CHAPTER 15

Solutions to selected exercises

Exercise 2.1: The concentration profile is decreasing with x, so the diffusive flux is from left to right, i.e., positive. The slope is decreasing in magnitude, i.e., the curve is steeper at $x = a$ than at $x = b$, so $J(a)$ is larger than $J(b)$. So there is a net influx into the interval $[a, b]$, and the amount of material increases.

More generally, the concentration profile is convex, i.e., $C'' > 0$, so from (2.4) we see that $\dot{C}$ is positive; the concentration increases everywhere in the region plotted.

Exercise 2.2: The verification is most easily done with a computer algebra system such as `Maple`, `Mathematica`, or `sage`. The following piece of `sage` does the verification:

```
var('x t D x0')
phi(x) = 1/sqrt(2*pi)*exp(-x^2/2)
C(x,t) = phi((x-x0)/sqrt(2*D*t))/sqrt(2*D*t)
res = factor(diff(C(x,t),t) - D*diff(C(x,t),x,x))
print(res)
```

The `factor` is there to simplify the expression. When run, the code will output the result 0 which verifies that the left and right hand side of the equation (2.5) agree.

The initial condition is satisfied in the sense that the solution $C(x, t)$ converges to a Dirac delta as $t \searrow 0$ in the weak sense. That is, if $f(x)$ is a continuous bounded function, then $\int f(x)C(x, t)\, dx \to f(x_0)$ as $t \to 0$.

Exercise 2.3: We define the diffusive length scale L as the standard deviation in the plume, i.e., $\sqrt{2DT}$ where T is the time scale of interest. We get:

Process	Diffusivity	1 sec	1 minute	1 hour	1 day
Salt in water at 293 K	$1 \cdot 10^{-9}$	$4.5 \cdot 10^{-5}$	$3.5 \cdot 10^{-4}$	$2.7 \cdot 10^{-3}$	$1.3 \cdot 10^{-2}$
Smoke in air at 293 K	$2 \cdot 10^{-5}$	$6.3 \cdot 10^{-3}$	$4.9 \cdot 10^{-2}$	$3.8 \cdot 10^{-1}$	$1.9 \cdot 10^{0}$
Carbon in iron at 1250 K	$2 \cdot 10^{-11}$	$6.3 \cdot 10^{-6}$	$4.9 \cdot 10^{-5}$	$3.8 \cdot 10^{-4}$	$1.9 \cdot 10^{-3}$
	m^2/s	m	m	m	m

Note that these length scales are all quite small, by everyday measures.

Exercise 2.4: We have

$$\dot{C} = -\lambda C \text{ and } C'' = -k^2 C$$

Combining, we get $\dot{C} = \lambda/k^2 \ C''$ which, with $D = \lambda/k^2$, agrees with the diffusion equation governing for C.

Exercise 2.7: We show that the bounds hold. Consider first the upper bound $1 - \Phi(x) \leq \phi(x)/x$. This trivially holds for $x \to \infty$, as both sides of the inequality converge to 0. The differential version of the inequality reads

$$-\phi(x) \geq -\phi(x) - \phi(x)/x^2$$

which obviously holds. Thus, the conclusion follows. Next, for the lower bounds, we repeat the procedure. The differential version of the inquality is

$$\left[\frac{-x^2}{1+x^2} + \frac{1+x^2-2x^2}{(1+x^2)^2}\right]\phi(x) \geq -\phi(x)$$

or

$$\frac{x^4+2x^2-1}{(1+x^2)^2}\phi(x) \leq \phi(x)$$

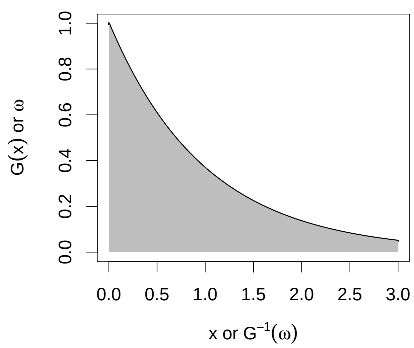
which obviously holds. The conclusion follows.

Exercise 3.1: Write $I := \{\mathbf{E}X_s : X_s \text{ simple}, 0 \leq X_s \leq X\}$. Since X is non-negative, it is bounded below by the simple random variable $X_s = 0$, for which $\mathbf{E}X_s = 0$, so $0 \in I$. Clearly, all $X_s \geq 0$ have $\mathbf{E}X_s \geq 0$. Finally, assume that $X_s \leq X$ and $\alpha \in [0,1]$; then also $\alpha X_s \leq X$. Therefore, if $x \in I$, then also $\alpha x \in I$. The result follows.

Exercise 3.2: We have $X(\omega) > x \Leftrightarrow \omega < G(x)$ and hence

$$\mathbf{P}(X > x) = \mathbf{P}(\omega < G(x)) = G(x)$$

since $G(\cdot)$ is decreasing. To see that $\mathbf{E}X = \int_0^\infty G(x)\,dx$, consider the following figure, based on the specific example $G(x) = \exp(-x)$:



By definition,

$$\mathbf{E}X = \int_\Omega X(\omega)\ \mathbf{P}(d\omega) = \int_0^1 G^{-1}(\omega)\ d\omega$$

which corresponds to the gray shaded area. Finding the area of this set by integrating along the abscissa (using Fubini's theorem), we see that

$$\mathbf{E}X = \int_0^\infty G(x)\ dx.$$

Note that this result is standard, but is usually verified by integration by parts.

The extension to the case where G is allowed to jump, and to be constant on subintervals, follows with a careful analysis and what happens at the end points of these intervals.

Exercise 3.4: If X is $\mathcal{H}$-measurable, then also X^2 is $\mathcal{H}$-measurable. We get

$$\begin{aligned}\mathbf{V}\{X|\mathcal{H}\} &= \mathbf{E}\{X^2|\mathcal{H}\} - (\mathbf{E}\{X|\mathcal{H}\})^2 \\ &= X^2 - X^2 = 0\end{aligned}$$

Exercise 3.5:

$$\begin{aligned}\mathbf{V}X &= \mathbf{E}X^2 - (\mathbf{E}X)^2 \\ &= \mathbf{E}\mathbf{E}\{X^2|\mathcal{H}\} - (\mathbf{E}\mathbf{E}\{X|\mathcal{H}\})^2 \\ &= \mathbf{E}\mathbf{V}\{X|\mathcal{H}\} + \mathbf{E}(\mathbf{E}\{X|\mathcal{H}\})^2 - (\mathbf{E}\mathbf{E}\{X|\mathcal{H}\})^2 \\ &= \mathbf{E}\mathbf{V}\{X|\mathcal{H}\} + \mathbf{V}\mathbf{E}\{X|\mathcal{H}\}\end{aligned}$$

Exercise 3.6: By the Law of Total Expectation we have

$$\mathbf{E}Y = \mathbf{E}\mathbf{E}\{Y|N\} = \mathbf{E}\mu N = \mu\lambda.$$

If you are skeptical about the claim $\mathbf{E}\{Y|N\} = \mu N$, then write $Y = \sum_{i=1}^{\infty} X_i \cdot \mathbf{1}\{N \geq i\}$. For the variance, we use the Law of Total Variance, and independence of X_i, to find

$$\mathbf{V}Y = \mathbf{E}\mathbf{V}\{Y|N\} + \mathbf{V}\mathbf{E}\{Y|N\} = \mathbf{E}N\sigma^2 + \mathbf{V}\mu N = (\mu^2 + \sigma^2)\lambda.$$

Again, if you are skeptical about the claim $\mathbf{V}\{Y|N\} = \sigma^2 N$, then write $Y = \sum_{i=1}^{\infty} X_i \cdot \mathbf{1}\{N \geq i\}$.

Exercise 3.7: If we just want to show that $\mathbf{E}|X|^q < \infty$, note that $|x|^q \leq 1 + |x|^p$ for any $x \in \mathbf{R}$, so $\mathbf{E}|X|^q < 1 + \mathbf{E}|X|^p < \infty$.

For the stronger result $\|X\|_q \leq \|X\|_p$, set $Y = |X|^q$ and $Z = |X|^p$. Then $Z = g(Y)$ with $g(y) = y^{p/q}$. The function g is convex so Jensen's inequality gives

$$\mathbf{E}Z \geq g(\mathbf{E}Y)$$

or $\mathbf{E}|X|^p \geq (\mathbf{E}|X|^q)^{p/q}$, from which the result follows.

Equality is obtained, for example, when $X = 1$ w.p. 1. An example where $\|X\|_q \ll \|X\|_p$ is obtained when X is a Bernoulli variable with probability parameter r. Then $\|X\|_p = r^{1/p}$, $\|X\|_q = r^{1/q}$ and the ratio $\|X\|_q/\|X\|_p \to 0$ as $r \to 0$.

Exercise 3.8:

1. Let X be standard Gaussian. Let I be an independent Bernoulli variable, taking values 0 and 1 with probability 1/2, independently of X. Define $Y = X(2I - 1)$.

2. We have $\mathbf{E}XY = \mathbf{E}\mathbf{E}\{XY|X\} = \mathbf{E}X\ \mathbf{E}Y$ under the conditions. Next, let X be uniform on $[-1, 1]$ and define $Y = |X|$. In greater generality, let X be symmetrically distributed around 0 and let $Y = g(X)$ for some even function g.

Exercise 3.10: It suffices to show this for $c = 0$. The idea is to start with rectangular (box) sets in the plane and construct the set A using countably many set operations. So define

$$B_n^m = \{(x, y) : x > m/n,\ y > -m/n\}$$

for natural n and integer m. You should sketch B_n^m, if you can't visualize it directly. Clearly B_n^m is Borel. Now define $B = \cup_{n \in \mathbf{N}} \cup_{m \in \mathbf{Z}} B_n^m$, then B is also Borel. Now it is easy to see that $(x, y) \in B$ if and only if $x + y > 0$. If you don't agree, then you should set out to find a n, m so that $(x, y) \in B_n^m$, for given (x, y) such that $x + y > 0$. So $A = \mathbf{R}^2\ B$. Hence A is also Borel.

Exercise 3.11: First, note that Z follows the same distribution as X. To see that X and Z are independent, we have $\mathbf{P}(X = 1, Z = 1) = \mathbf{P}(X = 1, Y = 1) = 1/4 = \mathbf{P}(X = 1)\mathbf{P}(Z = 1)$; similarly for the other combinations. The same argument yields that Y and Z are independent. Finally, we have $P(X = 1, Y = 1, Z = -1) = 0$ but $P(X = 1)\mathbf{P}(Y = 1)\mathbf{P}(Z = -1) = 1/8$.

Exercise 3.12:

1. For $s > 0$, we have $\mathbf{P}(S > s) = \mathbf{P}(\{\omega : \omega_1 < \exp(-s/2)\}) = \exp(-s/2)$.
2. First, brute force: The p.d.f. of (S, Θ) is $f_{S\Theta}(s, \theta) = \exp(-s/2)/(4\pi)$. We have $f_{XY}(x, y) = f_{S\Theta}(s, \theta)/|J|$ where J is the Jacobian of the map from (s, θ) to (x, y). This determinant is constant and equal to 1/2, which can be verified with standard multivariate calculus. Thus $f_{XY}(x, y) = \exp(-(x^2 + y^2)/2)/(2\pi)$.

   Alternatively, since the map $(s, \theta) \mapsto (x, y)$ is one-to-one, there is a one-to-one mapping between the distribution of (S, Θ) and that of (X, Y). This mapping is in fact given by the calculations in the previous, but we do not need the specifics, only the existence. Next, if (X, Y) are i.i.d. and standard Gaussian, then $S = X^2 + Y^2$ is $\chi^2(2)$-distributed; this is also an exponential distribution with mean 2. The angle Θ is uniformly distributed due to the joint density of (X, Y) being invariant under rotations. We conclude that (X, Y) are jointly standard Gaussian if and only if (S, Θ) are independent, S being exponentially distributed with mean 2 and Θ being uniform on $[0, 2\pi)$.

Exercise 3.13: We find

$$\begin{aligned}
\mathbf{E}|X|^p &= \int_{-\infty}^{+\infty} |x|^p \frac{1}{\sqrt{2\pi}}\ dx \\
&= 2\int_0^{\infty} x^p \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2) \\
&= \sqrt{2/\pi}\int_0^{\infty} (2u)^{p/2-1/2} \exp(-u)\ du \qquad = \sqrt{2^p/\pi}\Gamma(p/2+1/2)
\end{aligned}$$

Finally, $\mathbf{V}(X^2) = \mathbf{E}X^4 - (\mathbf{E}X^2)^2 = 3 - 1 = 2$.

The following `R` sniplet evaluates the result numerically and compares with Monte Carlo:

```
N <- 1e6
X <- rnorm(N)
sapply(1:4,function(p)mean(abs(X)^p))
sapply(1:4,function(p)sqrt(2^p/pi)*gamma((p+1)/2))
var(X^2)
```

Exercise 3.14: Write

$$\begin{pmatrix} X \\ Y \end{pmatrix} = \begin{bmatrix} a & 0 \\ b & c \end{bmatrix} \begin{pmatrix} U \\ V \end{pmatrix}$$

where $a = \sigma_X$, $b = \sigma_{XY}/\sigma_X$, $c = \sqrt{\sigma_Y^2 - \sigma_{XY}^2/\sigma_X^2}$, and U, V are jointly Gaussian with mean 0, variances 1, and independent. The condition

$$X > 0, Y > 0$$

can be stated in terms of U, V as

$$U > 0, bU + cV > 0,$$

which identifies a sector (or a cone) in the (u, v)-plane, the angle of which is ϕ given by

$$\cos\phi = -\frac{b}{\sqrt{b^2 + c^2}} = -\frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}}.$$

Since the distribution of U, V is invariant under rotations, the probability of this sector is

$$\mathbf{P}\{X > 0, Y > 0\} = \frac{\phi}{2\pi} = \frac{1}{2\pi} \arccos \frac{-\sigma_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}} = \frac{1}{2\pi}\left(\frac{\pi}{2} + \arcsin \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}}\right).$$

From this we easily find the probabilities

$$\mathbf{P}\{XY > 0\} = \frac{1}{2} + \frac{1}{\pi} \arcsin \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}},$$

and

$$\mathbf{P}\{XY < 0\} = \frac{1}{2} - \frac{1}{\pi} \arcsin \frac{\sigma_{XY}}{\sqrt{\sigma_X^2 \sigma_Y^2}}.$$

Exercise 3.16: The following table shows the various conditional expectations. To find e.g. $\mathbf{E}\{X|\mathcal{G}\}$, an easy approach is to consider the atomic sets in $\mathcal{G}$, i.e., $\{1,3\}, \{2\}, \{4,6\}, \{5\}$, and average X over each of these sets.

x	$\mathbf{E}\{X\|\mathcal{G}\}$	$\mathbf{E}\{X\|\mathcal{H}\}$	$\mathbf{E}\{\mathbf{E}\{X\|\mathcal{G}\}\|\mathcal{H}\}$	$\mathbf{E}\{\mathbf{E}\{X\|\mathcal{H}\}\|\mathcal{G}\}$
1	2	3	3	3
2	2	4	4	4
3	2	3	3	3
4	5	4	4	4
5	5	3	3	3
6	5	4	4	4

Exercise 3.17: The situation involves two observers, G and H, with information $\mathcal{G}$ and $\mathcal{H}$, so that G knows everything that H knows. $\mathbf{V}\{X|\mathcal{H}\}$ describes observer H's uncertainty about X. The result state that this can be decomposed into two terms: The first, $\mathbf{E}[\mathbf{V}\{X|\mathcal{G}\}|\mathcal{H}]$, is H's assessment of G's uncertainty. The next, $\mathbf{V}[\mathbf{E}\{X|\mathcal{G}\}|\mathcal{H}]$, represents that H is unsure about G's estimate.

1. We have, for the left hand side

$$\mathbf{V}\{X|\mathcal{H}\} = \mathbf{E}\{X^2|\mathcal{H}\} - (\mathbf{E}\{X|\mathcal{H}\})^2.$$

   For the right hand side, we get first

$$\begin{aligned}\mathbf{E}\{\mathbf{V}(X|\mathcal{G})|\mathcal{H}\} &= \mathbf{E}\left[\mathbf{E}\{X^2|\mathcal{G}\} - (\mathbf{E}\{X|\mathcal{G}\})^2|\mathcal{H}\right]\\ &= \mathbf{E}\{X^2|\mathcal{H}\} - \mathbf{E}\left[(\mathbf{E}\{X|\mathcal{G}\})^2|\mathcal{H}\right]\end{aligned}$$

   and next

$$\mathbf{V}\{\mathbf{E}(X|\mathcal{G})|\mathcal{H}\} = \mathbf{E}\left[(\mathbf{E}\{X|\mathcal{G}\})^2|\mathcal{H}\right] - (\mathbf{E}\{X|\mathcal{H}\})^2$$

   Combining these two terms on the right hand side, we get the desired result.

2. The result follows directly from the decomposition, since $\mathbf{V}[\mathbf{E}\{X|\mathcal{G}\}|\mathcal{H}] \geq 0$.

3. This situation may seem counter-intuitive, but can occur if the extra information available to G is that X is more uncertain than on average: Let Y be a Bernoulli variable with parameter $p \in (0,1)$ and let $X|Y$ be Gaussian distributed with variance Y. Let $\mathcal{G} = \sigma(Y)$ and let $\mathcal{H} = \{\emptyset, \Omega\}$, i.e., H has no information about the outcome of the stochastic experiment. Then $\mathbf{V}\{X|\mathcal{H}\} = \mathbf{V}\{X\} = \mathbf{EV}\{X|Y\} + \mathbf{VE}\{X|Y\} = p$ but $\mathbf{V}\{X|\mathcal{G}\} = Y$ which exceeds p when $Y = 1$.

Exercise 3.18: Since S and Θ are independent by construction, we have $\mathbf{E}\{S|\Theta\} = \mathbf{E}S = 2$. Since X and Y are independent, we have $\mathbf{E}\{S|Y\} = \mathbf{E}\{X^2 + Y^2|Y\} = \mathbf{E}\{Y^2|Y\} + \mathbf{E}\{X^2|Y\} = Y^2 + 1$.

We therefore get $\mathbf{E}\{S|\Theta \in \{0,\pi\}\} = 2$ while $\mathbf{E}\{S|Y = 0\} = 1$.

Ignoring the null event $S = 0$, we have that $\Theta \in \{0,\pi\}$ if and only if $Y = 0$. We see that the "conditional expectation of S given that the point (X,Y) is on the x-axis" is not well defined, because the event we condition on has probability 0. To get a well-defined conditional expectation, we must specify not just the actual observation that (X,Y) is on the x-axis, but also how the observation was made: By measuring Y or by measuring Θ?

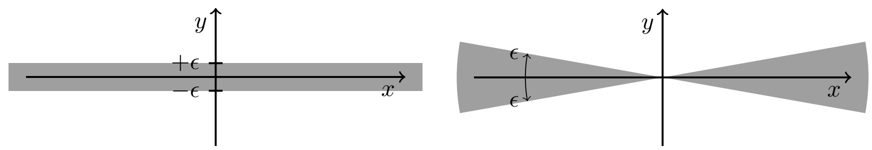
From a practical point of view, we never measure that the point (X,Y) is *on* the x-axis; we measure that the point is so close to the axis that we cannot distinguish it from the axis. The difference between the two σ-algebras $\sigma(Y)$ and $\sigma(\Theta)$, and thus the two different results, arise from different models of the measurement uncertainty: In the first, we have a measurement error on Y so that

$$\mathbf{E}(S \mid -\epsilon < Y < \epsilon) = 1 + O(\epsilon)$$

In the other, we have a measurement error on the angle Θ so that, for example

$$\mathbf{E}(S \mid \Theta < \epsilon) = 2 + O(\epsilon)$$

The two situations can be illustrated as follows:



In the left panel, we see the set $|Y| \leq \epsilon$. Averaging S over this set, and letting $\epsilon \to 0$, yields $\mathbf{E}(S|Y=0) = 1$. The right panel shows the set where Θ is ϵ-near 0, π, or 2π. Averaging S over this set, and letting $\epsilon \to 0$, yields $\mathbf{E}(S|\Theta = 0 \vee \Theta = \pi) = 2$. Note that the latter case set puts more weight to points far from the origin, i.e., with large S. Hence $\mathbf{E}(S|\Theta \in \{0,\pi\}) > \mathbf{E}(S|Y=0)$.

The paradox shows that it is not enough to report an observation, we should also report the observation process and, in particular, the nature of the observation error.

Exercise 3.19: First, we see from the rectangular geometry that $\mathbf{P}(A \cap B|Y) = \mathbf{P}(A|Y)\mathbf{P}(B|Y)$. In contrast, we have $\mathbf{P}(A \cap B|\neg Y) = 0$ while $\mathbf{P}(A|\neg Y) > 0$, $\mathbf{P}(B|\neg Y) > 0$.

Next, let X, Y and Z be three independent random variables, each uniformly distributed on $(0,1)$. Let $\mathcal{G} = \sigma(X)$ and define $A = \{Y \leq X\}$, $B = \{Z \leq X\}$. Then A and B are conditionally independent given $\mathcal{G}$. But $\Omega \in \mathcal{G}$ and A and B are not unconditionally independent.

Exercise 3.20: Without loss of generality, we can assume $\mu_X = 0$ and $\mu_Y = 0$. Set $Z := \Sigma_{xy}\Sigma_{yy}^{-1}Y$. Define $\tilde{X} := X - Z$; this is the estimation error when using Z as an estimate of X based on Y. Then $\tilde{X}$ is uncorrelated with Y:

$$\mathbf{E}\tilde{X}Y^\top = \mathbf{E}XY^\top - \mathbf{E}ZY^\top = \Sigma_{xy} - \mathbf{E}\Sigma_{xy}\Sigma_{yy}^{-1}YY^\top = \Sigma_{xy} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yy} = 0.$$

Using theorem 3.7.1, this shows that $Z = \mathbf{E}\{X|Y\}$. To show the result for the conditional variance, we evaluate the variance directly:

$$\begin{aligned}
\mathbf{V}\{X|Y\} &= \mathbf{V}\{\tilde{X}\} \\
&= \mathbf{V}(X - \Sigma_{xy}\Sigma_{yy}^{-1}Y) \\
&= \mathbf{E}(X - \Sigma_{xy}\Sigma_{yy}^{-1}Y)(X - \Sigma_{xy}\Sigma_{yy}^{-1}Y)^\top \\
&= \Sigma_{xx} - 2\Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} + \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yy}\Sigma_{yy}^{-1}\Sigma_{yx} \\
&= \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx}.
\end{aligned}$$

The first equality comes from $X = Z + \tilde{X}$ where Z is known given Y, and $\tilde{X}$ is independent of Y. Finally, to see that the conditional distribution is Gaussian it suffices to note that the logarithm to the conditional density of X is a quadratic form in x.

Exercise 3.21:

1. We must show that $Z = \mathbf{E}X$ satisfies the conditions in definition 3.5.1. Clearly Z is Y-measurable, since it is deterministic. Next, we must show that

   $$\mathbf{E}\{Z \cdot \mathbf{1}_H\} = \mathbf{E}\{X \cdot \mathbf{1}_H\}$$

   for every $H \in \sigma(Y)$. The right hand size equals $Z \cdot \mathbf{P}H$ since Z is deterministic.

   Assume that X is simple, i.e., $\mathbf{P}(X = x_i) = p_i$ where $x_1 < \cdots < x_n$ and $p_1 + \cdots + p_n = 1$. Then

   $$\mathbf{E}\{X \cdot \mathbf{1}_H\} = \sum_{i=1}^{n} x_i \mathbf{P}\{H \cap (X = x_i)\}$$

   By independence, $\mathbf{P}\{H \cap (X = x_i)\} = \mathbf{P}\{H\}p_i$ so the right hand side evaluates to $\mathbf{P}H \cdot \mathbf{E}X$.

   To show the result for general X, we approximate X with a sequence of simple random variables.

2. Let X be standard Gaussian and let $\mathcal{H} = \sigma(|X|)$.

Exercise 4.1: By the definition of Brownian motion, the increments

$$B_{t_1} - B_0, B_{t_2} - B_{t_1}, \ldots, B_{t_n} - B_{t_{n-1}}$$

are Gaussian and independent; hence they constitute a jointly Gaussian stochastic vector. The process itself $(B_{t_1}, B_{t_2}, \ldots, B_{t_n})$ (sampled at these time points) is found through a linear operation on this vector of increments (viz. the cumulative sum) and hence the process is Gaussian. The mean of each element is $\mathbf{E}B_t = \mathbf{E}(B_t - B_0) = 0$ and the variance is $\mathbf{V}B_t = \mathbf{V}(B_t - B_0) = t$. For the covariance, we find

$$E(B_sB_t) = \mathbf{E}[B_s((B_t - B_s) + B_s)] = \mathbf{E}B_s^2 = s.$$

Exercise 4.2: The following R code implements one way to simulate Brownian motion, and does the verification.

```
test.rBM <- function(N=1e4, t=c(0,0.5,1.5,2))
{
  B <- sapply(1:N,function(i)rBM(t))
  print("Theoretical covariance:")
  print(sapply(t,function(s)pmin(s,t)))
  print("Empirical covariance:")
  print(cov(t(B)))

  %% QQ-plot
  qqnorm(B[length(t)/sqrt(tail(t,1)),])
}
```

Exercise 4.3: We have

$$\mathbf{E}|B_{t+h} - B_t|^2 = h.$$

so $B_{t+h} - B_t \to 0$ in mean square as $h \to 0$. However,

$$\mathbf{E}\frac{1}{h}(B_{t+h} - B_t) = 0, \quad \mathbf{V}\frac{1}{h}(B_{t+h} - B_t) = 1/h$$

which diverges to ∞ as $h \searrow 0$. Since the variance diverges, there can be no mean square limit.

Exercise 4.5: It is clear that $\{W_t\}$ has independent increments. To find the distribution of one such increment, let $0 < s < t$. Then

$$W_t - W_s = tB_{1/t} - sB_{1/s}$$

is obviously Gaussian with mean 0. For the variance we find

$$\mathbf{V}(W_t - W_s) = (t, \ -s)\begin{bmatrix} 1/t & 1/t \\ 1/t & 1/s \end{bmatrix}\begin{pmatrix} t \\ -s \end{pmatrix} = t - s - s + s = t - s$$

as required.

It is clear that $\{W_t\}$ is continuous at any point $t > 0$. To see continuity at $t = 0$, note that $\lim_{t\searrow 0} W_t = \lim_{s\to\infty} s^{-1}B_s$, which is 0 according to exercise 4.4.

Exercise 4.6: By time inversion we have

$$\limsup_{t\searrow 0} \frac{1}{t}B_t = \limsup_{s\to\infty} s \cdot \frac{1}{s}W_s$$

which clearly is ∞. The result for the limes inferior follows, and we conclude that the sample path of the difference quotient does not converge as the time step vanishes.

Exercise 4.7: Conditions 1 and 2 are trivial. Condition 3 follows directly from the independence of the increments: Let $0 \geq s < t$, then $\mathbf{E}\{B_t|\mathcal{F}_s\} = B_s + \mathbf{E}\{B_t - B_s|\mathcal{F}_s\} = B_s$.

Exercise 4.8: As in the previous exercise, conditions 1 and 2 are trivial. Condition 3 follows from a direct calculation:

$$\begin{aligned}\mathbf{E}\{B_t^2 - t|\mathcal{F}_s\} &= \mathbf{E}\{(B_s + (B_t - B_s))^2|\mathcal{F}_s\} - t \\ &= B_s^2 + 2B_s\mathbf{E}\{B_t - B_s|\mathcal{F}_s\} + \mathbf{E}\{(B_t - B_s)^2|\mathcal{F}_s\} - t \\ &= B_s^2 + (t-s) - t \\ &= B_s^2 - s\end{aligned}$$

as required.

Exercise 4.9: This follows directly from the martingale inequality: Let $\{M_t : t \in \mathbf{N}_0\}$ denote the gambler's fortune after game t. Then

$$\mathbf{P}\{\sup(M_0, M_1, \ldots, M_N) \geq c\} \leq \frac{c}{M_0}$$

and with $M_0 = 1$, $c = 100$ the result follows.

Exercise 4.10: First note that since all variances are finite, the covariance exists as a finite number. Now we use the hint:

$$\begin{aligned}\mathbf{E}(M_v - M_u)(M_t - M_s) &= \mathbf{E}\mathbf{E}\{(M_v - M_u)(M_t - M_s)|\mathcal{F}_u\} \\ &= \mathbf{E}\left[(M_t - M_s)\mathbf{E}\{M_v - M_u|\mathcal{F}_u\}\right] \\ &= 0.\end{aligned}$$

The second claim follows directly from the increments $M_u - M_t$ and $M_t - M_s$ being uncorrelated. Finally, let $0 \leq s \leq t$, then

$$\mathbf{V}M_t = \mathbf{V}M_s + \mathbf{V}(M_t - M_s) \geq \mathbf{V}M_s.$$

Exercise 4.12: First, it is clear that the return value $\tilde{X}$ is conditionally Gaussian given Y, as it should be.

Next, the conditional expectation of $\tilde{X}$ is

$$\mathbf{E}\{\tilde{X}|Y\} = \mathbf{E}\{X|Y\}.$$

To see this, note that $\mathbf{E}\{\mathbf{E}\{\bar{X}|\bar{Y}\} + \bar{X}|Y\} = 0$, since $\bar{X}$ and $\bar{Y}$ are independent of Y.

Finally, we compute the conditional variance of $\tilde{X}$ given Y. Noting that $\mathbf{E}\{X|Y\}$ is Y-measurable and that $(\bar{X}, \bar{Y})$ is independent of Y, we get

$$\mathbf{V}\{\tilde{X}|Y\} = \mathbf{V}(\bar{X} - \mathbf{E}\{\bar{X}|\bar{Y}\}) = \mathbf{V}\{\bar{X}|\bar{Y}\}.$$

Now, since $(\bar{X}, \bar{Y})$ has the same distribution as (X, Y), the result follows.

Exercise 4.13: See source code `Wiener-expansion.R`.

Exercise 4.14: Loosely, Brownian motion scales with the square root of time, so as time becomes large, the term ut dominates over B_t. More strictly, we use

$$\liminf_{t\to\infty} \frac{B_t}{2\sqrt{t\log\log t}} = -1$$

and

$$\liminf_{t\to\infty} \frac{ut}{2\sqrt{t\log\log t}} = \infty$$

together with the general fact

$$\liminf_{t\to\infty}[X_t + Y_t] \geq \liminf_{t\to\infty} X_t + \liminf_{t\to\infty} Y_t$$

to conclude

$$\liminf_{t\to\infty} \frac{B_t + ut}{2\sqrt{t\log\log t}} = \infty$$

from which result follows. (An alternative is to use the result from exercise 4.4).

Exercise 4.15: From Exercise 4.14, we know that $B_t - t/2 \to -\infty$. Hence $X_t \to 0$. In contrast, the factbox on page 86 regarding log-normal distributions gives that $\mathbf{E}X_t^2 = \exp(t) \to \infty$ as $t \to \infty$.

Exercise 4.17:

1. With the candidate solution, we compute $\sum_k S_{ik}P_{kj}$, for the different cases of (i,j). With $i = j = 1$, we get

   $$\frac{t_2}{t_2 - t_1} - \frac{t_1}{t_2 - t_1} = 1.$$

   With $1 \leq i < j \leq n$, we get

   $$t_i\left(-\frac{1}{t_j - t_{j-1}} + \frac{t_{j+1} - t_{j-1}}{(t_{j+1} - t_j)(t_j - t_{j-1})} - \frac{1}{t_{j+1} - t_j}\right) = 0.$$

   With $1 < i = j < n$, we get

   $$-\frac{t_{i-1}}{t_i - t_{i-1}} + \frac{t_i(t_{i+1} - t_{i-1})}{(t_{i+1} - t_i)(t_i - t_{i-1})} - \frac{t_i}{t_{i+1} - t_i} = 1$$

   (after some cleaning up). Finally, with $i = j = n$, we get

   $$-t_{n-1}\frac{1}{t_n - t_{n-1}} + t_n\frac{1}{t_n - t_{n-1}} = 1.$$

2. The joint density of X is

$$f_X(x) = \frac{1}{(2\pi|S|)^{1/2}} \exp\left(-\frac{1}{2}x^\top P x^\top\right).$$

Now, the conditional density of X_i at x_i, given $X_{-i} = x_{-i}$, is

$$\begin{aligned} f_{X_i|X_{-i}}(x_i, x_{-i}) &= \frac{f_{X_i,X_{-i}}(x_i, x_{-i})}{f_{X_{-i}}(x_{-i})} \\ &= \frac{1}{Z}\exp\left(-\frac{1}{2}P_{ii}x_i^2 - x_i P_{i,-i}x_{-i} - \frac{1}{2}x_{-i}^\top P_{-i,-i}x_{-i}\right). \end{aligned}$$

Here, the normalization constant Z depends on x_{-i} but not on x_i. As a function of x_i, we recognize this as the p.d.f. of a Gaussian random variable with variance $1/S_{ii}$. The conditional mean can e.g. be found by maximizing the conditional p.d.f. w.r.t. x_i, using that the mean and the mode coincide in the Gaussian distribution. We get

$$-P_{ii}x_i - P_{i,-i}x_{-i} = 0,$$

i.e., the conditional mean is $-P_{ii}^{-1}P_{i,-i}x_{-i}$, as claimed.

3. Since the precision matrix is tridiagonal, we have that $X_i = B_{t_i}$ is conditionally independent of $X_j = B_{t_j}$ for $|j - i| > 1$, which is what the graphical model illustrates. The variances in the Brownian bridge, $\mathbf{V}\{B_{t_i}|B_{t_{i-1}}, B_{t_{i+1}}\}$, are found from the diagonal elements in the precision matrix. From the precision matrix, we also see that the conditional expectation interpolates the two neighbors linearly:

$$\begin{aligned} \mathbf{E}\{B_{t_i}|B_{t_{i-1}}, B_{t_{i+1}}\} &= P_{i,i-1}B_{t_{i-1}}/P_{i,i} + P_{i,i+1}B_{t_{i+1}}/P_{i,i} \\ &= \frac{t_{i+1} - t_i}{t_{i+1} - t_{i-1}}B_{t_{i-1}} + \frac{t_i - t_{i-1}}{t_{i+1} - t_{i-1}}B_{t_{i+1}}. \end{aligned}$$

Exercise 4.18: The filtration in question must be that generated by the process itself, so the process is obviously adapted. We have

$$\mathbf{E}|M_n| \le \sum_{i=1}^{n} \mathbf{E}|X_i| < \infty$$

As for the martingale property itself, we have for $1 \le s < t$

$$\mathbf{E}\{M_t|\mathcal{F}_s\} = M_s + \mathbf{E}\{\sum_{i=s+1}^{t} X_i|\mathcal{F}_s\} = M_s$$

as required.

Exercise 4.19: The process $\{M_t\}$ is adapted to the filtration, since $\mathbf{E}X|\mathcal{F}_t\}$ is $\mathcal{F}_t$-measurable by the definition of conditional expectations. To show $\mathbf{E}|M_t| < \infty$, note that we have $|M_t| \leq \mathbf{E}\{|X|\ |\mathcal{F}_t\}$ by Jensen's inequality (theorem 3.3.1). Hence

$$\mathbf{E}|M_t| \leq \mathbf{E}\mathbf{E}\{|X|\ \mathcal{F}_t\} = \mathbf{E}|X| < \infty$$

as required. Finally, we use the Tower property (theorem 3.5.1) to see that

$$\mathbf{E}\{M_t|\mathcal{F}_s\} = \mathbf{E}\{\mathbf{E}(X|\mathcal{F}_t)|\mathcal{F}_s\} = \mathbf{E}\{X|\mathcal{F}_s\} = M_s$$

since $\mathcal{F}_s \subset \mathcal{F}_t$.

Exercise 4.20: V_n is the sum of n independent and identically distributed random variables, which each has expectation $\sqrt{2/\pi}/\sqrt{n}$. Here we have used exercise 3.13 with $p = 1$ and the fact that $\mathbf{E}|\Delta B|$ scales with $\sqrt{|\Delta t|}$. The result $\mathbf{E}V_n = \sqrt{2n/\pi}$ follows.

Exercise 4.21: Using the same reflection argument as in Theorem 4.3.2, we find that the probability that the path crosses the origin in the interval $(t, 1]$ is twice the probability that B_t and B_1 have opposite sign:

$$\mathbf{P}\{\tau > t\} = 2\mathbf{P}\{B_t B_1 < 0\}.$$

From Exercise 3.14, and using that B_t and B_1 have variances t and 1, respectively, and covariance t (Section 4.3), we find

$$\mathbf{P}\{\tau > t\} = 2\left(\frac{1}{2} - \frac{1}{\pi}\arcsin\frac{t}{\sqrt{t}}\right),$$

or

$$\mathbf{P}\{\tau \leq t\} = \frac{2}{\pi}\arcsin\sqrt{t}.$$

Exercise 5.1: This is a standard, if somewhat tedious, exercise in manipulation of the integrals that appear in the Fourier transform:

$$\begin{aligned}
S_X(\omega) &= \frac{1}{2\pi}\int_{-\infty}^{+\infty} \rho_X(l)\ \exp(-i\omega l)\ dl \\
&= \frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_0^\infty\int_0^\infty e^{Av} G\rho_U(l+v-w)G^\top e^{A^\top w}\ dw\ dv\ \exp(-i\omega l)\ dl \\
&= \frac{1}{2\pi}\int_0^\infty\int_0^\infty\int_{-\infty}^{+\infty} e^{Av} G\rho_U(l+v-w)G^\top e^{A^\top w}\ \exp(-i\omega l)\ dl\ dw\ dv.
\end{aligned}$$

In the inner integral, we make the substitution $s = l + v - w$, $l = s + w - v$,

$dl = ds$, to get

$$\begin{aligned}
S_X(\omega) &= \frac{1}{2\pi}\int_0^\infty \int_0^\infty \int_{-\infty}^{+\infty} e^{Av} G\rho_U(s) G^\top e^{A^\top w} \ \exp(-i\omega(s+w-v)) \ ds \ dw \ dv \\
&= \frac{1}{2\pi}\int_0^\infty \int_0^\infty \int_{-\infty}^{+\infty} e^{Av} e^{i\omega v} G\rho_U(s) e^{-i\omega s} G^\top e^{A^\top w} \ e^{i\omega w} \ ds \ dw \ dv \\
&= \int_0^\infty e^{Av} e^{i\omega v} \ dv \frac{1}{2\pi}\int_{-\infty}^{+\infty} G\rho_U(s) e^{-i\omega s} \ ds \int_0^\infty G^\top e^{A^\top w} \ e^{i\omega w} \ dw \\
&= H(-\omega) S_U(\omega) H^\top(\omega).
\end{aligned}$$

Exercise 5.2: From the definition of Brownian motion, we have that $B_{t+k} - B_t$ is distributed as $N(0,k)$ for all t, so $X_t \sim N(0, 1/k)$. To determine the covariance structure, let $0 \le s \le t$. Aiming to find $\mathbf{E}X_s X_t$, we must distinguish between two cases:

1. $t > s + k$. Then the intervals $[s, s+k]$ and $[t, t+k]$ are disjoint, so the increments $B_{s+k} - B_s$ and $B_{t+k} - B_t$ are independent. We get $\mathbf{E}X_s X_t = 0$.

2. $t \le s+k$. Then the intervals $[s, s+k]$ and $[t, t+k]$ are not disjoint. We single out the overlapping sub-interval $[t, s+k]$ by adding and subtracting the contributions from the end points:

$$\begin{aligned}
& \mathbf{E}(B_{s+k} - B_s)(B_{t+k} - B_t) \\
=\ & \mathbf{E}(B_{s+k} - B_t + B_t - B_s)(B_{t+k} - B_{s+k} - B_{s+k} - B_t) \\
=\ & \mathbf{E}(B_{s+k} - B_t)^2 \\
=\ & s + k - t.
\end{aligned}$$

We therefore get $\mathbf{E}X_s X_t = k^{-2}(s+k-t)^2$. We see that the covariance depends only on the time lag $t - s$, so $\{X_t\}$ is second order stationary. Using symmetry, we arrive at the expression (5.15).

Exercise 5.3:

1. First, note that the expression has $\exp(At) = I$ for $t = 0$. Next, differentiate w.r.t. time to get

$$\begin{aligned}
\frac{d}{dt}\exp(At) &= -\mu \exp(At) + e^{-\mu t} k \begin{bmatrix} -\sin kt & -\cos kt \\ \cos kt & -\sin kt \end{bmatrix} \\
&= \left(-\mu I + k \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}\right) \exp(At) \\
&= A \exp(At)
\end{aligned}$$

which defines the matrix exponential.

2. We insert the candidate solution in the algebraic Lyapunov equation (5.27) and verify that it holds.

3. We combine the two previous results and use (5.28) for a positive time lag; for negative time lags, we use $\rho_X(-t) = \rho_X^\top(t)$.

Exercise 5.4:

1. Since dB_t/dt has a constant variance spectrum of 1, we find $\sigma = \sqrt{2k_BTR}$. With the given numeric values, we get $2.9 \cdot 10^{-9}$ V$\sqrt{s}$.

2. The equation for Q_t is

$$\frac{dQ_t}{dt} = -\frac{1}{RC}Q_t + \frac{\sigma}{R}\frac{dB_t}{dt}$$

or

$$dQ_t = -\frac{1}{RC}Q_t\ dt + \frac{\sigma}{R}\ dB_t.$$

3. The stationary mean for Q_t is 0. The variance is

$$\mathbf{V}Q_t = \frac{\sigma^2}{2R^2/(RC)} = \frac{\sigma^2 C}{2R} = k_BTC.$$

Note that this is independent of R. The autocovariance function of $\{Q_t\}$ is

$$\rho_Q(h) = \mathbf{E}Q_tQ_{t+h} = k_BTCe^{-|h|/(RC)}$$

and the power spectrum of $\{Q_t\}$ is

$$S_Q(\omega) = \frac{\sigma^2}{R^2(\omega^2 + 1/(RC)^2} = \frac{2k_BTRC^2}{1 + R^2C^2\omega^2}$$

For the voltage over the capacitor, we find a zero mean, a variance of k_BT/C (which again is independent of R), and a power spectrum

$$S_{Q/C}(\omega) = \frac{\sigma^2}{C^2R^2(\omega^2 + 1/(RC)^2} = \frac{2k_BTR}{1 + R^2C^2\omega^2}$$

Note that the low-frequency asymptote of this variance spectrum is independent of C and equals $2k_BTR$, i.e., the same as the driving thermal noise.

4. With $C = 1$ nF, we get a r.m.s. charge of $\sqrt{k_BTC} = 2 \cdot 10^{-15}$ C, a r.m.s. voltage of $\sqrt{k_BT/C} = 2\ \mu$V, and a decorrelation time of $RC = 1\ \mu$s.

Exercise 5.6: The algebraic Lyapunov equation is

$$\begin{bmatrix} 0 & 1 \\ -k & -\mu \end{bmatrix} \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} + \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \begin{bmatrix} 0 & -k \\ 1 & -\mu \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & s^2 \end{bmatrix} = 0$$

or, elementwise, starting with the (1,1)-element

$$\Sigma_{21} + \Sigma_{12} = 0,$$

which, together with symmetry $\Sigma_{12} = \Sigma_{21}$, implies that $\Sigma_{12} = \Sigma_{21} = 0$. So in stationarity, the position and velocity is uncorrelated. Using this, we get for the (1,2)-element

$$\Sigma_{22} - k\Sigma_{11} = 0.$$

Since the average kinetic energy is $\frac{1}{2}\mathbf{E}V_t^2 = \frac{1}{2}\Sigma_{22}$, while the average potential energy is $\frac{1}{2}k\mathbf{E}X_t^2 = \frac{1}{2}k\Sigma_{11}$, equipartitioning follows.

Finally, we can find the average energies. For the (2,2)-element, we get

$$-2\mu\Sigma_{22} + s^2$$

or $\Sigma_{22} = \frac{1}{2}\mu^{-1}s^2$.

Exercise 5.7: We use the variance decomposition formula

$$\mathbf{V}X_{t+h} = \mathbf{V}\mathbf{E}\{X_{t+h}|X_t\} + \mathbf{E}\mathbf{V}\{X_{t+h}|X_t\}$$

where we take $\mathbf{V}$ to mean the variance-covariance matrix. With the Euler-Maruyama approximation, we get

$$\mathbf{E}\{X_{t+h}|X_t\} = (I + Ah)X_t \Rightarrow \mathbf{V}\mathbf{E}\{X_{t+h}|X_t\} = (I + Ah)\mathbf{V}X_t(I + Ah)^\top$$

and

$$\mathbf{V}\{X_{t+h}|X_t\} = GG^\top \cdot h$$

Inserting, cleaning up, and omitting second order terms in h, we get

$$\mathbf{V}X_{t+h} = \mathbf{V}X_h + (A\mathbf{V}X_t + \mathbf{V}X_t \cdot A^\top + GG^\top)h$$

Letting $h \searrow 0$, we get

$$\frac{d}{dt}\mathbf{V}X_t = A\mathbf{V}X_t + \mathbf{V}X_t \cdot A^\top + GG^\top.$$

Exercise 5.9: Since the drift is linear in the state, the mean $\mu_t = \mathbf{E}X_t$ satisfies the ordinary differential equation

$$\frac{d}{dt}\mu_t = \frac{b - \mu_t}{T - t}\ .$$

Inserting $b_t = bt/T$, we see that this satisfies the ordinary differental equation. Likewise, the variance $\Sigma_t = \mathbf{V}X_t$ is governed by the equation

$$\frac{d}{dt}\Sigma_t = -\frac{2}{T-t}\Sigma_t + 1$$

and we see that $\Sigma_t = t(1-t/T)$, which we found for the Brownian bridge, satisfies this equation. Finally, since $\{X_t\}$ is governed by a linear equation, it is a Gaussian process.

Exercise 5.10: We write the system in standard form

$$dZ_t = AZ_t\ dt + G\ dB_t$$

where

$$A = \begin{bmatrix} 0 & I \\ -K & -cI \end{bmatrix}, \quad G = \begin{bmatrix} 0 \\ \sigma I \end{bmatrix}.$$

The algebraic Lyapunov equation for the stationary variance S is

$$AS + SA^\top + GG^\top = 0$$

and with the Ansatz

$$S = \begin{bmatrix} P & 0 \\ 0 & Q \end{bmatrix}$$

we get the two equations

$$-2cQ + \sigma^2 I = 0, \quad Q = KP,$$

which have the solutions

$$Q = \frac{\sigma^2}{2c} I, \quad P = \frac{\sigma^2}{2c} K^{-1}.$$

Note, in particular, that the precision matrix of the position is exactly the stiffness matrix K, scaled. A simulation is seen in source code `pearls.R`.

Exercise 6.3: Since $\{B_t^{(N)}\}$ has smooth sample paths, we have

$$\int_0^1 B_s^{(N)}\ dB_s^{(N)} = \frac{1}{2}(B_1^{(N)})^2$$

so

$$\int_0^t B_s^{(N)}\ dB_s^{(N)} \to \frac{1}{2}B_t^2 = \int_0^t B_s \circ dB_s.$$

Exercise 6.4: As usual, let $\|\cdot\|$ denote the $\mathcal{L}_2$ norm, i.e., root mean square. For a given partition $\Delta = \{0 = t_0 < t_1 < \cdots < t_n = t\}$, we get

$$\begin{aligned}\sum_{i=1}^{n}(\Delta T_i)^2 &\le |\Delta| \sum_{i=1}^{n} \Delta T_i \\ &= |\Delta| t\end{aligned}$$

and

$$\begin{aligned}\left\| \sum_{i=1}^{n} \Delta T_i \ \Delta B_i \right\| &\le \sum_{i=1}^{n} \|\Delta T_i\| \|\Delta B_i\| \\ &\le \sum_{i=1}^{n} \Delta t_i \sqrt{\Delta t_i} \\ &\le \sqrt{|\Delta|} \sum_{i=1}^{n} \Delta t_i \\ &= t\sqrt{|\Delta|}\end{aligned}$$

which both converge to 0 as $|\Delta| \to 0$.

Alternatively, one could have shown that the quadratic variation of *biased* random walk equals time, $[B+T]_t = t$. From this follows $\langle T, B\rangle_t = ([B+T]_t - [T-B]_t)/4 = 0$.

Exercise 6.5: For a given partition $\Delta = \{0 = t_0 < t_1 < \cdots < t_n = t\}$, we get

$$\mathbf{E} \sum_{i=1}^{n} \Delta B_i \ \Delta W_i = 0$$

and

$$\mathbf{E} \left(\sum_{i=1}^{n} \Delta B_i \ \Delta W_i \right)^2 = \sum_{i=1} n \Delta t_i^2 \le t|\Delta|$$

which converges to 0 as $|\Delta| \to 0$.

Exercise 6.6: According to the triangle inequality for norms, we have $\| \sum (\Delta B_i)^3 \| \le \sum \|(\Delta B_i)^3\|$, where $\|\cdot\|$ denotes root mean square. By the scaling of Brownian motion, we have $\|(\Delta B_i)^3\| = C(\Delta t_i)^{3/2}$. The numerical value for the constant C can be found from exercise 3.13 but we do not need it. It follows that

$$\left\| \sum_{i=1}^{n} (\Delta B_i)^3 \right\| \le C \sum_{i=1}^{n} (\Delta t_i)^{3/2} \le C|\Delta|^{1/2} \sum_{i=1}^{n} \Delta t_i = C|\Delta|^{1/2} t \to 0$$

as $|\Delta| \to 0$.

Exercise 6.7: We first verify the result formally using the rules (6.16):

$$\begin{aligned} d\langle X,Y\rangle_t &= dX_t\ dY_t \\ &= F_t K_t (dt)^2 + (F_t L_t + G_t K_t) dt\ dB_t + G_t L_t (dB_t)^2 \\ &= G_t L_t\ dt. \end{aligned}$$

Since this computation is formal only at this point, we verify its validity. Setting $\Delta X_i = X_{t_i} - X_{t_{i-1}}$, $\Delta Y_i = Y_{t_i} - Y_{t_{i-1}}$, $\Delta t_i = t_i - t_{i-1}$, we have

$$\begin{aligned} \langle X,Y\rangle_t &= \lim_{|\Delta|\to 0} \sum_{i=1}^{\#\Delta} \Delta X_i\ \Delta Y_i \\ &= \lim_{|\Delta|\to 0} \sum_{i=1}^{\#\Delta} \left[F_{t_i} K_{t_i} \Delta t_i^2 + F_{t_i} L_{t_i} \Delta t_i \Delta B_i + G_{t_i} K_{t_i} \Delta t_i \Delta B_i + G_{t_i} L_{t_i} (\Delta B_i)^2\right] \\ &= \lim_{|\Delta|\to 0} \sum_{i=1}^{\#\Delta} G_{t_i} L_{t_i} (\Delta B_i)^2 \\ &= \int_0^t G_s L_s\ ds. \end{aligned}$$

Here, we have used $[B]_t = t$, so $d[B]_t = dt$, and $\langle T, B\rangle_t = 0$.

Exercise 6.8: The integral corresponds to $\mathbf{E}\tau^2$, where $F_t = \mathbf{P}\{\tau \le t\}$. I.e., τ is exponentially distributed with mean 1. From the properties of that distribution, we know that $\mathbf{E}\tau^2 = \mathbf{V}\tau + (\mathbf{E}\tau)^2 = 1 + 1 = 2$.

A direct evaluation of the integral is

$$\begin{aligned} \int_0^\infty t^2\ dF_t &= \int_0^\infty t^2 \exp(-t)\ dt \\ &= \Gamma(3) = 2. \end{aligned}$$

Exercise 6.9: For the mean, we find

$$\mathbf{E} \int_0^t B_s\ dB_s = \frac{1}{2}t - \frac{1}{2}t = 0,$$

which agrees with theorem 6.3.2. For the variance, we find:

$$\mathbf{V} I_t = \frac{1}{4}\mathbf{V} B_t^2 = \frac{t^2}{4}\mathbf{V} B_1^2 = \frac{1}{2}t^2$$

using that $\mathbf{V} B_1^2 = 2$ (Exercise 3.13). This agrees with the Itô isometry:

$$\mathbf{V} \int_0^t B_s\ dB_s = \int_0^t \mathbf{E}|B_s|^2\ ds = \int_0^t s\ ds = \frac{1}{2}s^2.$$

Exercise 6.11:

1. See the source code `Exercise-low-pass-solution.R` for simulation.
2. We have $d(Y_t - B_t) = X_t\ dt - B_t = -\lambda^{-1}\ dX_t$, and since $Y_0 = B_0 = X_0 = 0$, it follows that $Y_t - B_t = -X_t/\lambda$. $\{X_t\}$ is an Ornstein-Uhlenbeck process with stationary variance $\lambda^2/(2\lambda) = \lambda/2$, from which the result follows.
3. We know that $\int_0^t B_s\ dB_s = B_t^2/t - t/2$ and that $\int_0^t B_s \circ dB_s = B_t^2/t$. Since $\{Y_t\}$ is a low-pass filtering of the Brownian motion, it has bounded total variation, so the Itô and Stratonovich interpretation coincide, and the usual rules of calculus apply: $\int_0^t Y_s\ dY_s = \int_0^t Y_s \circ dY_s = Y_t^2/2$. Since $Y_t \approx B_t$, these integrals are very close to $\in t_0^t B_s \circ dB_s = B_t^2/2$.
4. Since the Itô integral depends continuously on the integrand, and $\{Y_t\}$ is near $\{B_t\}$, we have that $\int_0^t Y_s\ dB_s$ is near $\int_0^t B_s\ dB_s$. Since the cross-variation between $\{Y_t\}$ and $\{B_t\}$ is 0, the Itô and Stratonovich interpretation of this integral coincides.

   On the other hand, we have $d(Y_t B_t) = Y_t\ dB_t + B_t\ dY_t$ using that $dY_t\ dB_t = 0$. We therefore get the approximation

   $$\int_0^t B_s\ dY_s = Y_t B_t - \int_0^t Y_t\ dB_t \approx B_t^2 - B_t^2/2 + t/2 = B_t^2/2 + t/2.$$

   Finally, the Itô and Stratonovich interpretation of this integral coincides, again since $\{B_t\}$ and $\{Y_t\}$ have vanishing cross-variation.

Exercise 6.12: The Euler-Maruyama scheme is

$$X_{t+h} = X_t + \lambda(\xi - X_t)h + \gamma\sqrt{X_t}(B_{t+h} - B_t)$$

so

$$\mathbf{E}\{X_{t+h}|X_t\} = X_t + \lambda(\xi - X_t)h$$

and therefore, with $\mu_t = \mathbf{E}X_t$,

$$\mu_{t+h} = \mu_t + \lambda(\xi - \mu_t)h.$$

In the limit $h \to 0$, we get

$$\dot{\mu}_t = \lambda(\xi - \mu_t)$$

with the solution $\mu_t = \xi + (\mu_0 - \xi)\exp(-\lambda t)$. For the variance $\Sigma_t = \mathbf{V}X_t$, we get

$$\mathbf{V}\{X_{t+h}|X_t\} = \gamma^2 X_t h$$

and therefore, with the law of total variance,

$$\begin{aligned}\Sigma_{t+h} &= \mathbf{E}\mathbf{V}\{X_{t+h}|X_t\} + \mathbf{V}\mathbf{E}\{X_{t+h}|X_t\}\\ &= \gamma^2\mu_t h + (1-\lambda h)^2\Sigma_t.\end{aligned}$$

In the limit $h \to 0$, we find

$$\dot{\Sigma}_t = -2\lambda\Sigma_t + \gamma^2\mu_t.$$

In particular, the stationary variance is $\Sigma = \gamma^2\xi/(2\lambda)$.

For the autocovariance function, we use

$$\mathbf{E}\{X_t|X_0\} = \xi + (X_0 - \xi)e^{-\lambda t}$$

and therefore

$$\mathbf{E}(X_t - \xi)X_0 = \mathbf{E}\mathbf{E}\{(X_t - \xi)X_0|X_0\} = \mathbf{E}(X_0 - \xi)X_0 e^{-\lambda t} = \Sigma e^{-\lambda t}.$$

Exercise 7.1: We apply Itô's lemma to $Y_t = h(t, X_t)$ where $h(t,x) = y\exp(x)$ where $X_t = (r - \sigma^2/2)t + \sigma B_t$, i.e., $dX_t = (r - \sigma^2/2)\ dt + \sigma\ dB_t$. We find

$$\dot{h} = 0, \quad h' = h, \quad h'' = h,$$

so that

$$\begin{aligned} dY_t &= \dot{h}\ dt + h'\ dX_t + \frac{1}{2}h''\ dX_t^2 \\ &= 0 + Y_t(r - \frac{1}{2}\sigma^2)\ dt + Y_t\sigma\ dB_t + \frac{1}{2}Y_t\sigma^2\ dt \\ &= rY_t\ dt + \sigma Y_t\ dB_t \end{aligned}$$

as required. Next, $\log Y_t$ is Gaussian with mean $\log y + (r - \frac{1}{2}\sigma^2)t$ and variance $\sigma^2 t$, Y_t is log-Gaussian distributed

$$Y_t \sim \mathrm{LN}(\log y + (r - \frac{1}{2}\sigma^2)t, \sigma^2 t).$$

It follows from the properties of the log-Gaussian distributions that Y_t has mean

$$\mathbf{E}Y_t = y\exp((r - \frac{1}{2}\sigma^2)t + \frac{1}{2}\sigma^2 t) = y\exp(rt)$$

and variance

$$\mathbf{V}Y_t = y^2(\exp(\sigma^2 t) - 1)\exp(2(r - \frac{1}{2}\sigma^2)t + \sigma^2 t) = y^2(\exp(\sigma^2 t) - 1)\exp(2rt).$$

The mean square is $\mathbf{E}Y_t^2 = (\mathbf{E}Y_t)^2 + \mathbf{V}Y_t = y^2\exp((2r + \sigma^2)t)$.

Exercise 7.4: Using the hint, define $Y_t = h(t, X_t) = \exp(-At)X_t$, then

$$Y_t = x + \int_0^t e^{-As}(w_s\ ds + G\ dB_s)$$

so $dY_t = e^{-At}(w_t\ dt + G\ dB_t)$. By Itôś lemma, this implies

$$dX_t = e^{At}dY_t + Ae^{At}Y_t\ dt = AX_t\ dt + w_t\ dt + G\ dB_t$$

as required. We find the mean by removing the Itô integral:

$$\mathbf{E}X_t = e^{At}x + \int_0^t e^{A(t-s)} w_s \, ds,$$

and note that the mean satisfies the ordinary differential equation

$$\frac{d}{dt}\mathbf{E}X_t = A\mathbf{E}X_t + w_t.$$

For the variance-covariance matrix $\Sigma(t)$, we get the same result as (5.20), i.e.

$$\Sigma(t) = \int_0^t e^{A(t-v)} GG^\top e^{A^\top (t-v)} \, dv$$

since linearity implies that the deterministic input $\{w_t\}$ does not affect the variance.

Exercise 7.5: Define $Y_t = h(t, x)$ where $h(t, x) = e^{-F_t}x$, then

$$Y_t = x + \int_0^t e^{-F_s} \sigma_s \, dB_s$$

or

$$dY_t = e^{-F_t} \sigma_t \, dB_t$$

By Itôś lemma applied to $X_t = g(F_t, Y_t)$ with $g(f, y) = \exp(f)y$, using $\partial^2 g / \partial y^2 = 0$ and $(dF_t)^2 = dF_t \cdot dY_t 0$, we get

$$dX_t = X_t \, dF_t + e^{F(t)} \, dY_t = \lambda_t X_t \, dt + \sigma_t \, dB_t$$

So $X_t = g(F_t, Y_t) = e^{F_t}x + \int_0^t e^{F_t - F_s} \sigma_s \, dB_s$ satisfies the stochastic differential equation.

Exercise 7.6: We let $h(y) = \sqrt{y}$; thus $h'(y) = 1/(2\sqrt{y})$ and $h''(y) = -y^{-3/2}/4$. Thus

$$dZ_t = h'(Y_t) \, dY_t + \frac{1}{2}'' (Y_y)(dY_t)^2 = \frac{n-1}{2} \frac{1}{Z_t} \sigma^2 \, dt + \sigma dW_t.$$

Exercise 7.7: For the process (X_t, Y_t), the drift is linear so the expectation satisfies

$$\begin{pmatrix} d\mathbf{E}X_t \\ d\mathbf{E}Y_t \end{pmatrix} = \frac{1}{2} \begin{pmatrix} \mathbf{E}X_t \\ \mathbf{E}Y_t \end{pmatrix} \, dt$$

from which we find $\mathbf{E}X_t = \exp(-\frac{1}{2}t)\mathbf{E}X_0$. With $X_0 = 1$, the conclusion follows.

Exercise 7.8: The transform is

$$h(x) = \int^x \frac{1}{\sigma v}\, dv = \sigma^{-1} \log x$$

The transformed process is $\{Y_t : t \geq 0\}$ given by $Y_t = \sigma^{-1}\ \log X_t$, which is governed by the SDE

$$dY_t = (\frac{rX_t}{\sigma X_t} - \frac{1}{2}\sigma)\ dt + dB_t = (\frac{r}{\sigma} - \frac{1}{2}\sigma)\ dt + dB_t$$

Alternatively, we could prefer to not scale with σ, so define a transformed process $Z_t = \log X_t$, corresponding to the SDE

$$dZ_t = (r - \frac{1}{2}\sigma^2)\ dt + \sigma\ dB_t$$

which has a constant noise intensity, which is not equal to 1 in general.

Exercise 7.9: The scale function s satisfies $s'\mu + \frac{1}{2}\sigma^2 s'' = 0$; the non-trivial solution is

$$s(x) = \exp\left(-2\mu x/\sigma^2\right)$$

and the transformed process $Y_t = s(X_t)$ satisfies

$$dY_t = s'(X_t)\sigma\ dB_t = -\frac{2\mu}{\sigma^2} Y_t\ dB_t.$$

In the x-coordinate, starting from a deterministic initial condition, a positive drift $\mu > 0$ implies that mode and mean of X_t is increasing. In the y-coordinate, the mode of Y_t is decreasing, but the convexity of s in concert with the variance of X_t implies that the expected value of Y_t is constant.

Exercise 7.10: The equation governing the scale function is

$$rxs'(x) + \frac{1}{2}\sigma^2 x^2 s''(x) = 0.$$

We verify the scale function directly by inserting the candidates in this equation. For $\nu \neq 0$, we use $s'(x) = x^{\nu-1}$, $s''(x) = (\nu - 1)x^{\nu-2}$. For $\nu = 0$, we use $s'(x) = 1/x$, $s''(x) = -1/x^2$.

Exercise 7.11: Define $X_t = k(t, B_t)$ with $k(t, b) = b/\sqrt{t}$; then Itô's lemma gives

$$dX_t = -\frac{1}{2} B_t t^{-3/2}\ dt + t^{-1/2}\ dB_t = -\frac{1}{2} X_t t^{-1}\ dt + t^{-1/2}\ dB_t$$

With $F_t = -\frac{1}{2} X_t t^{-1}$ and $G_t = t^{-1/2}$, and with the time change $U_t = \log t$, we get $H_t = 1/t$ and

$$dY_u = \frac{F_{T_u}}{H_{T_u}}\ du + \frac{G_{T_u}}{\sqrt{H_{T_u}}}\ dW_u = -\frac{1}{2} Y_u\ du + dW_u.$$

To avoid singularities in the time transform, we first restrict the time t to a compact interval $[a, b]$ with $0 < a < b < \infty$; we can later let $a \to 0$ and $b \to \infty$ and cover the entire time semiaxis $t \geq 0$. Thus, by rescaling both the dependent and the independent variable, we can transform Brownian motion to an Ornstein-Uhlenbeck process. Finally, the stationary variance of this Ornstein-Uhlenbeck process is 1, which follows from the algebraic Lyapunov equation, so the stationary distribution is a standard Gaussian, which coincides with the distribution of $Y_0 = B_1$.

Exercise 7.14: For X_t we get

$$dX_t = -sinB_t \circ dB_t = -Y_t \circ dB_t$$

while for Y_t we get

$$dY_t = \cos B_t \circ dB_t = X_t \circ dB_t.$$

Combining, we have

$$\begin{pmatrix} dX_t \\ dY_t \end{pmatrix} = \begin{pmatrix} -Y_t \\ X_t \end{pmatrix} \circ \ dB_t.$$

Exercise 7.15: We apply the transformation

$$Y_t = h(X_t) \text{ with } h(x) = \int^x \frac{1}{g(v)} \ dv$$

and find

$$dY_t = \frac{f(X_t)}{g(X_t)} \ dt + dB_t$$

Next, we rewrite the original equation governing $\{X_t\}$ as an Itô equation:

$$dX_t = (f(X_t) + \frac{1}{2} g(X_t) g'(X_t)) \ dt + g(X_t) \ dB_t$$

We Lamperti transform this equation with the same transformation and find

$$dY_t = \left(\frac{f(X_t)}{g(X_t)} + \frac{1}{2} g'(X_t) - \frac{1}{2} g'(X_t) \right) \ dt + dB_t = \frac{f(X_t)}{g(X_t)} \ dt + dB_t,$$

i.e., the same as we obtained with the direct transformation applied to the Stratonovich equation. Notice that the noise intensity is constant, which was the purpose of the transformation, so that the Itô and Stratonovich interpretation of the equation coincide. This explains why we reach the same result.

Exercise 7.16: Using Itô's lemma, we have

$$h'(x) = px^{p-1}, \quad h''(x) = p(p-1)x^{p-2},$$

so

$$dY_t = rpX_t^p\ dt + \sigma pX^p\ dB_t + \frac{1}{2}\sigma^2 p(p-1)X_t^p\ dt$$

or

$$dY_t = \left(rp + \frac{1}{2}\sigma^2 p(p-1)\right)\ dt + \sigma pY_t\ dB_t.$$

Hence, $\{Y_t\}$ is geometric Brownian motion with the drift parameter $rp + \sigma^2 p(p-1)/2$ and noise parameter σp.

Exercise 7.17: The verification consists of the following steps: We write $Y_t = h(B_t)$ and we use Itô's lemma to write $\{Y_t\}$ as an Itô process. Then we eliminate B_t from the drift and noise intensity. The following piece of `Maple` code does the computations:

```
h := sinh:

## Identify the inverse
hi := unapply(solve(y=h(b),b),y):

## Determine the increment of Y using Ito's lemma
dY := diff(h(B),B) * dB + 1/2 * diff(h(B),B,B) * dt:

## Eliminate B
dY := simplify( subs( B = hi(Y), dY ) );
```

The code produces the following output:

$$dY := (Y^2+1)^{1/2}\ dB + \frac{1}{2}Y\ dt$$

which confirms that $\{Y_t\}$ satisfied the SDE, as claimed.

The numerical verification is found in R-code `sinhB.R`.

Exercise 7.18: With Itô's lemma, we get

$$dZ_t = \dot{h}(X_t,t)\ dt + h'(X_t,t)\ dX_t + \frac{1}{2}h''(X_t,t)\ (dX_t)^2$$

and therefore

$$d\langle Z,Y\rangle_t = dZ_t\ dY_t = h'(X_t,t)\ dX_t\ dY_t = h'(X_t,t)\ d\langle X,Y\rangle_t$$

since the terms involving $dt\ dY_t$ and $(dX_t)^2\ dY_t$ vanish.

Exercise 7.19: The product rule applied to $Z_t = X_tY_t$ with $X_t = t$, $Y_t = B_t$ immediately gives

$$d(tB_t) = t\ dB_t + B_t\ dt$$

from which the integral formulation follows. The Itô integral $\int_0^t s \; dB_s$ has expectation 0 and variance $\int_0^t s^2 \; ds = t^3/3$ according to the Itô isometry. The integral $\int_0^t B_s \; ds$ has expectation 0 and a variance

$$\mathbf{E}\left(\int_0^t B_s \; ds \; \int_0^t B_u \; du\right) = \int_0^t \int_0^t (s \wedge u) \; ds \; du = 2\int_0^t \int_0^s u \; du \; ds = t^3/3,$$

i.e., the same. Here we have used that $EB_sB_u = s \wedge u = \min(s,u)$; compare (4.1). Since tB_t has expectation 0 and variance t^3, we find

$$\mathbf{Cov}(\int_0^t s \; dB_s, \int_0^t B_s \; ds) = \frac{1}{2}(t^3 - 2t^3/3) = t^3/6,$$

using the general formula $\mathbf{V}(X+Y) = \mathbf{V}X + \mathbf{V}Y + 2\mathbf{Cov}(X,y)$.

The code `int_s_dB_and_int_B_ds.R` verifies the result with a Monte Carlo simulation.

Exercise 7.20: By the Brownian bridge, $\mathbf{E}\{B_s|B_t\} = sB_t/t$ for $0 \leq s \leq t$. By Fubini's theorem, $\mathbf{E}\left\{\int_0^t B_s \; ds|B_t\right\} = \int_0^t (sB_t/t) \; ds = B_t/t \cdot \frac{1}{2}t^2 = \frac{1}{2}tB_t$.

By the product formula, $\int_0^t s \; dB_s = tB_t - \int_0^t B_s \; ds$, so $\mathbf{E}\left\{\int_0^t s \; dB_s|B_t\right\} = \frac{1}{2}tB_t$.

Exercise 7.21: The result follows directly from (7.10) with $f(x) = rx$, $g(x) = \sigma x$, and $h(x) = x$.

Exercise 7.22:

1. We have $\mathbf{E}X_t = \int_{-\infty}^{+\infty} (2\pi t)^{-1/2} \exp(b^2(1 - 1/(2t))) \; db$. The integrand is a Gauss bell if $t < 1/2$ and hence the expectation exists. If $t \geq 1/2$, then the integrand does not vanish, so the integral is $+\infty$.
2. We have $\mathbf{E}X_t = \int_{-\infty}^{+\infty} (2\pi t)^{-1/2} \exp(-b^2/(2t) + \exp b) \; db$. The integrand diverges to $+\infty$ as $b \to \infty$, for all values of $t > 0$.

Exercise 7.23: Set $h(t,b) = \exp(At + Gb)x$, then we have

$$\frac{\partial h}{\partial t} = Ah, \quad \frac{\partial h}{\partial b} = Gh, \quad \frac{\partial^2 h}{\partial b^2} = G^2 b$$

so Itô's lemma gives

$$dX_t = AX_t \; dt + GX_t \; dB_t + \frac{1}{2}G^2 X_t \; dt$$

as claimed.

Exercise 7.24:

1. Define $h(x) = |x|^2$ and $Y_t = h(X_t)$ and note that $\nabla h(x) = 2x^\top$. According to the chain rule of Stratonovich calculus, we have
$$dY_t = \frac{\partial h}{\partial x} \circ dX_t = 2X_t^\top \circ dX_t = 0.$$
2. Define $Y_t = UX_t$. Then $|Y_t| = |X_t|$ and
$$dY_t = U\ dX_t = U(I - \frac{1}{|X_t|^2} X_t X_t^\top) U^\top U \circ dB_t = (U - \frac{1}{|Y_t|^2} Y_t Y_t^\top) \circ dW_t$$
where $UB_t = dW_t$. Note that $\{W_t\}$ is Brownian motion; for example $W_t - W_s \sim N(0, I(t-s))$ for $0 \leq s \leq t$.
3. With Itô's formula, and using that $\mathbf{H}h = 2I$, we get
$$dY_t = \nabla h\, dX_t + \frac{1}{2} dX_t^\top \mathbf{H}h\, dX_t = \left[2X_t^\top \frac{1-n}{2|X_t|^2} X_t + \mathrm{tr}(I - \frac{1}{|X_t|^2} X_t X_t^\top) \right] dt = 0.$$
Here we have used that $g(x) = I - |x|^{-2} x x^\top$ is a symmetric projection matrix, i.e. $gg = g$.
Taking expectation in the Itô equation for $\{X_t\}$, and using that $|X_t|^2 = |x|^2$, we find that
$$\mathbf{E}^x X_t = x e^{t(1-n)/(2|x|^2}.$$
See source code `BM-on-the-sphere.R` for the numerical verification.

Exercise 7.25: We apply Itô's lemma applied to $Y_t = h(X_t)$ where $h(x) = \exp x$. We get
$$dY_t = Y_t(dX_t + \frac{1}{2}(dX_t)^2) = Y_t(G_t\ dB_t - \frac{1}{2} G_t^2\ dt + \frac{1}{2} G_t^2\ dt) = Y_t G_t\ dB_t.$$
See source code `ExponentialMartingale.R` for the numerical verification.

Exercise 8.2: Our first (somewhat naive) guess on the solution is the constant process $\{X_t^{(0)} : t \geq 0\}$ given by $X_t^{(0)} = 1$. Next, the Picard iteration gives us
$$X_t^{(i)} = 1 + \int_0^t r X_s^{(i-1)}\ ds + \sigma X_s^{(i-1)} \circ dB_s \text{ for } i = 1, 2, \ldots$$
and with the Stratonovich calculus, we get:
$$\begin{aligned} X_t^{(1)} &= 1 + \int_0^t r\ ds + \int_0^t \sigma \circ dB_s \\ &= 1 + rt + \sigma B_t \\ X_t^{(2)} &= 1 + \int_0^t r(1 + rs + \sigma B_s)\ ds + \int_0^t \sigma(1 + rs + \sigma B_s) \circ dB_s \\ &= 1 + (rt + \sigma B_t) + \frac{1}{2}(rt + \sigma B_t)^2 \end{aligned}$$

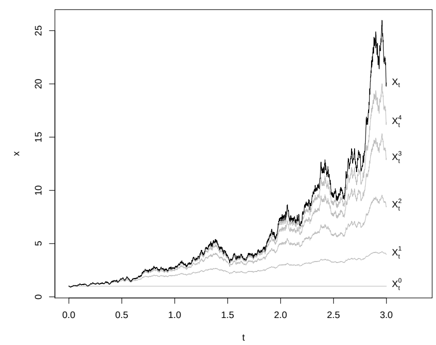
This motivates us to guess

$$X_t^{(n)} = \sum_{i=0}^{n} \frac{1}{i!}(rt + \sigma B_t)^i$$

which is the truncated Taylor expansion of an exponential function, i.e.,

$$X_t^{(n)} \to \exp(rt + \sigma B_t)$$

for all t and all ω. The following figure shows the first few iterates:



To show that this guess is correct, assume that it holds for n. We then get

$$X_t^{(n+1)} = 1 + \int_0^t \sum_{i=0}^{n} \frac{1}{i!}(rs + \sigma B_s)^i (r\ ds + \sigma\ dB_s) = \sum_{i=0}^{n+1} \frac{1}{i!}(rs + \sigma B_s)^i$$

which was to be shown.

Exercise 8.3: For both models, the noise intensity g is constant, so it is all about the drift. In both cases, the drift is continuously differentiable, so locally Lipschitz, and uniqueness is guaranteed. For the double well model, we have

$$xf(x) = rx^2 - qx^4 \leq rx^2$$

so the condition in theorem 8.3.1 is satisfied with $C = r$.

For the van der Pol oscillator, we have

$$(x\ v) \begin{pmatrix} v \\ (\mu(1 - x^2)v - x \end{pmatrix} = xv + \mu v^2 - \mu x^2 v^2 - xv \leq \mu v^2$$

so the condition in theorem 8.3.1 is satisfied with $C = r$, if we use Euclidean norm for the state (x, v).

Exercise 8.4: The noise intensity is globally Lipschitz, so it is all about the drift $f(x) = x(1-x)$. This satisfies $xf(x) \leq 1 + x^2$ as long as $x \geq 0$. So existence and uniqueness is guaranteed as long as the process stays non-negative. But if the process ever hits 0, then it will stay there, since 0 is an equilibrium. Therefore existence and uniqueness is guaranteed. *Note:* In fact, the process never hits 0, as we will see in Section 11.7 and exercise 12.7.

Exercise 8.5: The functions f, g are Lipschitz continuous on any interval $[\epsilon, \infty)$ with $\epsilon > 0$, so existence and uniqueness holds up to the time the process hits ϵ. Since $\epsilon > 0$ is arbitrary, existence and uniqueness hold until the process hits $x = 0$.

Exercise 8.6: We have

$$g_1(x) = 1, \quad g_2(x) = x$$

so

$$L_1 V = V', \quad L_2 V = xV'$$

and hence

$$L_1 L_2 V = (xV')' = xV'' + V'$$

while

$$L_2 L_1 V = xV''$$

Therefore, $(L_2 L_1 - L_1 L_2)V$ is not identically 0, and the noise terms do not commute. More concisely, we have $g_1' g_2 = 0 \neq 1 = g_2' g_1$.

Exercise 8.7: We have

$$(\nabla g_1) g_2 = 0, \quad (\nabla g_2) g_1 = \begin{bmatrix} 0 & 0 & 0 \\ -\sin\theta & 0 & 0 \\ \cos\theta & 0 & 0 \end{bmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -\sin\theta \\ \cos\theta \end{pmatrix}.$$

Since these two results differ, the noise terms do not commute. The interpretation of this result is that the unicycle will diffuse not just in the forward-backward direction, but also sideways. More elaborately, we have

$$g_1(\theta, x, y) = (1, 0, 0)^\top, \quad g_2(\theta, x, y) = (0, \cos\theta, \sin\theta)^\top$$

Hence, the operators L_i are

$$L_1 V = \frac{\partial V}{\partial \theta}, \quad L_2 V = \cos\theta \frac{\partial V}{\partial x} + \sin\theta \frac{\partial V}{\partial y}.$$

We therefore get

$$\begin{aligned} L_1 L_2 V &= \frac{\partial}{\partial \theta}\left(\cos\theta \frac{\partial V}{\partial x} + \sin\theta \frac{\partial V}{\partial y}\right) \\ &= -\sin\theta \frac{\partial V}{\partial x} + \cos\theta \frac{\partial^2 V}{\partial x \partial \theta} + \cos\theta \frac{\partial V}{\partial y} + \sin\theta \frac{\partial^2 V}{\partial y \partial \theta} \end{aligned}$$

while

$$L_2L_1V = \cos\theta\frac{\partial^2 V}{\partial x\partial\theta} + \sin\theta\frac{\partial^2 V}{\partial y\partial\theta}$$

so that

$$(L_1L_2 - L_2L_1)V = -\sin\theta\frac{\partial V}{\partial x} + \cos\theta\frac{\partial V}{\partial y}$$

which is not identically 0. Hence the noise terms do not commute.

Exercise 8.8: We have

$$Z_{t+h} = X_t(1 + rh + \sigma\Delta B)$$

and thus

$$\bar{f} = \frac{1}{2}(rX_t + rZ_{t+h}) = (r + \frac{1}{2}r^2h + \frac{1}{2}r\sigma\Delta B)X_t$$

and

$$\bar{g} = \frac{1}{2}(\sigma X_t + \sigma Z_{t+h}) = (\sigma + \frac{1}{2}r\sigma h + \frac{1}{2}\sigma^2\Delta B)X_t$$

and therefore

$$X_{t+h} = X_t(1 + rh + \sigma\Delta B + \frac{1}{2}(rh + \sigma\Delta B)^2).$$

Exercise 9.1: The stationary distribution can be written as

$$\begin{aligned}
\rho(x) &= \frac{1}{Z}\exp\left(\int_{x_0}^{x}\frac{u(y)}{D(y)}\,dy\right)\\
&= \frac{1}{Z}\exp\left(\int_{x_0}^{x}\frac{f(y) - D'(y)}{D(y)}\,dy\right)\\
&= \frac{1}{Z}\exp\left(\int_{x_0}^{x}\frac{f(y)}{D(y)}\,dy - \log(D(x)) + \log(D(x_0))\right)\\
&= \frac{D(x_0)}{Z\cdot D(x)}\exp\left(\int_{x_0}^{x}\frac{f(y)}{D(y)}\,dy\right)
\end{aligned}$$

Redefining $Z := Z/D(x_0)$, we obtain the desired result.

Exercise 9.2: The forward Kolmogorov equation can be written in terms of advection and diffusion:

$$D(x) = \frac{1}{2}g^2(x) = \frac{1}{2}\gamma^2 x, \quad u(x) = f(x) - D'(x) = \lambda(\xi - x) - \frac{1}{2}\gamma^2.$$

To find the stationary distribution using (9.6) on p. 207), we first identify the antiderivative

$$\int_{x_0}^{x}\frac{u(y)}{D(y)}\,dy = \int_{x_0}^{x} 2\frac{\lambda\xi - \gamma^2}{\gamma^2}\frac{1}{y} - \frac{2\lambda}{\gamma^2}\,dy = \left(\frac{2\lambda\xi -}{\gamma^2} - 1\right)\log(x/x_0) - \frac{2\lambda}{\gamma^2}(x - x_0)$$

so that the un-normalized density is

$$\phi(x) = \frac{1}{Z} x^{2\lambda\xi/\gamma^2 - 1} e^{-2\lambda/\gamma^2 \; x} = \frac{1}{Z} x^{\nu-1} e^{-\omega x};$$

i.e., a Gamma distribution with rate and shape parameter as claimed. The normalization constant follows from the properties of the Gamma distribution, which also give us the stationary mean and variance:

$$\mathbf{E} X_t = \nu/\omega = \xi, \quad \mathbf{V}^\rho X_t = \nu/\omega^2 = \xi\gamma^2/(2\lambda).$$

This agrees with the results of exercise 8.5, as expected.

Exercise 9.3:

1. We have $Lh(x) = r(1 - x/K) - \sigma^2/2$. The result follows from taking expectation, setting equal to 0, and isolating $\mathbf{E}X_t$.
2. We have $Lh(x) = rx(1 - x/K)$. The first result follows from taking expectation and setting equal to 0. Then use the definition of the variance, $\mathbf{V}X_t = \mathbf{E}X_t^2 - (\mathbf{E}X_t)^2$.
3. We first maximize $\mathbf{V}X_t$ w.r.t. $\mathbf{E}X_t \in [0, K]$ and find that this is maximal at $\mathbf{E}X_t = K/2$. Next, we see that this expectation is obtained with $\sigma^2 = r$. At lower noise intensities, the noise is insufficient to excite large fluctuations. At higher noise intensities, the noise hampers the growth so that the entire process live at lower levels. Ultimately, when the noise intensity reaches $\sigma^2 = 2r$, the expectation drops to 0 so that the population has died out, in which case there are no more fluctuations.

Exercise 9.4: The solution to the state equation is geometric Brownian motion, starting at $X_0 = x$:

$$X_t = x e^{(r - \frac{1}{2}\sigma^2)t + \sigma \; B_t}$$

and, using that $B_t \sim N(0, t)$, we can write $B_t = \Xi\sqrt{t}$ where $\Xi \sim N(0, 1)$, so

$$k(0, x) = \int_{-\infty}^{+\infty} (x e^{(r - \frac{1}{2}\sigma^2)t + \sigma\xi\sqrt{t}} - K)^+ \phi(\xi) \; d\xi$$

where $\phi(\cdot)$ is the p.d.f. of a standard Gaussian random variable. We first identify the threshold $\xi = -d$ where the integrand touches 0:

$$x e^{(r - \frac{1}{2}\sigma^2)t - \sigma d\sqrt{t}} = K \Leftrightarrow -d = \frac{\log(K/x) - (r - \frac{1}{2}\sigma^2)t}{\sigma\sqrt{t}}$$

so

$$k(0, x) = \int_{-d}^{+\infty} (x e^{(r - \frac{1}{2}\sigma^2)t + \sigma\xi\sqrt{t}} - K)\phi(\xi) \; d\xi.$$

We next use

$$\int_{-d}^{\infty} \phi(\xi)\ d\xi = \Phi(d) \text{ and } \int_{-d}^{\infty} e^{\sigma\xi\sqrt{t}}\phi(\xi)\ d\xi = e^{\frac{1}{2}\sigma^2 t}\Phi(d+\sigma\sqrt{t}).$$

where $\Phi(\cdot)$ is the standard Gaussian c.d.f. Combining these terms, we get the final expression:

$$k(0,x) = xe^{rt}\Phi(d+\sigma\sqrt{t}) - K\Phi(d).$$

The discounted price follows, multiplying with $\exp(-rt)$. Note that $\Phi(d)$ denotes the probability that the option is "in the money", i.e., the price of the stock exceeds the strike price at the time of expiry, so the option will have a positive value. Thus, $K\Phi(d)$ is the expected price we will pay for the stock at the time of expiry. The first term is the expected revenue we will get from selling the stock: $x\exp(rt)$ is the expected price of the stock, and the term $\Phi(d+\sigma\sqrt{t})$ corrects for the fact that we only sell the stock if the price exceeds K.

See source code `BlackSholes.R` for the numerical computation and visualization.

Exercise 9.5:

1. We have
$$L\psi = -x\psi' + \psi'', \quad L^*\phi = (x\phi)' + \phi''.$$

2. We find
$$\begin{array}{ll} LH_0 = 0, & LH_1 = -x = -H_1 \\ LH_2 = -2x^2 + 2 = -2H_2, & LH_3 = -3x^3 + 3x + 6x = -3H_3 \end{array}$$
so these are eigenfunctions corresponding to the eigenvalues 0,1,2 and 3, respectively.

3. For the mean, we know from Section 7.4.1 that $\mathbf{E}^x X_t = x\exp(-t)$. This is consistent with
$$\mathbf{E}^x H_1(X_t) = (e^{Lt}H_1)(x) = e^{-t}H_1(x) = e^{-t}x.$$
For the variance, we know from Section 7.4.1 that
$$\mathbf{V}^x X_t = 1 - e^{-2t}$$
which implies that the mean square satisfies
$$\mathbf{E}^x X_t^2 = (\mathbf{E}^x X_t)^2 + \mathbf{V}^x X_t = 1 + e^{-2t}(x^2-1).$$
This can also be written
$$\mathbf{E}^x H_2(X_t) = e^{-2t}H_2(x)$$
which is consistent with $LH_2 = -2H_2$ and $\mathbf{E}^x H_2(X_t) = (e^{Lt}H_2)(x)$.

4. Let k be an arbitrary smooth test function with bounded support and let $\langle h, k\rangle = \int hk\ dx$ denote the usual inner product on $\mathcal{L}_2$. Then

$$\begin{aligned}\langle L^*(\rho\psi), k\rangle &= \langle \rho\psi, Lk\rangle \\ &= \langle \psi, Lk\rangle_\rho \\ &= \langle L\psi, k\rangle_\rho \\ &= \langle \lambda\psi, k\rangle_\rho \\ &= \langle \lambda\rho\psi, k\rangle.\end{aligned}$$

   Since k is arbitrary, we conclude that $L^*(\rho\psi) = \lambda\rho\psi$.

5. See source code `OU-spectrum.R`

Exercise 9.6: We have

$$\mathbf{E}[h(X_0)h(X_t)] = \mathbf{E}\left[h(X_0)\mathbf{E}\{h(X_t)|X_0\}\right]$$

and the backward Kolmogorov equation gives us

$$\mathbf{E}\{h(X_t)|X_0 = x\} = (e^{Lt}h)(x).$$

Combining, we get

$$\mathbf{E}\left[h(X_0)\mathbf{E}\{h(X_t)|X_0\}\right] = \mathbf{E}\left[h(X_0)(e^{Lt}h)(X_0)\right] = \int_{\mathbf{X}} \rho(x)h(x)(e^{Lt}h)(x)\ dx$$

as desired. For the Ornstein-Uhlenbeck process, with $h(x) = x$, we have $Lh = -\lambda h$, so $(\exp(Lt)h) = \exp(-\lambda t)h$. Therefore, we get

$$\mathbf{E}X_0X_t = \mathbf{E}X_0^2 e^{-\lambda t} = e^{-\lambda t}\frac{\sigma^2}{2\lambda}.$$

Here we have used the stationary distribution (e.g., example 9.9.2). The result is consistent with what we found earlier (Section 5.10). Finally, if $Lh = \lambda h$, then

$$\mathbf{E}h(X_t)h(X_0) = \int_{\mathbf{X}} \rho(x)h(x)e^{\lambda t}h(x)\ dx = e^{\lambda t}\mathbf{E}|h(X_0)|^2$$

for $t \geq 0$.

Exercise 9.7: The forward Kolmogorov equation is

$$\dot{\rho} = -(\rho\sin x)' + \frac{1}{2}\sigma^2\rho''.$$

The stationary distribution is the canonical distribution

$$\rho(x) = \frac{1}{Z}\exp(-U(x)/D)$$

where, as always, $D = \sigma^2/2$ and the potential U is an antiderivative to $-f(x) = \sin x$, i.e. $U(x) = -\cos x$. So

$$\rho(x) = \frac{1}{Z}\exp(D^{-1}\cos x)$$

as claimed, with $\kappa = 1/D = 2/\sigma^2$.

The question does not ask us to find Z, but we can: We normalize the distribution over $[0, 2\pi)$, obtaining

$$Z = \int_0^{2\pi} \exp(-U(x)/D)\ dx = 2\pi I_0(2/\sigma^2)$$

where $I_0(\cdot)$ is the modified Bessel function of the first kind of order 0.

For the numerical part of the exercise, see source code `vonMises.R`.

Exercise 9.8:

1. The forward Kolmogorov equation is

$$\dot{\phi} = -(\lambda(\xi - y)\phi)' + (\frac{1}{2}\gamma^2 y\phi)''.$$

2. We take as starting point the setting of Section 7.5.2, i.e., we have n independent Ornstein-Uhlenbeck processes. Assume that the state of these processes are known at time $t = 0$, and that their sum of squares is Y_0. At time t, these processes have relaxed partially towards the origin and are Gaussian distributed. We can now form Y_t as the sum of squares of these n random variables, which implies that a re-scaling of Y_t (to ensure that each Ornstein-Uhlenbeck variables has unit variance) will follow a non-central chi-squared distribution.

3. We pose the model for n Ornstein-Uhlenbeck processes, as in Section 7.5.2:

$$dX_t^{(i)} = -\mu X_t^{(i)}\ dt + \sigma\ dB_t^{(i)}$$

with $\mu = \lambda/2$, $\sigma = \gamma/2$, and $n = 4\lambda\xi/\gamma^2$. Here we assume that n is integer. Let $Y_0 = y_0$ be given; set $X_t^{(i)} = \sqrt{y_0/n}$. Note that it does not matter how we distribute the energy y_0 over the n modes. Then, from the properties of the Ornstein-Uhlenbeck process (Section 7.4.1), we have

$$X_t^{(i)} \sim N(\sqrt{y_0/n}e^{-t\lambda/2}, \frac{\gamma^2}{4\lambda}(1 - e^{-\lambda t})).$$

Since the non-central chi-squared distribution concerns sum-of-squares of Gaussian variables, which each have unit variance, we define

$$c = \frac{2\lambda}{\gamma^2(1 - e^{-\lambda t})}$$

so that
$$\sqrt{2c}X_t^{(i)} \sim N(\sqrt{2cy_0/n}e^{-t\lambda/2}, 1).$$

Therefore, $2cY_t$ is non-central chi-squared distributed with n degrees of freedom, and a non-centrality parameter ν which is the sum of the squared means, i.e.
$$\nu = 2cy_0 e^{-\lambda t}.$$

Let $f(z)$ be the p.d.f. of this non-central chi-squared distribution (available in `R` and other software environments), then the p.d.f. of Y_t at y is
$$f(2cy)2c$$

Using the analytical expression for the p.d.f. of the non-central chi-squared distribution, we find that that the p.d.f. of Y_t at y is
$$ce^{(-2cy+\nu)/2}\left(\frac{2cy}{\nu}\right)^{n/4-1/2} I_{n/2-1}(\sqrt{2c\nu y})$$

where I is a modified Bessel function of the first kind.

Exercise 9.9: We have
$$f(x) = u(x) + \nabla D(x), \quad g(x) = (2D)^{1/2}(x)$$

and the Euler-Maruyama method is
$$X_{t+h} = X_t + (u(X_t) + \nabla D(X_t))\ h + (2D(X_t))^{1/2}\ (B_{t+h} - B_t)$$

The term ∇D pushes particles in direction of higher diffusivity; without this term, particles will tend to aggregate in regions with low diffusivity, which is not in agreement with Fickian diffusion.

Exercise 9.10: From Section 7.7, we know that the equation governing $\{Y_u\}$ is $dY_u = h^{-1}(Y_u)f(Y_u)\ du + h^{-1/2}(Y_u)g(Y_u)\ dW_u$. Let $\rho(y)$ be the stationary density of Y_u, then
$$\begin{aligned}\rho(y) &= \frac{2h(y)}{\bar{Z}_Y g^2(y)} \exp\left(\int_{y_0}^{y} \frac{2f(x)}{g^2(x)}\ dx\right) \\ &= \frac{\bar{Z}_X}{\bar{Z}_Y} h(y)\phi(y)\end{aligned}$$

which should be shown; here $\bar{Z}_X$ and $\bar{Z}_Y$ are the two normalization constants. Note that this result can be explained as follows: The sample path of $\{Y_u : u \geq 0\}$ visits exactly the same points as $\{X_t : t \geq 0\}$, but the u-time the process $\{Y_u : u \geq 0\}$ spends in a certain region dy is a factor $h(y)$ larger than the t-time that $\{X_t : t \geq 0\}$ spends in the same region.

Exercise 9.11: We have $A\Sigma + \Sigma A^\top + GG^\top = 0$, and the stationary density is

$$\rho(x) = |2\pi\Sigma|^{-1/2} \exp\left(-\frac{1}{2}x^\top \Sigma^{-1} x\right).$$

The flux of probability in stationarity at x is

$$\rho A x - \frac{1}{2}GG^\top \nabla\rho = \rho\left(A + \frac{1}{2}GG^\top \rho\Sigma^{-1}\right)x.$$

This vanishes everywhere iff

$$A + \frac{1}{2}GG^\top \rho\Sigma^{-1} = 0 \Leftrightarrow 2A\Sigma + GG^\top = 0$$

which, combined with the stationary Lyapunov equation, gives $A\Sigma = \Sigma A^\top$.

This holds, as we have already shown, for the Ornstein-Uhlenbeck process, and more generally and interestingly for the equation

$$dX_t = -QX_t\ dt + \sigma\ dB_t$$

in n dimensions, where $\sigma > 0$ is a scalar, $\{B_t\}$ is n-dimensional Brownian motion, and $Q = Q^\top$. Here, the stationary distribution is a Gibbs canonical distribution corresponding to the potential $U(x) = \frac{1}{2}x^\top Qx$, i.e., a Gaussian with variance $Q^{-1}\sigma^2/2$.

One example where detailed balance does *not* hold is the noisy harmonic oscillator (Section 5.11).

Exercise 9.12:

1. The generator is

   $$(Lh)(x) = \nabla h\ Ax + \frac{1}{2}\mathrm{tr}[\mathbf{H}hGG^\top],$$

   and with the particular form $h(x) = x^\top Qx + q$, we get

   $$(Lh)(x) = x^\top[QA + A^\top Q]x + \mathrm{tr}[QGG^\top].$$

   Here, we have used $2x^\top QAx = x^\top(QA + A^\top Q)x$ so that the matrix in the quadratic form is symmetric.

2. We get

   $$x^\top[\dot P_t + P_tA + A^\top P_t]x + \dot p_t + \mathrm{tr}[P_tGG^\top] = 0.$$

   This holds for all x iff

   $$\dot P_t + P_tA + A^\top P_t = 0, \quad \dot p_t + \mathrm{tr}[P_tGG^\top] = 0.$$

   In addition, we have the terminal conditions $P_T = Q$, $p_T = q$.

Exercise 9.13:

1. We have $g_1(\theta, x, y) = (1, 0, 0)^\top$ and $g_2(\theta, x, y) = (0, \cos\theta, \sin\theta)^\top$. So

$$[g_1, g_2](x) = \nabla g_1 \cdot g_2 - \nabla g_2 \cdot g_1 = \begin{bmatrix} 0 & 0 & 0 \\ -\sin\theta & 0 & 0 \\ \cos\theta & 0 & 0 \end{bmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ -\sin\theta \\ \cos\theta \end{pmatrix}.$$

It is now easy to see that the three vectors span $\mathbf{R}^3$ - for example, we can compute the determinant

$$|[g_1 \; g_2 \; [g_1, g_2]| = \left| \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & -\sin\theta \\ 0 & \sin\theta & \cos\theta \end{bmatrix} \right| = 1.$$

Exercise 9.15: See source code `von-Mises-identification.R`.

Exercise 10.1: The stationary Riccati equation is

$$-2\lambda P + \sigma^2 - P^2 c^2/d^2 = 0,$$

and the maximal solution is

$$\Sigma = \frac{-2\lambda + \sqrt{4\lambda^2 + 4\sigma^2 c^2/d^2}}{2c^2/d^2}.$$

Note that $\Sigma > 0$. For fixed λ, Σ is a decreasing function of c and an increasing function of d. For fixed c, d, Σ is a decreasing function of λ. When $d \to 0$, we have $\Sigma \approx \sigma d/c \to 0$. When $d \to \infty$, there are two situations: If $\lambda > 0$, then the system dynamics is stable, and with high measurement noise, the best we can do is to use the estimate $\hat{X}_t = 0$, which will lead to an estimation error with variance $\sigma^2/(2\lambda)$. To obtain this result, we use $\sqrt{1+x} = 1 + x/2 + o(x)$. Conversely, with $\lambda < 0$, the state diverges, so we need to use measurements, even if they are noisy. We obtain $\Sigma \approx -2d^2\lambda/c^2 \to \infty$ as $d \to \infty$.

The Kalman gain is

$$K = \Sigma c/d^2 = \frac{-2\lambda + \sqrt{4\lambda^2 + 4\sigma^2 c^2/d^2}}{2c}$$

which is positive. For fixed λ, it is an decreasing function of d and an increasing function of c. For fixed c, d, it is a decreasing function of λ. For $d \to 0$, we have $K \to \infty$. For $d \to \infty$ and $\lambda > 0$, we have $K \to 0$, which agrees with what we found for Σ. For $d \to \infty$ and $\lambda < 0$, we have $K \to -2\lambda/c$.

Finally, we have

$$A - KC = -\sqrt{\lambda^2 + \sigma^2 c^2/d^2}$$

which is always negative, and becomes more negative when we increase $|\lambda|$, σ or the signal-to-noise ratio c/d. When $d \to 0$, we have $A - CK \to -\infty$, so that we use measurements agressively to make the error decorrelate fast. When $\lambda > 0$ and $d \to \infty$, we have $A - KC \to -\lambda$, which agrees with our previous results that in this case we do not use measurements at all. When $\lambda < 0$ and $d \to \infty$, we find $A - KC \to \lambda$. In this case the optimal filter flips the unstable eigenvalue $-\lambda$ of the system dynamics into a stable eigenvalue λ of the error dynamics.

Exercise 10.2: See the source code `kalman.R`.

Exercise 10.3:

1. We have

$$\begin{aligned} d\tilde{X}_t &= dX_t - d\hat{X}_t \\ &= AX_t\ dt + u_t\ dt + G\ dB_t \\ &\quad - [(A-KC)\hat{X}_t\ dt + u_t\ dt + K(dY_t - C\hat{X}_t\ dt)] \\ &= (A-KC)\tilde{X}_t + G\ dB_t + KD\ dW_t. \end{aligned}$$

2. The algebraic Lyapunov equation is

$$(A-KC)\Sigma + \Sigma(A-KC)^\top + GG^\top + KDD^\top K^\top = 0.$$

3. Minimizing $\mathrm{tr}(\Sigma Q)$ subject to the constraint of the algebraic Lyapunov equations, we include this in the objective using a Lagrange multiplier, say Λ:

$$\mathrm{tr}(\Sigma Q) + \mathrm{tr}\Lambda[(A-KC)\Sigma + \Sigma(A-KC)^\top + GG^\top + KDD^\top K^\top].$$

This must be minimized w.r.t. Σ, K, and Λ. It is linear in Σ, so at the optimum, the criterion must be stationary w.r.t. Σ:

$$Q + \Lambda(A-KC) + (A-KC)^\top\Lambda = 0.$$

Since $Q > 0$ and $A - KC$ must be stable, this implies that $\Lambda > 0$. For given Σ, Λ, we now identify a stationary point w.r.t. K by taking matrix derivative:

$$-2C\Sigma\Lambda + 2DD^\top K^\top \Lambda = 0$$

which is satisfied iff

$$K = \Sigma C^\top (DD^\top)^{-1}.$$

Inserting this in the algebraic Lyapunov equation, we obtain the algebraic Riccati equation (10.8).

Exercise 10.4: See source code `diffusion-bridge.R`

Exercise 11.1: The quickest way to confirm this expression is to see that it agrees with the original expression when $\lambda = 1$, $\sigma = \sqrt{2}$, and that it is dimensionally correct in that time has been rescaled with $1/\lambda$ and space with $\sqrt{\sigma^2/2\lambda}$. More elaborately, we would pose the equation for k:

$$\frac{1}{2}\sigma^2 k' = xk + 1, \quad k(0) = 0,$$

which has the solution

$$k(x) = -\frac{2}{\sigma^2} \int_0^x e^{\frac{\lambda}{\sigma^2}(x^2-y^2)}\, dy = -\frac{2\sqrt{\pi}}{\sigma\sqrt{\lambda}} e^{\lambda x^2/\sigma^2} (\Phi(x\sqrt{2\lambda/\sigma^2}) - \frac{1}{2})$$

To reach the final approximation for $h(0) = -\int_0^l k(x)\, dx$, we use that the integral is dominated by the contribution from the region where $\Phi \approx 1$ (Exercise 2.7 provides bounds which can be used to assess the error made here; it is negligible). Finally, we use the asymptotic expansion of the integral, with the substitution $y = x\sqrt{2\lambda}/\sigma$.

Exercise 11.2: We have $\mathbf{E}^x \tau = h(x)$ where

$$-uh' + Dh'' + 1 = 0, \quad h(R) = 0, \quad h'(0) = 0.$$

Here, $D = \sigma^2/2$. With $k(x) = h'(x)$, we find $-uk + Dk' + 1 = 0$, which has the general solution is $k(x) = 1/u + c\exp(ux/D)$. With the boundary condition $k(0) = 0$, we get $c = -1/u$. Now integrate k using the boundary condition $h(R) = 0$.

Exercise 11.3: Since $f = 0$, we find $\phi(x) = 1$. Thus $s(x) = x$ and $h(x) = x/l$. The varying diffusivity affects the time it takes to reach the boundary, but does not affect the eventual outcome.

Exercise 11.4: We claim that h is concave, i.e., $h'' < 0$: Note first that ϕ is positive, so the scale function s is increasing. Therefore $c_1 > 0$ and hence h is increasing, which should not come as a surprise: The further one starts to the right, the greater is the probability of exit to the right. Now, from the equation $fh' + \frac{1}{2}g^2h'' = 0$ we find

$$h'' = -\frac{2fh'}{g^2}$$

Therefore, in any region where f is positive, h'' must be negative and thus h is concave. Here, since $f(x) = D'(x) > 0$, this applies to the entire interval $[0, l]$.

The graph of a concave function lies above any chord, and hence $h(l/2) > (h(0) + h(l))/2 = 1/2$. So the process is more likely to exit at the boundary point where the diffusivity is high. This result adds to our understanding that pure Fickian diffusion, when the diffusivity varies with space, is biased towards regions with high diffusivity.

Exercise 11.5: When h is constant-quadratic, as assumed, the sub-generator has the form

$$(Lh - mh)(x) = x^\top \left[A^\top P + PA - mP\right] x + \mathrm{tr}PGG^\top - mp,$$

(compare exercise 9.12) so the steady-state equation is

$$x^\top \left[A^\top P + PA - mP + Q\right] x + \mathrm{tr}PGG^\top - mp = 0.$$

This holds for all x iff

$$A^\top P + PA - mP + Q = 0, \quad \mathrm{tr}PGG^\top = mp.$$

This equation has a solution with the given characterization, provided that $A - mI/2$ is strictly stable (all eigenvalues in the open left half of the complex plane). Here I is the identity matrix.

Exercise 11.6: When $x \geq 1$, we have $\tau = 0$ and $h(X_\tau) = x$. When $x < 1$, we have $\tau < \infty$ a.s. (Theorem 4.3.3), and $X_\tau = 1$ so that $h(X_\tau) = 1$. In summary $\mathbf{E}h(X_\tau) = h(X_\tau) = x \wedge 1$.

We also have $Lh \equiv 0$ so $h(x) + \mathbf{E}^x \int_0^\tau Lh(X_t)\, dt = h(x) = x$.

This h does not have bounded support and τ does not have finite expectation, so Dynkin's formula needs not apply. In this situation, the Itô integral that appears in the proof of Dynkin's formula is indeed a martingale, so the problem is not that h does not have bounded support, but rather that τ has infinite expectation.

Exercise 11.7: The probability that $h(x) = \mathbf{P}^x\{\tau < \infty\}$ is governed by the equation $Lh = 0$ with boundary conditions $h(0) = 1$, $h(\infty) = 0$, and the solution to this boundary value problem is $h(x) = \exp(-ux/D)$. Note that we have to be somewhat careful with these boundary value problems on infinite domains: If u had been negative, we would have had $h(x) \equiv 1$. A more stringent approach would be to first truncate the domain to $[0, l]$, apply the results of Section 11.4.1, and then let $l \to \infty$.

To show that S is exponentially distributed, simply note that the random variable S does not depend on the parameter x, and that $S \geq x$ if and only if $\tau < \infty$, i.e., $\mathbf{P}\{S \geq x\} = \exp(-ux/D)$.

Exercise 11.8: See the source code `Time-to-boundary.R`.

Exercise 11.9: The result is obvious since $\{cX_t\}$ is a martingale and $\mathbf{E}\tau < \infty$, and also follows from Dynkin's formula with $h(x) = cx$. It is still noteworthy, since it holds regardless of the shape of the domain, and regardless of g (as long as the diffusion is regular).

Exercise 11.10: Consider the extended state $Y_t = (X_t, T_t)$ with $dT_t = dt$. Then the generator of $\{Y_t\}$, say M, is $Mf = \dot{f} + Lf$. Set

$f(x,t) = k(x) + 2th(x) + t^2$, then $Mf = -2h + 2h + 2tLh + 2t = 0$. It follows that $k(x,t) = \mathbf{E}\{T_\tau^2 | X_0 = x, T_0 = t\}$. Now the result follows with $t = 0$ since $T_\tau = t + \tau$.

Exercise 11.11:

1. If the population abundance $\{X_s : s \geq 0\}$ were known, the survival function would be
$$\mathbf{P}\{\tau > t\} = e^{-\int_0^t \mu(X_s)\ ds}$$
where τ is the random time of death, and so the expected remaining lifetime would be
$$\mathbf{E}\tau = \int_0^\infty e^{-\int_0^t \mu(X_s)\ ds}\ dt.$$
With the population size unknown but following the stochastic logistic growth equation, we have
$$\mathbf{E}^x\tau = \mathbf{E}^x \int_0^\infty e^{-\int_0^t \mu(X_s)\ ds}\ dt.$$
This corresponds to the Feynman-Kac formula with cumulated rewards, i.e., (11.11), with $r \equiv 1$. We therefore find that the expected lifetime $h(x) = \mathbf{E}^x\tau$ is governed by (11.12.

2. See source code `life-expectancy.R` for the implementation.

3. When the parameters μ_0 is large, the individual animal has a short lifespan relative to the time scales of population growth. For very large μ_0, the animal effectively dies before the population size has changed, so that its life expectancy is $\mathbf{E}^x\tau \approx 1/\mu(x)$. As μ_0 is decreased, the animal lives to see the population size approach its stationary distribution, and therefore the effect of the initial condition $X_0 = x$ is less pronounced.

Exercise 12.1: Define $B = A - I\lambda$ where I is the identity matrix. Then $S_t = \exp(\lambda t)\exp(Bt)$ and therefore $\bar{\sigma}(S_t) = \exp(\lambda t) \cdot \bar{\sigma}(\exp(Bt))$. B has all eigenvalues in the closed left half of the complex plane, and therefore all elements of $\exp(Bt)$ can grow at most polynomially in time. It follows that
$$\limsup_{t\to\infty} \frac{1}{t} \log \bar{\sigma}(S_t) \leq \lambda.$$
On the other hand, let v_1 be a right eigenvector of A corresponding to an eigenvalue λ_1 such that $\mathbf{Re}\lambda_1 = \lambda$. Then $S_t v_1 = \exp(\lambda_1 t) v_1$ which implies that $\bar{\sigma}(S_t) \geq |\exp(\lambda_1 t)| = \exp(\lambda t)$, for all t. Hence
$$\liminf_{t\to\infty} \frac{1}{t} \log \bar{\sigma}(S_t) \geq \lambda.$$
The conclusion follows.

Exercise 12.3: Define $Y_t = \log X_t = (r - \sigma^2/2)t + \sigma B_t$. Then we know that $\sup\{Y_t : t \geq 0\}$ is exponentially distributed with rate parameter $\lambda := 1 - 2r/\sigma^2$. It follows that

$$\mathbf{P}\{S > s\} = \mathbf{P}\{\log S > \log s\} = \exp(-\lambda \log s) = s^{-\lambda}.$$

I.e., S is Pareto distributed with scale parameter 1 and shape parameter λ. See the source code `Pareto.R` for a simulation which verifies the result.

Exercise 12.4: Using that $G^2 = -4I$, we see from exercise 7.23 that the solution is

$$X_t = \exp(It + GB_t)X_0$$

so

$$\|X_t\| = e^t \|X_0\|$$

from which the conclusion follows.

Exercise 12.5: See source code `double-well-lyap.R`. With $\lambda = 1,\ 0,\ 1$, we find Lyapunov exponents of approximately -0.13, -0.7, and -1.5, respectively.

Exercise 12.6: With $f(x,v) = (v, \mu v(1 - x^2) - x)$, the Jacobian is

$$\nabla f = \begin{bmatrix} 0 & 1 \\ -2\mu x v - 1 & \mu(1 - x^2) \end{bmatrix}$$

while $g = (0, \sigma)$, $\nabla g = 0$. So the sensitivity equation is $dS_t = \nabla f \ S_t \ dt$.
See source code `vanderPol-lyap.R` for the numerical analysis.

Exercise 12.7:

1. V is convex and we have $V(x) \to \infty$ as $x \to \infty$ and as $x \to 0$. So the set $\{x : 0 \leq V(x) \leq K\}$ is a closed interval.

   Finding the stationary point $V'(1/\sqrt{2}) = 0$, we conclude that $V(x) \geq (1 + \log 2)/2$. Next, we have $LV(x) = 2x^2(1 - x) + \sigma^2 x^2 + x - 1 + \sigma^2/2$. As a function of x, this is a polynomial in which the leading term, the cubic, has negative coefficient, $2x^3$. So $LV(x)$ is continuous and bounded above (but not below).

2. $LV(x) - \gamma V(x)$ is bounded above by $m := \max\{LV(x)\} - \min\{V(x)\}$. Then $LV(x) \leq V(x) + m$.

3. We have

   $$dY_t = e^{-t}\left[-V(X_t) - \delta + LV(X)_t)\right] \ dt + e^{-t}V'(X_t) \ dB_t$$

   and the drift term is negative.

4. It follows from the previous that $\mathbf{E}^x Y_t \leq Y_0 = V(x)$ and so $\mathbf{E}^x V(X_t) \leq e^t V(x) < \infty$.

5. When x is near 0, we have $LV(x) = -1 + \sigma^2/2 + O(x)$, so negative near 0. For x large, we already established that $LV(x)$ is a third order polynomial in x with a negative cubic term, hence negative for sufficiently large x. It follows that for every $\epsilon > 0$, there exists an interval $[a, b]$ such that $LV(x) < -1 + \sigma^2/2 + \epsilon$ for $x \notin [a, b]$. This interval is then positively recurrent.

Exercise 13.2:

1. We have $dZ_t = U_t\ dt + \sigma\ dB_t - s\ dW_t$, and the performance objective $\frac{1}{2}\mathbf{E}[qZ_t^2 + U_t^2]$. The problem is therefore stationary LQR control of $\{Z_t\}$, i.e., section 13.6.1.

2. The stationary Riccati equation is $-S^2 + q = 0$; compare (13.13), where $a = 0$, $f = 1$. The stabilizing solution is $S = \sqrt{q}$, i.e., the feedback law $U_t = -\sqrt{q} Z_t$. With this, the stationary tracking error has mean 0 and variance $(\sigma^2 + s^2)/2/\sqrt{q}$, and the stationary control signal has mean 0 and variance $sqrtq(\sigma^2 + s^2)/2$. The total cost in stationarity has expectation
$$\frac{1}{2}\mathbf{E}(qZ_t^2 + U_t^2) = \frac{1}{2}(\sigma^2 + s^2)\sqrt{q},$$
in agreement with (13.13).

3. The two-dimensional problem can now be written
$$d\begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} U_t + \begin{bmatrix} \sigma & 0 \\ 0 & s \end{bmatrix} \begin{pmatrix} dB_t \\ dW_t \end{pmatrix}$$
with the performance objective
$$\frac{1}{2}\mathbf{E}\left[(X_t\ Y_t)\begin{bmatrix} 0 & -1/2 \\ -1/2 & 0 \end{bmatrix}\begin{pmatrix} X_t \\ Y_t \end{pmatrix} + U_t^2\right].$$
This problem is in the form of Theorem 13.6.1, with
$$A = 0, \quad F = \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \quad Q = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}, \quad R = \begin{bmatrix} \sigma & 0 \\ 0 & s \end{bmatrix}$$
and "$B_t := (B_t\ W_t)^\top$". However, the system is not stabilizable: The system matrix A is the zero matrix, so both eigenvalues are 0, and the left eigenvector $p = (0\ 1)$ is unaffected by the control ($pF = 0$). Also, the system is not detectable: The right eigenvector $v = (1\ 1)^\top$ is unobservable in the performance criterion ($Qv = 0$).

4. We insert the candidate solution in the algebraic Riccati equation:

$$-q\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}\begin{pmatrix}1\\0\end{pmatrix}(1\ 0)\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} + q\begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix} = 0$$

which we see holds, after doing the matrix multiplication.

The closed-loop system is

$$d\begin{pmatrix}X_t\\Y_t\end{pmatrix} = -q\begin{pmatrix}1\\0\end{pmatrix}(1\ 0)\begin{bmatrix} 1 & -1 \\ -1 & 0 \end{bmatrix}\begin{pmatrix}X_t\\Y_t\end{pmatrix}\,dt + \begin{bmatrix} \sigma & 0 \\ 0 & s \end{bmatrix}\begin{pmatrix}dB_t\\dW_t\end{pmatrix}$$

or

$$d\begin{pmatrix}X_t\\Y_t\end{pmatrix} = q\begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}\begin{pmatrix}X_t\\Y_t\end{pmatrix}\,dt + \begin{bmatrix} \sigma & 0 \\ 0 & s \end{bmatrix}\begin{pmatrix}dB_t\\dW_t\end{pmatrix}$$

and we see that the closed-loop eigenvalues are $-q$ and 0. The eigenvalue $-q$ corresponds to the tracking error, while the eigenvalue 0 corresponds to the reference.