{rh}\right).$$

The pattern is now fairly obvious:

$$D_h(\ell h) = e^{\ell rh} - A(h) \sum_{k=0}^{\ell-1} e^{krh}, \tag{151}$$

and summing the geometric series[2]

$$D_h(\ell h) = e^{\ell rh} - A(h) \left(\frac{e^{\ell rh} - 1}{e^{rh} - 1} \right). \tag{152}$$

In order to achieve $D(T) = 0$, we choose

$$A(h) = \frac{e^{rh} - 1}{1 - e^{-rT}}. \tag{153}$$

Exercise 20.1. *What happens if $T \to \infty$?*

Exercise 20.2. *Prove that*

$$D_h(\ell h) = \frac{1 - e^{-r(T - \ell h)}}{1 - e^{-rT}}. \tag{154}$$

Thus, if $t = \ell h$ is constant (so we increase ℓ as we reduce h), then

$$D_h(t) = \frac{1 - e^{-r(T-t)}}{1 - e^{-rT}}. \tag{155}$$

Almost all mortgages are repaid by 300 monthly payments for 25 years. However, until recently, many mortgages calculated interest *yearly*, which means that we choose $h = 1$ in Exercise 20.1 and then divide $A(1)$ by 12 to obtain the monthly payment.

Exercise 20.3. *Calculate the monthly repayment $A(1)$ when $M = 10^5$, $T = 25$, $r = 0.05$ and $h = 1$. Now repeat the calculation using $h = 1/12$. Interpret your result.*

In principle, there's no reason why our repayment could not be continuous, with interest being recalculated on our constantly decreasing debt. For continuous repayment, our debt $D : [0, T] \to \mathbb{R}$ satisfies the relations

$$D(0) = 1, \quad D(T) = 0 \quad \text{and} \quad D(t + h) = D(t)e^{rh} - hA. \tag{156}$$

[2] Many students forget the simple formula. If $S = 1 + a + a^2 + \cdots + a^{m-2} + a^{m-1}$, then $aS = a + a^2 + \cdots + a^{m-1} + a^m$. Subtracting these expressions implies $(a - 1)S = a^m - 1$, all other terms cancelling.

Exercise 20.4. *Prove that*

$$D'(t) - rD(t) = -A, \tag{157}$$

where, in particular, you should prove that (156) implies the differentiability of $D(t)$. Solve this differential equation using the integrating factor e^{-rt}. You should find the solution

$$D(t)e^{-rt} - 1 = A\int_0^t \left(-e^{-r\tau}\right) d\tau = A\left(\frac{e^{-rt}-1}{r}\right). \tag{158}$$

Hence show that

$$A = \frac{r}{1-e^{-rT}} \tag{159}$$

and

$$D(t) = \frac{1-e^{-r(T-t)}}{1-e^{-rT}}, \tag{160}$$

agreeing with (155)*, i.e.* $D_h(kh) = D(kh)$*, for all* k*. Prove that* $\lim_{r\to\infty} D(t) = 1$ *for* $0 < t < T$ *and interpret.*

Observe that

$$\frac{A(h)}{Ah} = \frac{e^{rh}-1}{rh} \approx 1 + (rh/2), \tag{161}$$

so that continuous repayment is optimal for the borrower, but that the mortgage provider is making a substantial profit. Greater competition has made yearly recalculations much rarer, and interest is often paid daily, i.e. $h = 1/250$, which is rather close to continuous repayment.

Exercise 20.5. *Construct graphs of $D(t)$ for various values of r. Calculate the time $t_0(r)$ at which half of the debt has been paid.*

20.1. **Pricing Mortgages via lack of arbitrage.**

There is a very slick arbitrage argument to deduce the continuous repayment mortgage debt formula (160). Specifically, the simple fact that $D(t)$ is a deterministic financial instrument implies, via arbitrage, that $D(t) = a + b\exp(rt)$, so we need only choose the constants a and b to satisfy $D(0) = 1$ and $D(T) = 1$, which imply $a + b = 1$ and $a + b\exp(rT) = 0$. Solving these provides $a = \exp(rT)/(\exp(rT) - 1)$ and $b = -1/(\exp(rT) - 1)$, and equivalence to (160) is easily checked.

21. Pensions

In this note we compute the requirements for Defined Benefit pensions, where the pension is a function of final salary and length of service. The numbers are chosen to mimic USS before 2016, when it was an undiminished final salary scheme: it is salutary to remember the pension we enjoyed so recently. The idea is solely to provide a simple (even simplistic) model for experiment, the focus being the finance: all actuarial details are therefore ignored (except for retirement years before death!).

We assume that a scheme member begins with an initial income of 1 at time zero, which grows at rate g until retirement at time R. The member saves at contribution rate c at interest rate r, so the pension fund at retirement is given by

$$F = c \int_0^R e^{gt} e^{r(R-t)}\, dt \tag{162}$$

if we assume continuous time, for simplicity. Thus

$$F = c\left(\frac{e^{gR} - e^{rR}}{g-r}\right) = ce^{gR}\left(\frac{1 - e^{-(g-r)R}}{g-r}\right). \tag{163}$$

At retirement, the aim is to pay the member a lump sum of $1.5\exp(gR)$, followed by a pension of $(R/80)\exp(gR)$ until death D years later. This is where the true actuarial difficulties begin, but we shall avoid this by treating D as a parameter. We use the fund to create an annuity to pay the pension, and a standard argument implies the pension cost

$$P = \frac{3}{2}e^{gR} + \frac{R}{80}e^{gR}\int_0^D e^{-rt}\, dt = e^{gR}\left(\frac{3}{2} + \frac{R}{80}\left[\frac{1 - e^{-rD}}{r}\right]\right). \tag{164}$$

Equating (163) and (164), and dividing by $\exp(gR)$, we obtain

$$c\left(\frac{1 - e^{-(g-r)R}}{g-r}\right) = \frac{3}{2} + \frac{R}{80}\left[\frac{1 - e^{-rD}}{r}\right] \tag{165}$$

or

$$c = \frac{\frac{3}{2} + \frac{R}{80}\left(\frac{1-e^{-rD}}{r}\right)}{\left(\frac{1-e^{-(g-r)R}}{g-r}\right)}. \tag{166}$$

How do we choose the parameters? I have chosen $r = 0.02$ in most experiments, although USS sometimes boasts of much higher yields in the equity investments. The reason for the low value is that much of the USS fund is in UK bonds, whose yield has been near zero since the financial crisis of 2007–2008: it is the near-zero yield, imposed by central banks as part of Quantitative Easing, which is at the root of the problem for all pension funds. You will also see further reasons for this choice in the examples below.

I have provided MATLAB code at the end of this Section for experimentation. As we should all know, the fund imposed a £55K cap on its Defined Benefit scheme in 2016, and this was probably necessary (see the examples), and it's easy to alter the code to include caps. Some of my experiments strongly suggest that £55K is not sustainable while r remains so low, unless contributions increase substantially. The alternative is a lower cap, e.g. £45K, which does seem feasible. Of course, the present decision is to remove the Defined Benefit component entirely in April 2019. The usefulness of caps is further suggested by the next simple result.

Lemma 21.1. *The contribution rate c is an increasing function of the salary increase rate g.*

Proof. If we let $v(g)$ denote the denominator of (166) as a function of the salary increase rate g, then we can write this as the integral

$$v(g) := \frac{1 - e^{-(g-r)R}}{g-r} = \int_0^R e^{-(g-r)s}\, ds.$$

We can now differentiate under the integral sign with respect to g, obtaining

$$\frac{\partial v}{\partial g} = -\int_0^R s e^{-(g-r)s}\, ds < 0.$$

Thus $v(g)$ is a strictly decreasing function of g, which implies that $c(g)$, being a multiple of its reciprocal, is a strictly increasing function of g. □

Here are some examples of the MATLAB code's output for different parameters.

Example 21.1. *For an academic who triples their salary over* 40 *years, here is one result.*

```
Interest rate r = 0.02, service length R = 40 years
Final salary/Initial salary = exp(g*R) = 3
The contribution percentage needed for D = 20 funded years c = 0.281721
The number of funded years for the USS contribution rate is 17.8077
```

Example 21.1 illustrates several USS problems, not least of which is the high required value of c. At present, each month USS members contribute 8% of gross salary, whilst the employer contributes 18%. The choices are to increase r, fund higher contributions, reduce g (effectively capping the salary) or reduce D (effectively ending Defined Benefit, unless our VCs introduce a euthanasia component).

Example 21.2. *We can model a salary cap by imposing* $\exp(gR) = 2$. *If we assume an initial salary of £27K, then this is close to our current salary cap of £55K.*

```
Interest rate r = 0.02, service length R = 40 years
Final salary/Initial salary = exp(g*R) = 2
The contribution percentage needed for D = 20 funded years c = 0.23077
The number of funded years for the USS contribution rate is 23.8243
```

This is better news than Example 21.1.

Example 21.3. *We can model a more stringent salary cap by imposing* $\exp(gR) = 1.7$. *If we assume an initial salary of £27K, then this is close to a salary cap of £45K.*

```
Interest rate r = 0.02, service length R = 40 years
Final salary/Initial salary = exp(g*R) = 1.7
The contribution percentage needed for D = 20 funded years c = 0.212218
The number of funded years for the USS contribution rate is 27.0143
```

Finally, some of our colleagues see *much* larger salary increases during their career. The following example is designed for these successful individuals.

Example 21.4. *Let us consider a highly successful academic who increases their salary from £27K to £400K:*

```
Interest rate r = 0.02, service length R = 40 years
Final salary/Initial salary = exp(g*R) = 14.8148
The contribution percentage needed for D = 20 funded years c = 0.543296
The number of funded years for the USS contribution rate is 6.76152
```

Until the mid-1990s, it was still usual for most academics to remain lecturers throughout their academic careers, with a small minority achieving higher rank. Thus the typical value of g was smaller, say $\exp(gR) = 2$ for $R = 40$, or $g \approx 0.017$. In other words, limited salary progression is good for Defined Benefit schemes based on final salary. For that reason, the SAUL pension fund (Superannuation Arrangements of the University of London) is much more likely to be able to keep funding a Defined Benefit scheme, since it is mostly limited to lower salary ranges.

Example 21.5. *Why have I chosen the relatively small value of* $r = 0.02$ *in my examples. Surely USS is better than that, given the tremendous bonuses paid to their fund managers? Let's try an example:*

```
Interest rate r = 0.04, service length R = 40 years
Final salary/Initial salary = exp(g*R) = 2
The contribution percentage needed for D = 20 funded years c = 0.128724
The number of funded years for the USS contribution rate is 36.2423
```

This would be a world in which USS pension benefits should increase not decrease! Clearly this is not our world, so USS is achieving something in the range $0.01 \leq r \leq 0.03$.

Example 21.6. *As a simple check on our calculations, let us consider the special case when* g *and* r *are tiny. In this case, Taylor expansions of the exponential yield*

$$cR = \frac{3}{2} + \frac{RD}{80},$$

or

$$c = \frac{3}{2R} + \frac{D}{80}.$$

If we consider an academic who retires after $R = 40$ *years of service, then*

$$c = \frac{3 + D}{80}.$$

Thus $D = 20$ *years of retirement require* $c = 23/80 \approx 29\%$ *of gross salary, as expected.*

Finally, on the next page there is the promised MATLAB to calculate c and D.

```
%
% Defined Benefit Pensions
%
% r is the interest rate
%
r=0.02;
%
% R = number of years of service, exp(g*R)=3.0
%
R=40;
%
% Assumption: Final salary/initial salary =  exp(gR)
%
% g=log(3.0)/R;
g=log(2.0)/R;
% g = log(1.7)/R;
%g = log(400/27)/R;
%
% D = number of years in retirement (i.e. before death)
% Employer pays 18% and employee pays 8% of gross salary
%
%
D=20;
c=0.26;
%
printf('Interest rate r = %d, service length R = %d years\n', r, R)
printf('Final salary/Initial salary = exp(g*R) = %d\n', exp(g*R))
%
% F1 is multiplied by c to obtain the actual
% pension fund total
%
F1=(1-exp(-(g-r)*R))/(g-r);
%
% We first find the annuity base cost required
% for D years of retirement, together with the
% corresponding contribution rate.
%
P=1.5 + (R/80)*(1-exp(-r*D))/r;
c1=P/F1;
%
printf('The contribution percentage needed for D = %d funded years c = %d\n', D, c1)
%
% We next compute the number D1 of retirement years
% our fund can actually provide for the USS contribution rate c.
%
F=c*F1;
D1=-(1.0/r)*log(1-r*(F-3/2)/(R/80));
printf("The number of funded years for the USS contribution rate is %d\n", D1)
```

22. Complex Numbers

22.1. **Rotation–Enlargements.** For any $r \in \mathbb{R}$ and $\theta \in \mathbb{R}$, we define the *rotation–enlargement*

$$R(r,\theta) = r \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}, \tag{167}$$

and we let $\mathcal{C}$ denote the set of all rotation-enlargement matrices. If we multiply rotation–enlargement matrices $R(r_1,\theta_1)$ and $R(r_2,\theta_2)$, then we obtain $R(r_1r_2,\theta_1+\theta_2)$, i.e.

$$R(r_1r_2,\theta_1+\theta_2) = R(r_1,\theta_1)R(r_2,\theta_2). \tag{168}$$

Exercise 22.1. *Set $r_1 = r_2 = 1$ in* (168) *and obtain the trigonometric addition formulae for* $\cos(\theta_1+\theta_2)$ *and* $\sin(\theta_1+\theta_2)$.

Lemma 22.1. *The rotation–enlargement matrices $\mathcal{C}$ comprise the matrices of the form*

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}, \tag{169}$$

for $a, b \in \mathbb{R}$,

Proof. Given any rotation–enlargement matrix $R(r,\theta)$, we let $a = R\cos\theta$ and $b = R\sin\theta$. Conversely, given any matrix of the form (169), we let $R = \sqrt{a^2+b^2}$ and define θ by $\cos\theta = a/R$, $\sin\theta = b/R$. □

It is not difficult to check that $\mathcal{C}$ satisfies the axioms for a field. Further, setting

$$\mathbf{1} = R(1,0) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \quad \text{and} \quad \mathbf{J} = R(1,\pi/2) = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \tag{170}$$

we see that every rotation–enlargment matrix can be uniquely written in the form

$$a\mathbf{1} + b\mathbf{J}. \tag{171}$$

Further, the relation

$$\mathbf{J}^2 = R(1,\pi) = -\mathbf{1} \tag{172}$$

implies the multiplication rule

$$(a_1\mathbf{1} + b_1\mathbf{J})(a_2\mathbf{1} + b_2\mathbf{J}) = (a_1a_2 - b_1b_2)\mathbf{1} + (a_1b_2 + b_1a_2)\mathbf{J}. \tag{173}$$

By this point it should come as no surprise that the rotation–enlargment matrices are really the complex numbers in very light disguise.

23. A Geometrical Interpretation of limsup

Let ℓ_∞ denote the vector space of real sequences $\mathbf{x} = (x_1, x_2, \ldots)$ endowed with the norm

$$\|\mathbf{x}\|_\infty = \sup_{k \ge 1} |x_k|. \tag{174}$$

The $(\ell_\infty, \|\cdot\|_\infty)$ is the Banach space of bounded real sequences. It contains the proper closed subspace c_0 of null sequences, i.e. sequences which converge to zero. We shall show that

$$\operatorname{dist}(\mathbf{x}, c_0) = \lim_{n\to\infty} \sup_{k \ge n} |x_k|. \tag{175}$$

Another way of stating (175) is that the quotient space ℓ_∞ / c_0 inherits the quotient norm

$$\|\mathbf{x} + c_0\|_\infty \equiv \operatorname{dist}(\mathbf{x}, c_0) = \lim_{n\to\infty} \sup_{k \ge n} |x_k|. \tag{176}$$

Firstly, given any $\mathbf{x} \in \ell_\infty \setminus c_0$, define the sequence

$$\mathbf{z}^{(n)} = (x_1, x_2, \ldots, x_{n-1}, 0, 0, \ldots) \in c_0, \qquad n \ge 2.$$

Thus

$$\|\mathbf{x} - \mathbf{z}^{(n)}\|_\infty = \sup_{k \ge n} |x_k|$$

and thus

$$\operatorname{dist}(\mathbf{x}, c_0) \le \lim_{n\to\infty} \sup_{k \ge n} |x_k|. \tag{177}$$

Conversely, observe that $\|\mathbf{x}\|_\infty \ge \sup_{k \ge n} |x_k|$, for each positive integer n, which implies the simple inequality

$$\|\mathbf{x}\|_\infty \ge \lim_{n\to\infty} \sup_{k \ge n} |x_k|.$$

Further, $\mathbf{x}$ and $\mathbf{x} - \mathbf{z}$ possess the same convergent subsequences, with the same limits, for any $\mathbf{z} \in c_0$, and a classical result states that $\lim_{n\to\infty} \sup_{k \ge n} |x_k|$ is also the largest real number that is the limit of any convergent subsequence of $\mathbf{x}$. Hence

$$\|\mathbf{x} - \mathbf{z}\|_\infty \ge \lim_{n\to\infty} \sup_{k \ge n} |x_k - z_k| = \lim_{n\to\infty} \sup_{k \ge n} |x_k|. \tag{178}$$

Combining (177) and (178) yields (175).

24. Variation of Parameters

This is my own treatment of an entirely classical technique.

The linear second order ODE

$$y'' + Q(x)y' + R(x)y = S(x) \tag{179}$$

has general solution

$$y = a_1y_1 + a_2y_2 + u, \qquad \text{for any } a_1, a_2 \in \mathbb{R}, \tag{180}$$

where the *complementary functions* (CFs) y_1 and y_2 satisfy the homogeneous equation

$$y'' + Q(x)y' + R(x)y = 0 \tag{181}$$

and the *particular integral* (PI) u satisfies (179). If both linearly independent CFs are known, then we shall show that

$$u(x) = c_1(x)y_1(x) + c_2(x)y_2(x) \tag{182}$$

where

$$c_1(x) = -\int_{x_0}^{x} \frac{y_2(u)S(u)}{W(u)}\,du, \quad c_2(x) = \int_{x_0}^{x} \frac{y_1(u)S(u)}{W(u)}\,du \tag{183}$$

and the *Wronskian* $W(u)$ is defined by

$$W(u) = y_1(u)y_2'(u) - y_1'(u)y_2(u). \tag{184}$$

It is possible to simply substitute (182) in (179) and calculate, but it is more enlightening, in my view, to see some further analysis. The first key point is that the second order ODE (179) in y can be transformed into an equivalent first order ODE in terms of an associated vector function

$$\mathbf{z} = \begin{pmatrix} y \\ y' \end{pmatrix}. \tag{185}$$

Then (179) implies

$$\mathbf{z}' = \begin{pmatrix} y' \\ y'' \end{pmatrix} = \begin{pmatrix} y' \\ S - Qy' - Ry \end{pmatrix} = \mathbf{f}(x) + A(x)\mathbf{z}, \tag{186}$$

where

$$A(x) = \begin{pmatrix} 0 & 1 \\ -R(x) & -Q(x) \end{pmatrix} \quad \text{and} \quad \mathbf{f}(x) = \begin{pmatrix} 0 \\ S(x) \end{pmatrix}. \tag{187}$$

Now the general solution to the homogeneous ODE (181) can be restated in the form

$$\mathbf{z} = Y(x)\mathbf{c}, \tag{188}$$

where

$$Y(x) = \begin{pmatrix} y_1(x) & y_2(x) \\ y_1'(x) & y_2'(x) \end{pmatrix} \tag{189}$$

and $\mathbf{c} \in \mathbb{R}^2$ can be any constant vector. Thus $\mathbf{z}$ satisfies the first order ODE

$$(D - A(x))\,\mathbf{z} = 0, \qquad \text{or equivalently} \qquad Y'(x) = A(x)Y(x), \tag{190}$$

where $D := d/dx$. The idea of Variation of Parameters is to solve (179) in its first order form, i.e.

$$(D - A(x))\,\mathbf{z} = \mathbf{f}, \tag{191}$$

by substituting

$$\mathbf{z} = Y(x)\mathbf{c}(x), \tag{192}$$

where the parameter vector $\mathbf{c}(x)$ is now a function of x, rather than being a constant, hence the name of the method. Thus we obtain

$$\begin{aligned}\mathbf{f} &= (D - A(x))\left(Y(x)\mathbf{c}(x)\right) \\ &= D\left(Y(x)\mathbf{c}(x)\right) - A(x)Y(x)\mathbf{c}(x) \\ &= Y'(x)\mathbf{c}(x) + Y(x)\mathbf{c}'(x) - A(x)Y(x)\mathbf{c}(x) \\ &= \left(Y'(x) - A(x)Y(x)\right)\mathbf{c}(x) + Y(x)\mathbf{c}'(x).\end{aligned}$$

Now

$$Y'(x) - A(x)Y(x) = 0,$$

by (190), so we find the first order ODE

$$Y(x)\mathbf{c}'(x) = \mathbf{f}(x)$$

or

$$\mathbf{c}'(x) = Y(x)^{-1}\mathbf{f}(x). \tag{193}$$

Integrating this first order ODE, we obtain

$$\mathbf{c}(x) = \mathbf{c}(x_0) + \int_{x_0}^{x} Y(u)^{-1}\mathbf{f}(u)\,du. \tag{194}$$

Now it is elementary that the Wronskian $W(u) = \det Y(u)$, so that the inverse matrix $Y(u)^{-1}$ is given by

$$Y(u)^{-1} = \frac{1}{W(u)}\begin{pmatrix} y_2'(u) & -y_2(u) \\ -y_1'(u) & y_1(u)\end{pmatrix}. \tag{195}$$

Hence

(196)

$$Y(u)^{-1}\mathbf{f}(u) = \frac{1}{W(u)}\begin{pmatrix} y_2'(u) & -y_2(u) \\ -y_1'(u) & y_1(u)\end{pmatrix}\begin{pmatrix} 0 \\ S(u)\end{pmatrix} = \frac{1}{W(u)}\begin{pmatrix} -y_2(u)S(u) \\ y_1(u)S(u)\end{pmatrix}$$

Substituting (196) in (194) yields (182) and (183). Further, using (191) we deduce

$$\mathbf{z} = Y(x)\mathbf{c}(x) = Y(x)\mathbf{c}(x_0) + Y(x)\int_{x_0}^{x} Y(u)^{-1}\mathbf{f}(u)\,du. \tag{197}$$

Example 24.1. *Suppose*

$$xy'' + 2y' + xy = 4\cos x$$

and it is easily (if tediously!) checked that the CFs are $y_1(x) = x^{-1}\cos x$ and $y_2(x) = x^{-1}\sin x$. The Wronskian for these two functions is given by $W(x) = x^{-2}$, so the CFs are linearly independent. We must recast the ODE in the form (179)*, that is,*

$$y'' + 2x^{-1}y' + y = 4x^{-1}\cos x,$$

i.e. $Q(x) = 2x^{-1}$, $R(x) = 1$ and $S(x) = 4x^{-1}\cos x$. Then

$$c_1(x) = -4\int_{x_0}^{x} \cos u \sin u\,du = 2\cos^2 x + c_1(x_0)$$

and

$$c_2(x) = 4\int_{x_0}^{x} \cos^2 u\,du = 2\left(x + \sin x\cos x\right) + c_2(x_0).$$

Combining these, and ignoring the constant terms because they correspond to the CFs not the PI, we find

$$u(x) = 2y_1(x)\cos^2 x + 2y_2(x)\left(x + \sin x\cos x\right) = 2\left(\sin x + x^{-1}\cos x\right).$$

Of course, $y_1(x) = x^{-1}\cos x$, so the PI is simply

$$u(x) = 2\sin x.$$

One particularly interesting case is that of constant coefficients, i.e. $Q(x) \equiv Q$ and $R(x) \equiv R$. In this case, we find $A \equiv A(x)$ is a constant matrix and the solution of $Y' = AY$ is the matrix exponential

$$Y(x) = \exp(Ax)\mathbf{c},$$

for some constant vector $\mathbf{c}$.

Example 24.2. *Suppose*

$$y'' + 3y' + 2y = e^{\alpha x},$$

where $\alpha \neq -2, -1$, for reasons which will become apparent. Then a simple calculation provides $y_1(x) = \exp(-2x)$, $y_2(x) = \exp(-x)$ and $W(x) = e^{-3x}$. Hence (183) *yields*

$$c_1(x) = -\int_{x_0}^{x} e^{3u} e^{-u} e^{\alpha u}\, du = -\int_{x_0}^{x} e^{\alpha+2} u\, du = \frac{e^{(\alpha+2)x_0} - e^{(\alpha+2)x}}{\alpha+2},$$

since $\alpha \neq -2$. Further

$$c_2(x) = \int_{x_0}^{x} e^{3u} e^{-2u} e^{\alpha} u\, du = \int_{x_0}^{x} e^{(\alpha+1)u}\, du = \frac{e^{(\alpha+1)x} - e^{(\alpha+1)x_0}}{\alpha+1},$$

since $\alpha \neq -1$. Omitting the CF terms, we find the PI

$$u(x) = c_1(x)y_1(x) + c_2(x)y_2(x) = -\frac{e^{\alpha x}}{\alpha+2} + \frac{e^{\alpha x}}{\alpha+1} = \frac{e^{\alpha x}}{(\alpha+1)(\alpha+2)}.$$

Example 24.3. *More generally, suppose*

$$(D - \lambda_1)(D - \lambda_2) = e^{\alpha x},$$

where $\lambda_1, \lambda_2 \in \mathbb{C}$ and different and $\alpha \in \mathbb{R} \setminus \{\lambda_1, \lambda_2\}$. Then $y_k(x) = \exp(\lambda_k x)$, $k = 1, 2$, and it is easily checked that

$$W(x) = \frac{e^{(\lambda_1+\lambda_2)x}}{\lambda_2 - \lambda_1}.$$

Thus

$$c_1(x) = -\frac{1}{\lambda_2 - \lambda_1} \int_{x_0}^{x} e^{-(\lambda_1+\lambda_2)u} e^{\lambda_2 u} e^{\alpha u}\, du = -\frac{1}{\lambda_2 - \lambda_1} \left(\frac{e^{(\alpha-\lambda_1)x} - e^{(\alpha-\lambda_1)x_0}}{\alpha - \lambda_1} \right)$$

and

$$c_2(x) = \frac{1}{\lambda_2 - \lambda_1} \int_{x_0}^{x} e^{-(\lambda_1+\lambda_2)u} e^{\lambda_1 u} e^{\alpha u}\, du = -\frac{1}{\lambda_2 - \lambda_1} \left(\frac{e^{(\alpha-\lambda_2)x} - e^{(\alpha-\lambda_2)x_0}}{\alpha - \lambda_2} \right).$$

Hence, omitting the CF terms, we find

$$\begin{aligned} u(x) &= e^{\alpha x} \frac{1}{\lambda_2 - \lambda_1} \left(\frac{1}{\alpha - \lambda_2} - \frac{1}{\alpha - \lambda_1} \right) \\ &= \frac{e^{\alpha x}}{(\alpha - \lambda_1)(\alpha - \lambda_2)}. \end{aligned}$$

Heaviside would have simply observed that $D \exp(\alpha x) = \alpha \exp(\alpha x)$ and cheerfully replaced D by α in

$$\frac{e^{\alpha x}}{(D - \lambda_1)(D - \lambda_2)}.$$

25. Bond Pricing

We begin with the fundamental concept of the time value of money.

Example 25.1. *Suppose we need* 100 *in* 5 *years and the interest rate is a constant* 5%. *Then the amount M needed today must satisfy*

$$(1.05)^5 M = 100,$$

or

$$M = 100/(1.05^5) = 78.35.$$

Exercise 25.1. *What amount M should be invested today to obtain* 100 *in* 5 *years if the interest rate is* 2%.

More generally, if the interest rate is r%, then M today will grow to $M(1+r)^n$ after n years: we say its **future value** (FV) of today's M in n years is $M(1+r)^n$. Conversely, if we need D in n years, then we need to invest

$$\frac{D}{(1+r)n}$$

today: we say that the **present value** (PV) of D in n years is $D/(1+r)^n$ and we call r the **discount rate**.

Companies and Governments typically pay today's bills by borrowing via **bonds**. The Bank of England was founded in 1694 to sells bonds for the English Government (which became the British Government in 1707, following the union of Scotland and England). The Bank of England was one of the first central banks (the Swedish Bank was slightly earlier) and bonds rapidly replaced many alternative methods used by states to obtain funds (monopolies, lotteries, subscriptions and taxes were all used before this and still are). What is a bond?

The idea is quite simple: we provide the state with F today, at time $t = 0$. The state promises to pay us a series of payments (or **coupons**), say c each year, for n years, at which point the state returns F to us. We say that F is the **face value** of the bond. One fundamental question therefore arises: what is the PV of the bond?

This is a difficult question in practice, so we shall deal with a simple case, which nevertheless captures the crucial points. We introduce a single "rate" y called the **bond yield**, and the key idea is that

$$PV \text{ (future cash flows discounted by } y) = \text{bond price}. \tag{198}$$

At time $t = 0$, the bond price is F, so (198) gives

$$\frac{c}{1+y} + \frac{c}{(1+y)^2} + \cdots + \frac{c}{(1+y)^n} + \frac{F}{(1+y)^n} = F. \tag{199}$$

In reality, the bond may trade above or below its face value on the **bond market**. In other words, the market price $P(y)$ and the bond yield y are related by

$$P(y) = \frac{c}{1+y} + \frac{c}{(1+y)^2} + \cdots + \frac{c}{(1+y)^n} + \frac{F}{(1+y)^n}. \tag{200}$$

It's important to understand that y is determined by the bond market and the coupon payment c. The key point is that y decreases when $P(y)$ increases, while y increases when $P(y)$ decreases. In other words, a high bond price means a low yield, while a high yield means a low bond price. If the bond is issued by a country which wholly controls its own currency, then it's said to be nominally **risk free**, since the state can always create more money. There is a crucial difference between the **nominal value** and the **real value** (which might be vastly less).

In reality, the coupons can vary and their times of payment need not be annual, but we ignore this here.

Proposition 25.1. *If the bond trades at its face value F, i.e.* (199) *holds, then*

$$F = \frac{c}{y}, \tag{201}$$

that is, the face value F varies inversely with the yield y.

Proof. We have

$$\begin{aligned} F &= \frac{c}{1+y}\left(1 + \frac{1}{1+y} + \frac{1}{(1+y)^2} + \cdots + \frac{1}{(1+y)^{n-1}}\right) + \frac{F}{(1+y)^n} \\ &= \frac{c}{1+y}\left(\frac{1 - \frac{1}{(1+y)^n}}{1 - \frac{1}{1+y}}\right) + \frac{F}{(1+y)^n} \\ &= \frac{c}{y}\left(1 - \frac{1}{(1+y)^n}\right) + \frac{F}{(1+y)^n}. \end{aligned}$$

Rearranging this we obtain

$$F\left(1 - \frac{1}{(1+y)^n}\right) = \frac{c}{1+y}\left(1 - \frac{1}{(1+y)^n}\right),$$

or

$$F = \frac{c}{y}.$$

□

One key point here is that the bond price can be very sensitive to changes in the yield, and conversely. To say more we need some calculus: for any (differentiable) function $f(x)$, we have

$$f(x+h) \approx f(x) + hf'(x)$$

when h is small, and there are excellent techniques for calculating the **derivative** $f'(x)$ originating with Newton in the 17th century. For the relation (201), in the very spacial case when the bond trades at its face value, we have

$$F(y) = \frac{c}{y},$$

and it can be shown that

$$F'(y) = \frac{c}{y^2},$$

so that

$$F(y+h) \approx F(y) + \frac{hc}{y^2}.$$

If y is small, then $1/y^2$ can be enormous.

Example 25.2. *Suppose $c = 1$, $h = 10^{-2}$ and $y = 10^{-2}$. Then*

$$F(y+h) \approx F(y) + \frac{10^{-2}}{10^{-4}} = F(y) + 100.$$

A practical bond price is somewhat more complicated: the bond almost never trades at its face value and the coupons are all different:

$$P(y) = \frac{\alpha_1}{1+y} + \frac{\alpha_2}{(1+y)^2} + \cdots + \frac{\alpha_{n-1}}{(1+y)^{n-1}} + \frac{\alpha_n}{(1+y)^n}. \tag{202}$$

It can be shown that

$$P'(y) = -\frac{\alpha_1}{(1+y)^2} - \frac{2\alpha_2}{(1+y)^3} - \cdots - \frac{n\,(\alpha_n)}{(1+y)^{n+1}}. \tag{203}$$

This might look horrible, but the computer doesn't care, and we can use (203) to estimate bond price sensitivity to a change in yield, via

$$P(y+h) \approx P(y) + hP'(y). \tag{204}$$

In 1938, the economist Macaulay gave an interesting interpretation of (203) which is still used, termed **Macaulay's duration**. Specifically, we let

$$c_i(y) = \frac{\alpha_i}{(1+y)^i}, \qquad 1 \le i \le n, \tag{205}$$

which is the PV cash flow due at year i discounted at rate y. We then define

$$w_i(y) = \frac{c_i(y)}{c_1(y) + \cdots + c_n(y)}. \tag{206}$$

Thus $w_i(y)$ is the PV cash flow due at year i divided by the total bond price, so it's really the discounted PV percentage contribution of the cash flow due at year i.

Exercise 25.2. *Show that*

$$w_1(y) + w_2(y) + \cdots + w_n(y) = 1.$$

Macaulay's **duration** is then defined by

$$D(y) = w_1(y) + 2w_2(y) + 3w_3(y) + \cdots + nw_n(y) \tag{207}$$

and it can be shown that

$$P'(y) = -\left(\frac{D(y)}{1+y}\right) P(y). \tag{208}$$

Economists like (208) because it provides an interpretation they find congenial, but there is no difference if we simply use (203) and (204).

26. Bernstein Polynomials and Jensen's Theorem

Let $f : [0,1] \to \mathbb{R}$ be any continuous function. The Bernstein polynomials are defined by

$$B_n f(t) = \sum_{k=0}^{n} \binom{n}{k} f(k/n) t^k (1-t)^{n-k}, \qquad \text{for } 0 \le t \le 1 \text{ and } n \in \mathbb{N}. \tag{209}$$

Equivalently, if $X_1, X_2, \ldots$ are independent Bernoulli random variables satisfying

$$\mathbb{P}(X_k = 1) = t \text{ and } \mathbb{P}(X_k = 0) = 1 - t, \tag{210}$$

for all non-negative integer n and $t \in [0,1]$, then

$$B_n f(t) = \mathbb{E} f\left(\frac{X_1 + X_2 + \cdots + X_n}{n}\right). \tag{211}$$

This well-known probabilistic interpretation was at the heart of Bernstein's definition in 1912. It is easily checked that $B_n f(t) \equiv 1$ when $f(t) \equiv 1$ and $B_n f(t) \equiv t$ when $f(t) \equiv t$. Thus the Bernstein polynomial operator $f \mapsto B_n f$ is linear and preserves linear polynomials. Hence, we can always adjust f to satisfy

$$f(0) = f(1) = 0 \tag{212}$$

by addition of a linear polynomial; more formally, we observe that

$$g(t) := f(t) - f(0) - (f(1) - f(0))\, t$$

satisfies $g(0) = g(1) = 0$.

Theorem 26.1. *Let $f : [0,1] \to \mathbb{R}$ be any convex continuous function satisfying* (212)*. Then*

$$0 = B_0 f(t) = B_1 f(t) \ge B_2 f(t) \ge \cdots \ge B_n f(t) \ge B_{n+1} f(t) \ge \cdots \ge f(t). \tag{213}$$

The novelty of the proof is that it depends on Jensen's inequality: for any real-valued random variable Y and convex function $\phi : \mathbb{R} \to \mathbb{R}$, we have

$$\phi(\mathbb{E}X) \le \mathbb{E}(\phi(X)). \tag{214}$$

Proof. We have

$$\begin{aligned} B_{n+1} f(t) &= \mathbb{E} f\left(\frac{X_1 + \cdots + X_{n+1}}{n+1}\right) \\ &= \mathbb{E} f\left(\left(\frac{n}{n+1}\right)\left(\frac{X_1 + \cdots + X_n}{n}\right) + \left(\frac{1}{n+1}\right) X_{n+1}\right) \\ &\le \left(\frac{n}{n+1}\right) \mathbb{E} f\left(\frac{X_1 + \cdots + X_n}{n}\right) + \left(\frac{1}{n+1}\right) \mathbb{E} f(X_{n+1}) \\ &= \left(\frac{n}{n+1}\right) B_n f(t) \\ &\le B_n f(t). \end{aligned}$$

Further, since f is continuous, we must have $B_n f \to f$ uniformly in $[0,1]$, which implies the lower bound of f. □

Incidentally, a convex function $f : (0,1) \to \mathbb{R}$ is automatically continuous, which proof I will probably add to these notes at some point, but we must impose continuity at the endpoints. Consider, for example, the discontinuous convex function f which is identically zero in $(0,1)$ and satisfies $f(0) = f(1) = 1$.

27. The Economics of Naked Wine

At the time of writing, the wine merchant Naked Wine (NW) provided an interesting discount offer: for every £1 spent, NW deposits a cashback of £1/3 in your NW account, which can only be spent on further wine orders. At first sight, this seems to be a discount of one third, but further analysis is interesting.

Let's consider a slightly more general analysis, assuming a cashback of $£p$ for every £1 spent, where $p \in [0,1]$ is constant. We suppose that at the k^{th} purchase we use the existing NW account balance V_{k-1} and add a further $£M_k$. After the k^{th} purchase the new NW balance after cashback is $V_k = pM_k$; of course, $V_0 = 0$. If we let T_n denote our total spend after n purchases, then the total value of wine bought after n purchases is $T_n + pT_{n-1}$. Thus

$$P_n := \frac{\text{Total sterling spent after } n \text{ purchases}}{\text{Value of wine after } n \text{ purchases}} = \frac{T_n}{T_n + pT_{n-1}} = \frac{1}{1 + p\left(\frac{T_{n-1}}{T_n}\right)}$$

and the corresponding discount is given by

$$D_n = 1 - P_n = \frac{p\left(\frac{T_{n-1}}{T_n}\right)}{1 + p\left(\frac{T_{n-1}}{T_n}\right)}.$$

Now $T_{n-1}/T_n \le 1$, so

$$P_n \ge \frac{1}{1+p}$$

and

$$D_n \le \frac{p}{1+p}.$$

Example 27.1. *If* $p = 1/3$*, the NW value, then the discount*

$$D_n \le \frac{1/3}{1 + (1/3)} = \frac{1}{4},$$

i.e. the discount is at most 25%.

If we make the same purchase amount every time, then $T_n = na$, say, and $T_{n-1}/T_n = 1 - n^{-1} \to 1$, as $n \to \infty$. Thus $P_n \to 1/(1+p)$ and the discount satisfies $D_n \to p/(1+p)$, i.e. the upper bound on the discount is sharp.

Example 27.2. *What happens if our purchases grow algebraically, i.e.*

$$T_n = Cn^\alpha,$$

for some positive constants C *and* α*. The*

$$\frac{T_{n-1}}{T_n} = \left(\frac{n-1}{n}\right)^\alpha = \left(1 - \frac{1}{n}\right)^\alpha \to 1,$$

as $n \to \infty$*. Hence* $D_n \to p/(1+p)$ *for algebraically growing purchases too.*

Now consider true wine thirst: exponential purchase growth! If

$$T_n = Ce^{\alpha n},$$

for positive constants C and α, then

$$\frac{T_{n-1}}{T_n} = e^{-\alpha} < 1.$$

Thus

$$D_n = \frac{pe^{-\alpha}}{1 + pe^{-\alpha}}.$$

Example 27.3. *If $e^\alpha = 2$ and $p = 1/3$, i.e. our wine purchases double every month, then*

$$D_n = \frac{p/2}{1+p/2} = \frac{1/6}{7/6} = \frac{1}{7},$$

or a constant discount of about 14%. *If we had taken the more modest value of $e^\alpha = 1.1$, then*

$$D_n = \frac{1/3.3}{1+1/3.3} = 0.23256\ldots.$$

Thus higher exponential growth reduces *the discount.*

The reduced discount for higher exponential growth might seem a curiosity, since few individuals could afford exponential growth, whether in currency or health consequences. However, if we consider the discount afforded to *all* NW customers, then NW was reporting 40% growth in turnover in the mid-2010s. If we take $e^\alpha = 1.4$ and $p = 1/3$, then the discount is given by

$$D_n = \frac{1/4.2}{1+1/4.2} = 0.1923\ldots,$$

i.e. a discount of roughly 19%.

28. Lagrange's Identity

Theorem 28.1. *Let $(a_k)_1^n$ and $(b_k)_1^n$ be real sequences. Then*

$$\left(\sum_{k=1}^n a_k b_k\right)^2 = \left(\sum_{k=1}^n a_k^2\right)\left(\sum_{\ell=1}^n a_\ell^2\right) - \sum_{1\le k<j\le n} \left(a_k b_j - a_j b_k\right)^2. \tag{215}$$

The key observation is that

$$\sum_{1\le k<j\le n} \left(a_k b_j - a_j b_k\right)^2 = \frac{1}{2}\sum_{i,j=1}^n \left(a_i b_j - a_j b_i\right)^2$$

and

$$\begin{aligned}\frac{1}{2}\sum_{i,j=1}^n \left(a_i b_j - a_j b_i\right)^2 &= \frac{1}{2}\sum_{i,j=1}^n a_i^2 b_j^2 - 2a_i b_j a_j b_j + a_j^2 b_i^2 \\ &= \|a\|^2\|b\|^2 - (a^T b)^2.\end{aligned}$$

However, it is also possible to see this via the Frobenius norm. We define the skey-symmetric matrix $M \in \mathbb{R}^{n\times n}$

$$M_{jk} = a_j b_k - a_k b_j, \qquad 1 \le j, k \le n,$$

that is,

$$M = \mathbf{a}\mathbf{b}^T - \mathbf{b}\mathbf{a}^T.$$

Recall that the Frobenius inner product is defined by

$$\langle A, B\rangle_F = \sum_{j=1}^n \sum_{k=1}^n A_{jk} B_{jk}.$$

Further,

$$\operatorname{trace} AB^T = \sum_{j=1}^n \left(AB^T\right)_{jj} = \sum_{j=1}^n \sum_{k=1}^n A_{jk} B_{jk} = \langle A, B\rangle_F$$

and, for any vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^n$,

$$\operatorname{trace} \mathbf{u}\mathbf{v}^T = \mathbf{u}^T\mathbf{v}.$$

Hence

$$\begin{aligned}\|\mathbf{a}\mathbf{b}^T - \mathbf{b}\mathbf{a}^T\|_F^2 &= \operatorname{trace}\left[(\mathbf{a}\mathbf{b}^T - \mathbf{b}\mathbf{a}^T)(\mathbf{b}\mathbf{a}^T - \mathbf{a}\mathbf{b}^T)\right] \\ &= -\operatorname{trace}\left[(\mathbf{a}\mathbf{b}^T - \mathbf{b}\mathbf{a}^T)^2\right] \\ &= -\operatorname{trace} \mathbf{a}\mathbf{b}^T\mathbf{a}\mathbf{b}^T + 2\operatorname{trace} \mathbf{a}\mathbf{b}^T\mathbf{b}\mathbf{a}^T - \operatorname{trace} \mathbf{b}\mathbf{a}^T\mathbf{b}\mathbf{a}^T \\ &= -(\mathbf{b}^T\mathbf{a})\operatorname{trace} \mathbf{a}\mathbf{b}^T + 2\|\mathbf{b}\|^2 \operatorname{trace} \mathbf{a}\mathbf{a}^T - (\mathbf{a}^T\mathbf{b})\operatorname{trace} \mathbf{b}\mathbf{a}^T \\ &= 2\left(\|\mathbf{a}\|^2\|\mathbf{b}\|^2 - (\mathbf{a}^T\mathbf{b})^2\right).\end{aligned}$$

29. Sums of Powers of Integers

You have probably all seen the formulae

$$S_1(n) := \sum_{k=0}^{n} k = \frac{1}{2}n(n+1)$$

and

$$S_2(n) := \sum_{k=0}^{n} k^2 = \frac{1}{6}n(n+1)(2n+1)$$

but where do they come from? The answer lies in a fascinating borderland between series and the origins of calculus. We define

$$S_p(n) = \sum_{k=0}^{n} k^p,$$

for any non-negative integers n and p.

Definition 29.1. *The* **forward difference operator** Δ *is defined by*

$$\Delta a_n = a_{n+1} - a_n.$$

Example 29.1. *Here are some simple properties of* Δ.

(i) *If* $a_n = c$, *for all* n, *then* $\Delta a_n = 0$.
(ii) *If* $a_n = n$, *then* $\Delta a_n = n + 1 - n = 1$.
(iii) $\Delta n^2 = (n+1)^2 - n^2 = 2n + 1$
(iv) $\Delta n^3 = (n+1)^3 - n^3 = 3n^2 + 3n + 1$.

Example 29.2 ($S_1(n) = \sum_{k=0}^{n} k$)**.** *For* $S_1(n) = 0 + 1 + 2 + \cdots + n$, *we have* $\Delta S_1(n) = S_1(n+1) - S_1(n) = n + 1$. *Then*

$$\Delta\left(S_1(n) - An^2 - Bn\right) = 0,$$

if

$$n + 1 - A(2n+1) - B = 0, \qquad \textit{for all } n,$$

or $A = B = 1/2$. *Now* $\Delta(S_1(n) - An^2 - Bn) = 0$ *implies that* $S_n - An^2 - Bn = c$, *for some constant* c, *but setting* $n = 0$ *implies* $c = 0$. *Further, note that*

$$\begin{pmatrix} 2 & 0 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} A \\ B \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Theorem 29.1 ($S_2(n) = \sum_{k=0}^{n} k^2$)**.** *We have*

$$S_2(n) = 0^2 + 1^2 + 2^2 + \cdots + n^2 = \frac{1}{3}n^3 + \frac{1}{2}n^2 + \frac{1}{6}n.$$

Proof. Now

$$\Delta S_2(n) = S_2(n+1) - S_2(n) = (n+1)^2 = n^2 + 2n + 1.$$

We must therefore find constants P, Q and R for which

$$\Delta\left(Pn^3 + Qn^2 + Rn\right) = 2n^2 + 2n + 1,$$

i.e. equating coefficients of powers of n, we have

$$P(3n^2 + 3n + 1) + Q(2n+1) + R = n^2 + 2n + 1.$$

Hence $3P = 1$, $3P + 2Q = 2$ and $P + Q + R = 1$, i.e.

$$\begin{pmatrix} 3 & 0 & 0 \\ 3 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix}\begin{pmatrix} P \\ Q \\ R \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 1 \end{pmatrix}.$$

□

Example 29.3 ($S_3(n) = \sum_{k=0}^n k^3$)**.** *For $S_3(n) = \sum_{k=0}^n k^3$ we have*

$$\Delta S_3(n) = (n+1)^3 = n^3 + 3n + 3n + 1$$

and we need constants $A_1, A_2, A_3, A_4 \in \mathbb{R}$ for which

$$\Delta \left(A_4 n^4 + A_3 n^3 + A_2 n^2 + A_1 n \right) = n^3 + 3n^2 + 3n + 1.$$

Now

$$LHS = A_4(4n^3 + 6n^2 + 4n + 1) + A_3(3n^2 + 3n + 1) + A_2(2n + 1) + A_1$$

so that, equating coefficients of powers of n,

$$\begin{aligned} 4A_4 &= 1, \\ 6A_4 + 3A_3 &= 3, \\ 4A_4 + 3A_3 + 2A_2 &= 3, \\ A_4 + A_3 + A_2 + A_1 &= 1. \end{aligned}$$

or

$$\begin{pmatrix} 4 & 0 & 0 & 0 \\ 6 & 3 & 0 & 0 \\ 4 & 3 & 2 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} A_4 \\ A_3 \\ A_2 \\ A_1 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \\ 3 \\ 1 \end{pmatrix}.$$

Exercise 29.1. *Hence show that*

$$S_3(n) = \frac{1}{4} n^4 + \frac{1}{2} n^3 + \frac{1}{4} n^2 = S_1(n)^2.$$

Sadly the beautiful fact that $S_3(n) = S_1(n)^2$ does not extend to higher sums of powers but the binomial pattern is clear. Here's the linear system for $S_4(n)$:

$$\begin{pmatrix} 5 & 0 & 0 & 0 & 0 \\ 10 & 4 & 0 & 0 & 0 \\ 10 & 6 & 3 & 0 & 0 \\ 5 & 4 & 3 & 2 & 0 \\ 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} A_4(5) \\ A_4(4) \\ A_4(3) \\ A_4(2) \\ A_4(1) \end{pmatrix} = \begin{pmatrix} 1 \\ 4 \\ 6 \\ 4 \\ 1 \end{pmatrix}.$$

Theorem 29.2. *We have*

$$S_m(n) = \sum_{j=1}^{m+1} A_m(j) n^j, \tag{216}$$

where

$$\sum_{j=1}^{m+1} A_m(j) \sum_{k=0}^{j-1} \binom{j}{k} n^k = \sum_{\ell=0}^{m} \binom{m}{\ell} n^\ell. \tag{217}$$

In particular, $(m+1)A_m(m+1) = 1$ and $\sum_{j=1}^{m+1} A_m(j) = 1$.

Proof. We use the facts that

$$\Delta S_m(n) = (n+1)^m = \sum_{\ell=0}^{m} \binom{m}{\ell} n^\ell$$

and

$$\Delta n^j = (n+1)^j - n^j = \sum_{k=0}^{j-1} \binom{j}{k} n^k.$$

We must solve

$$\Delta \Big(A_m(m+1) n^{m+1} + A_m(m) n^m + \cdots + A_m(2) n^2 + A_m(1) n \Big) = \Delta S_m(n),$$

so

$$\sum_{j=1}^{m+1} A_m(j)\Delta n^j = \Delta S_m(n),$$

which provides (217). Equating the coefficient of n^m and the constant term provides $A_m(m+1)$ and the fact that the coefficients sum to unity. □

Now (217) is true for any non-negative integer n and both sides are polynomials of degree m in n. Thus we must have the polynomial identity

$$\sum_{j=1}^{m+1} A_m(j) \sum_{k=0}^{j-1} \binom{j}{k} z^k = \sum_{\ell=0}^{m} \binom{m}{\ell} z^\ell, \qquad \text{for all } z \in \mathbb{C}. \tag{218}$$

In other words we have

$$\sum_{j=1}^{m+1} A_m(j) \Big[(1+z)^j - z^j\Big] = \Big(1+z\Big)^m, \qquad z \in \mathbb{C}. \tag{219}$$

If we differentiate (219), then we obtain

$$\sum_{j=2}^{m+1} jA_m(j) \Big[(1+z)^{j-1} - z^{j-1}\Big] = m\Big(1+z\Big)^{m-1}.$$

Setting $k = j-1$ in the sum on the LHS, we obtain

$$\sum_{k=1}^{m} \Big(\frac{(k+1)A_m(k+1)}{m}\Big) \Big[(1+z)^k - z^k\Big] = \Big(1+z\Big)^{m-1}. \tag{220}$$

Comparing (219) and (220) when m is reduced by 1, we see that

$$A_{m-1}(k) = \frac{(k+1)A_m(k+1)}{m}, \qquad \text{for } 1 \le k \le m. \tag{221}$$

Conversely, we have

(222)

$$A_m(\ell) = \frac{m}{\ell} A_{m-1}(\ell), \qquad \text{for } \ell = 2, \ldots, m+1, \qquad \text{and } A_m(1) = 1 - \sum_{\ell=2}^{m+1} A_m(\ell).$$

Thus we can generate all of them using $A_1(1) = A_1(2) = 1/2$. Here's the MATLAB code to do just that:

```
n=10;
A=zeros(n,n+1);
A(1,1)=0.5; A(1,2)=0.5;
for m=2:n
  for l=2:m+1
    A(m,l)=(m/l)*A(m-1,l-1);
  end
  A(m,1) = 1 - sum(A(m,2:m+1));
end
```

30. Hamiltonians in Mathematical Economics

We discuss Example 11.3 in [2]: we wish to minimize the cost functional

$$J = \int_0^T (Q - I)^2 + \alpha^2 P^2 \, dt,$$

where Q and α are positive constants, $I_0 \equiv I(0)$ and $Q > I_0$. Here $I(t)$ is the inventory level and is the **state** variable, while $P(t)$ is the production level and is the **control** variable, which are related by $\dot{I} = P$.

The usual calculus of variations solution is to define the **augmented functional**

$$J^* = \int_0^T (Q - I)^2 + \alpha^2 P^2 + \lambda(\dot{I} - P) \, dt \tag{223}$$

and we let

$$F = (Q - I)^2 + \alpha^2 P^2 + \lambda(\dot{I} - P) \tag{224}$$

be the integrand. The Euler equations are then given by the state equation (I)

$$\frac{\partial F}{\partial I} - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{I}}\right) = 0 \tag{225}$$

and the control equation (P)

$$\frac{\partial F}{\partial P} - \frac{d}{dt}\left(\frac{\partial F}{\partial \dot{P}}\right) = 0. \tag{226}$$

Thus (226) implies

$$\lambda(t) = 2\alpha^2 P(t) \tag{227}$$

and (225) yields

$$-2(Q - I) - \dot{\lambda} = 0$$

or

$$\dot{\lambda} = -2(Q - I). \tag{228}$$

Hence differentiating (227) provides

$$\dot{\lambda} = 2\alpha^2 \dot{P} = 2\alpha^2 \ddot{I} \tag{229}$$

and substituting (229) in (228) gives

$$2\alpha^2 \ddot{I} = -2(Q - I)$$

or

$$\ddot{I} - \alpha^{-2} I = -\alpha^{-2} Q \tag{230}$$

with general solution

$$I(t) = Q + C\cosh(t/\alpha) + D\sinh(t/\alpha). \tag{231}$$

The **transversality condition** is

$$\frac{\partial F}{\partial \dot{I}} \mid_{t=T} = 0 \tag{232}$$

i.e. $\lambda(T) = 0$. We shall complete the solution below.

Now Mathematical Economics uses a slightly different formalism borrowed from Hamiltonian mechanics. We define the Hamiltonian

$$H = \lambda \dot{I} - F \tag{233}$$

i.e.

$$F = -H + \lambda \dot{I}. \tag{234}$$

The Euler equations then become

$$\frac{\partial H}{\partial P} = 0 \quad \text{and} \quad \frac{\partial H}{\partial I} = -\dot{\lambda}. \tag{235}$$

If we use these Hamiltonian equations, then we obtain

$$H = -(Q - I)^2 - \alpha^2 P^2 + \lambda P \tag{236}$$

where

$$0 = \frac{\partial H}{\partial P} = -2\alpha^2 P + \lambda \tag{237}$$

and

$$-\dot{\lambda} = \frac{\partial H}{\partial I} = 2(Q - I). \tag{238}$$

Hence (237) implies

$$P(t) = \frac{\lambda(t)}{2\alpha^2}. \tag{239}$$

Now $\dot{I} = P$ and (239) yield

$$-\dot{\lambda} = 2\alpha^2 \dot{P} = 2\alpha^2 \ddot{I} \tag{240}$$

and then (238) provides

$$\ddot{I} + \alpha^{-2} I = \alpha^{-2} Q. \tag{241}$$

The general solution is (231), of course, and the condition $I(0) = I_0 < Q$ becomes $I_0 = Q + C$, so that C is negative. We write

$$I(t) = Q - (Q - I_0)\cosh(t/\alpha) + D\sinh(t/\alpha). \tag{242}$$

How do we find D? The transversality condition

$$\frac{\partial F}{\partial \dot{I}} = 0$$

then becomes $\lambda(T) = 0$, or $P(T) = 0$, i.e. $\dot{I}(T) = 0$. Hence

$$0 = \dot{I}(T) = -(Q - I_0)\alpha^{-1}\sinh(T/\alpha) + (D/\alpha)\cosh(T/\alpha) \tag{243}$$

or

$$D = (Q - I_0)\tanh(T/\alpha). \tag{244}$$

Then

$$\begin{aligned} I(t) &= Q - (Q - I_0)\cosh(t/\alpha) + (Q - I_0)\tanh(T/\alpha)\sinh(t/\alpha) \\ &= Q - \frac{Q - I_0}{\cosh(T/\alpha)}\left[\cosh(t/\alpha)\cosh(T/\alpha) + \sinh(T/\alpha)\sinh(t/\alpha)\right] \\ &= Q - (Q - I_0)\frac{\cosh((T - t)/\alpha)}{\cosh(T/\alpha)}. \end{aligned}$$

References

[1] Beardon (2005), *Algebra and Geometry*, CUP.

[2] D. N. Burghes and A. M. Downs (1977), em Modern Introduction to Classical Mechanics and Control, Ellis Horwood, Chichester.

[3] W. F. Donoghue (1970), *Monotone Matrix Functions and Analytic Continuation*, Springer.

[4] R. A. Horn and C. R. Johnson (1990), *Matrix Analysis*, CUP.

[5] S. M. Roman (2005), *The Umbral Calculus*, Dover

[6] S. M. Roman and G.-C. Rota (1978), "The Umbral Calculus", *Advances in Mathematics* **27**: 95–120.

[7] D. Zeilberger (2005), "An Umbral Approach to the Hankel Transform for Sequences", *Personal Journal of Ekhad and Zeilberger* `http://www.math.rutgers.edu/~zeilberg/pj.html`.