mufit2 model

Peter Komar

June 1, 2019

We introduce a 2-level model, "hidden Gaussian process", for smoothening the inferred growth rate curves of optical density time series produced by a turbidostat.

Contents

1 Data

1.1 Raw optical density data

We start with the time series of the raw optical density (OD) measurements recorded during a single experimental run of a turbidostat. At consecutive (but not necessarily equidistant) time points, OD is recorded. This produces two vectors of real numbers:

- the time points, $\{\mathrm{tp}_n \ : \ n = 1, 2, \ldots N_{\mathrm{total}}\}$, and
- the OD values, $\{\mathrm{od}_n \ : \ n = 1, 2, \ldots N_{\mathrm{total}}\}$.

Under normal operating conditions, the time series (tp, od) has the following features:

- a long initial growth from a low OD value to the operating OD regime,
- sharp drops of OD, when it reaches a predefined threshold value,
- gradual growth of OD between sharps drops, and
- intermittent spikes of OD.

1.2 Preprocessing

Out of the four features of the raw (tp, od) time series, we wish to model only the gradual growth (and maybe the initial growth) phases. For this we filter the time series and partition it into non-overlapping regions by

1. Using heuristic filters to identify sudden changes of od, and remove the associated data points.
2. Group uninterrupted series of data points into non-overlapping regions.
3. Take the logarithm of OD.

This produces the cleaned time series (t, x) of N data points:

$$\begin{aligned} t &= (t_1, t_2, \ldots t_N), \quad \text{where } t_n \in \mathbb{R}, \quad \text{and } t_n < t_{n+1}, \\ x &= (x_1, x_2, \ldots x_N), \quad \text{where } x_n = \log(\mathrm{od}) \in \mathbb{R}, \end{aligned}$$

and a list of R regions, i.e. non-overlapping sets $\mathcal{R}_r$,

$$r \in \{1, 2, \ldots R\}, \qquad \text{where each } \mathcal{R}_r = \{s(r), s(r)+1, \ldots e(r)-1, e(r)\} \subset \{1, 2, \ldots N\}$$

is a list of consecutive indexes, where $s(r)$ is the first and $e(r)$ is the last. The bottom panel of Fig. 1 illustrates how a typical (x, t) time series look like.

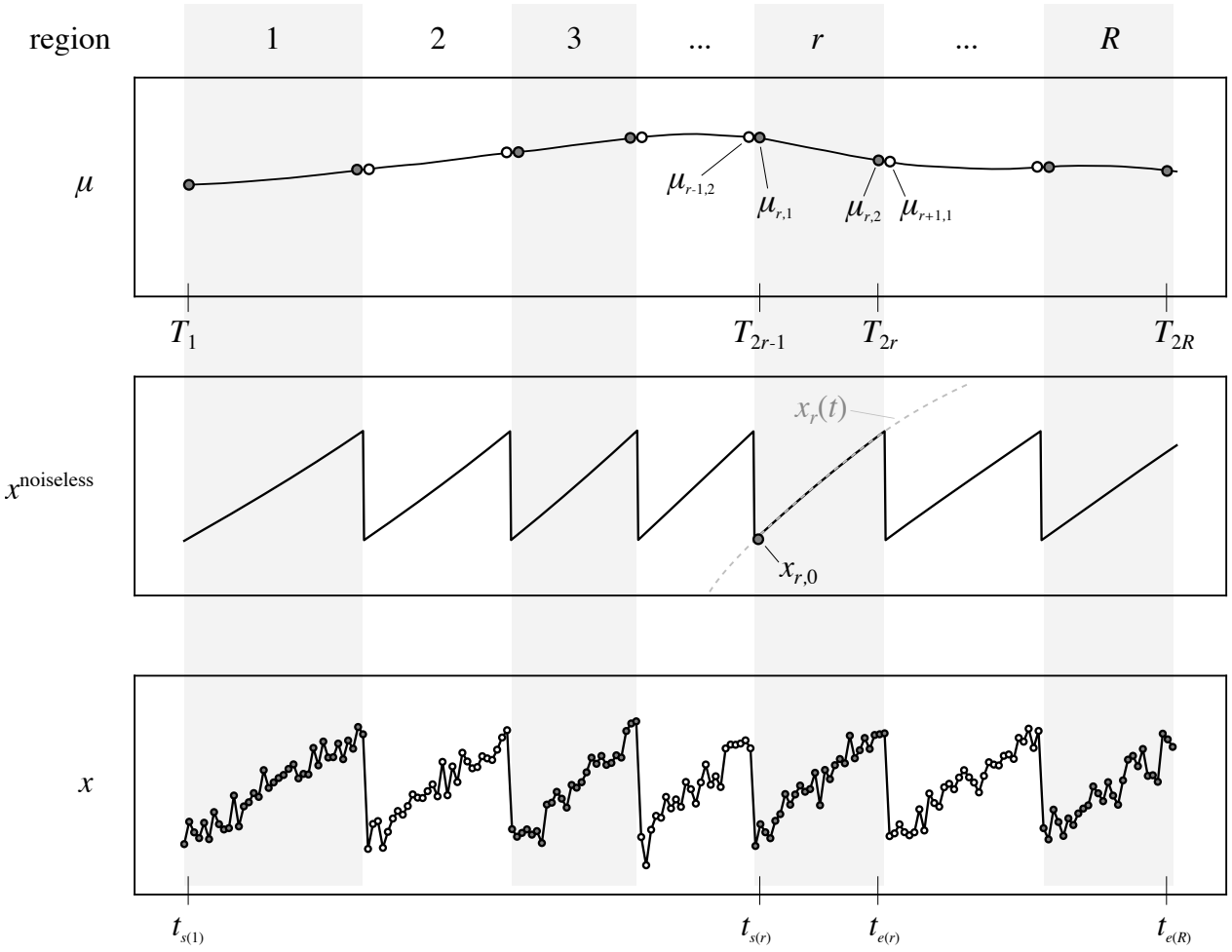


Figure 1: Illustration of the data expected by the model, which is described in the next section. **(Top panel)** The growth rate μ changes gradually with time. **(Middle panel)** The hidden $x^{\text{noiseless}}$ log-OD follows a saw-tooth-like behavior, where regions of gradual growth (between times $t_{s(r)}$ and $t_{e(r)}$ for each region r) are interrupted by sudden drops. **(Bottom panel)**Due to measurement noise, the recorded x log-OD values are distributed around the noiseless log-OD curve, where deviations are uncorrelated.

2 Model

We model all R regions of the cleaned time series (t, x) with a single model that captures the gradual growth within each region as well as two features of between-region change of the growth rate (μ), sudden changes and long-term drifts.

2.1 Parameters

We use the following unknown variables to describe the time series. The role of each value will become more clear in the next subsection.

- $x_0 = \{x_{r,0} \ : \ r = 1, 2, \dots R\} \in \mathbb{R}^R$ represents the starting log-OD value at the beginning of each region.
- $\mu_1 = \{\mu_{r,1} \ : \ r = 1, 2, \dots R\} \in \mathbb{R}^R$ represents the growth rate at the beginning of each region.
- $\mu_2 = \{\mu_{r,2} \ : \ r = 1, 2, \dots R\} \in \mathbb{R}^R$ represents the growth rate at the end of each region.
- $\sigma_x \in \mathbb{R}$ is the strength of measurement noise of log-OD.
- μ_0 is μ at the beginning of the experiment.
- ν_0 is the *rate of change* of μ at the beginning of the experiment
- D is the diffusion coefficient of the velocity of the long term drift of μ.
- τ is the time scale of the short-term fluctuations of μ.
- σ_y is the magnitude of the short-term fluctuations of μ.

2.2 Individual regions

We assume that for given x_0, μ_1, μ_2 values, the log-OD values in different regions become independent. This allows us to write the generating distribution of log-OD as

$$P(x \mid x_0, \mu_1, \mu_2, \sigma_x) = \prod_{r=1}^{R} P(x^{(r)} \mid x_{r,0}, \mu_{r,1}, \mu_{r,2}, \sigma_x) \quad ,$$

where $x^{(r)}$ and $t^{(r)}$ are the set of log-OD and time points belonging to region r.

Furthermore, within each region r, we assume that μ changes deterministically and linearly between $\mu_{r,1}$ (at $t_{s(r)}$) and $\mu_{r,2}$ (at $t_{e(r)}$), i.e.

$$\mu(t) = \mu_{r,1} \frac{t_{e(r)} - t}{t_{e(r)} - t_{s(r)}} + \mu_{r,2} \frac{t - t_{s(r)}}{t_{e(r)} - t_{s(r)}} \quad , \quad \text{if} \quad t_{s(r)} \le t \le t_{e(r)} \quad .$$

This leads to a quadratic time dependence for the "noiseless" log-OD,

$$\begin{aligned}
x_r^{\text{noiseless}}(t) &= x_{r,0} + \int_{t_{s(r)}}^{t} dt' \ \mu(t') = x_{r,0} + f_r(t)\, \mu_{r,1} + g_r(t)\, \mu_{r,2} \quad , \text{ where} \\
f_r(t) &= \frac{t_{e(r)}(t - t_{s(r)}) - \frac{1}{2}(t^2 - t_{s(r)}^2)}{t_{e(r)} - t_{s(r)}} \\
g_r(t) &= \frac{\frac{1}{2}(t^2 - t_{s(r)}^2) - t_{s(r)}(t - t_{s(r)})}{t_{e(r)} - t_{s(r)}}
\end{aligned}$$

The noiseless log-OD deterministic (given x_0, μ_1, μ_2). If we further assume that the measurement noise is uncorrelated between different time points, then each x_n becomes independent from every other $x_{n'}$. Assuming Gaussian noise with strength σ_x, we can write their generating distribution as

$$P(x^{(r)} \mid x_{r,0}, \mu_{r,1}, \mu_{r,2}, \sigma_x) = \prod_{n=s(r)}^{e(r)} \text{Normal}\Big(x_n \;\Big|\; \text{mean} = x_r^{\text{noiseless}}(t_n),\ \text{variance} = \sigma_x^2\Big) \tag{1}$$

2.3 Smooth growth rate

If measurement noise (modeled by σ_x) is small enough, and each region contains enough time points, fitting $x_{r,0}, \mu_{r,1}$ and $\mu_{r,2}$ for each region independently is a viable strategy. Maximizing Eq.(1) can be done in a single step, using the formula of ordinary least square regression. If, however, the measurement noise is considerable and separate fits to individual regions do not provide a robust prediction of μ, then it is worth considering the following model, which assumes that μ is generated by a Gaussian process.

We consider the time series (T, μ), where $T \in \mathbb{R}^{2R}$ is a subset of t, and $\mu \in \mathbb{R}^{2R}$ is the concatenation of consecutive $(\mu_{r,1}, \mu_{r,2})$ vectors, i.e.

$$\begin{aligned} T &= \big(t_{s(1)}, t_{e(1)}, \dots t_{s(r)}, t_{e(r)}, \dots t_{s(R)}, t_{e(R)}\big) = \bigoplus_{r=1}^{R} \big(t_{s(r)}, t_{e(r)}\big) \quad , \\ \mu &= \big(\mu_{1,1}, \mu_{1,2}, \dots \mu_{r,1}, \mu_{r,2}, \dots \mu_{R,1}, \mu_{R,2}\big) = \bigoplus_{r=1}^{R} \big(\mu_{r,1}, \mu_{r,2}\big) \quad . \end{aligned}$$

We model this time series with a Gaussian process, which prescribes that the joint distribution of all components of μ are normally distributed:

$$P(\mu \mid T, \dots) = \text{Multi-Normal}\Big(\mu \;\Big|\; \text{mean} = m(T, \dots),\ \text{covariance} = \Sigma(T, \dots)\Big),$$

where $m \in \mathbb{R}^{2R}$ and $\Sigma \in \mathbb{R}^{2R \times 2R}$ are functions of T and all other model parameters.

We model the expected long-term and short-term behaviors of μ with the sum of two Gaussian processes.

- The long-term drift with gradually changing slope is modeled by the integral of a Brownian motion. In other words we expect the *rate of change* of μ to walk randomly with a fixed diffusion rate. This process has the following mean and covariance

$$\begin{aligned} (m^{\text{int.BM}})_i &= \mu_0 + \nu_0 T_i \\ (\Sigma^{\text{int.BM}})_{i,j} &= D\, \frac{\Big(\min(T_i, T_j)\Big)^2}{2} \left(\max(T_i, T_j) - \frac{\min(T_i, T_j)}{3}\right) \end{aligned}$$

  where μ_0, ν_0 are the value and rate of change of μ at T_1, and D is the diffusion coefficient of the rate of change. (See appendix A.1 for derivation.)

- The short-term changes are modeled by a Gaussian process with squared exponential covariance:

$$\begin{aligned} (m^{\text{sq.exp}})_i &= 0 \\ (\Sigma^{\text{sq.exp}})_{i,j} &= \sigma_\mu^2 \exp\left(-\frac{(T_i - T_j)^2}{2\tau^2}\right) \end{aligned}$$

  where σ_μ is the typical *a priori* fluctuation of μ, and τ is the time scale on which the *a priori* autocorrelation of μ is high.

We assuming independence between the two processes (given their parameter values), their sum results in the Gaussian process

$$\begin{aligned} &P(\mu \mid \mu_0, \nu_0, D, \sigma_\mu, \tau) = \\ &\quad \text{Multi-Normal}\Big(\mu \;\Big|\; \text{mean} = m^{\text{int.BM}}(\mu_0, \nu_0) + m^{\text{sq.exp}},\ \text{covariance} = \Sigma^{\text{int.BM}}(D) + \Sigma^{\text{sq.exp}}(\sigma_\mu, \tau)\Big) \end{aligned} \tag{2}$$

2.4 Prior of x_0

We introduce a Gaussian prior for x_0,

$$P(x_0) = \prod_{r=1}^{R} \text{Normal}\Big(x_{r,0} \;\Big|\; \text{mean} = \bar{x},\ \text{variance} = \lambda^2 (\Delta x)^2\Big),$$

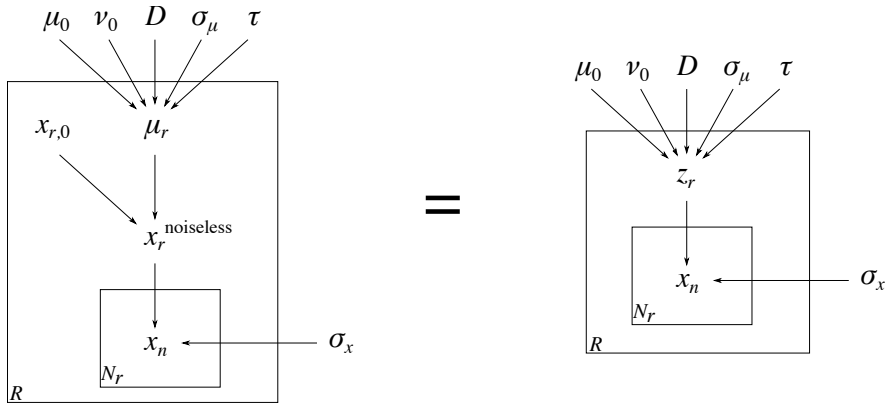
where $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$ is the mean and $(\Delta x)^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$ is the empirical standard deviation of all log-OD values. Constant λ is a large enough spread factor, whose exact value is unimportant as long as $\lambda > 3$. This prior choice does not introduce additional restrictions, but makes the following algebraic and numerical manipulations well-behaving.

2.5 Complete model

Putting the two components of the model (described by Eq.(1) and Eq.(2)) together yields the joint generating distribution

$$P(x, x_0, \mu \mid \mu_0, \nu_0, D, \sigma_\mu, \tau) = P(x_0)\ P\Big(\mu \;\Big|\; T, \mu_0, \nu_0, D, \sigma_\mu, \tau\Big) \prod_{r=1}^{R} P\Big(x^{(r)} \;\Big|\; x_{r,0}, \mu_{r,1}, \mu_{r,2}, \sigma_x\Big) \tag{3}$$

This can be represented graphically as shown on the left side of the figure below (where $N_r = e(r) - s(r) + 1$).



To simplify notation and calculations, we package the elements of x_0, μ_1, μ_2 in a single vector $\tilde{z} \in \mathbb{R}^{3R}$ in an interleaved fashion:

$$z_r = (x_{r,0},\ \mu_{r,1},\ \mu_{r,2}) \quad , \qquad \tilde{z} = \bigoplus_{r=1}^{R} z_r$$

With this notation, we can eliminate $x_r^{\text{noiseless}}$ from the graphical representation (because it is deterministic). This is shown on the right side of the figure, and can be written as

$$P(x, \tilde{z} \mid \mu_0, \nu_0, D, \sigma_\mu, \tau) = P\Big(\tilde{z} \;\Big|\; \mu_0, \nu_0, D, \sigma_\mu, \tau\Big) \prod_{r=1}^{R} P\Big(x^{(r)} \;\Big|\; z_r, \sigma_x\Big) \quad .$$

2.6 Re-ordering $\tilde{z}$

To make calculations in the next section easier, we define a re-ordered version of $\tilde{z}$:

$$z = \left[\bigoplus_{r=1}^{R} x_{r,0}\right] \oplus \left[\bigoplus_{r=1}^{R} (\mu_{r,1}, \mu_{r,2})\right] = S\tilde{z} \quad , \tag{4}$$

where $S = (S_{i,j} \ : \ i,j \in \{0,1,2 \ldots R-1\})$ is the permutation matrix expressed with zero-based indexes

$$S_{i,j} = \begin{cases} [j = 3i] & , \quad \text{if} \quad i \leq R-1 \\ [j = i - R + 1 + \lfloor (i-R)/2 \rfloor] & , \quad \text{if} \quad i \geq R \end{cases}$$

where $[\ldots]$ is one if the statement inside if true and false otherwise, and $\lfloor \ldots \rfloor$ indicates the floor function.

We will use the S matrix to convert between the ordering of $\tilde{z}$ and z. As an example, we expand $S\tilde{z} = z$ for $R = 4$ explicitly:

$$S\tilde{z} \quad = \quad \left[\begin{array}{c|c|c|c} 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \\ \hline 1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \\ \hline & & 1 & \\ & & & 1 \\ & & & \\ & & & \\ \hline & & & 1 \\ & & & \\ & & & \\ & & & \end{array}\right] \left[\begin{array}{c} x_{1,0} \\ \mu_{1,1} \\ \mu_{1,2} \\ \hline x_{2,0} \\ \mu_{2,1} \\ \mu_{2,2} \\ \hline x_{3,0} \\ \mu_{3,1} \\ \mu_{3,2} \\ \hline x_{4,0} \\ \mu_{4,1} \\ \mu_{4,2} \end{array}\right] \quad = \quad \left[\begin{array}{c} x_{1,0} \\ x_{2,0} \\ x_{3,0} \\ x_{4,0} \\ \hline \mu_{1,1} \\ \mu_{1,2} \\ \mu_{2,1} \\ \mu_{2,2} \\ \mu_{3,1} \\ \mu_{3,2} \\ \mu_{4,1} \\ \mu_{4,2} \end{array}\right] \quad = \quad z$$

3 Type II maximum likelihood solution

We wish to obtain robust estimates of μ, and its statistical uncertainty. To do this, we follow the "type II" maximum likelihood procedure:

1. We determine the maximum likelihood estimates of the parameters $\mu_0, \nu_0, D, \sigma_\mu, \tau$ and σ_x. (For this, we need a computationally efficient access to the marginal likelihood $P(x \mid \mu_0, \nu_0, D, \sigma_\mu, \tau, \sigma_x)$.)

2. Using the MLE parameter values, we calculate the mean and standard deviation of each element of μ.

In this section, we derive a formula for the logarithm of the marginal likelihood, which can directly implemented using matlab or python's numpy.

3.1 Expanding Eq.(1)

To better understand how data from each region $x^{(r)}$ determines the region-specific parameters $z_r = (x_{r,0}, \mu_{r,1}, \mu_{r,2})$, we expand Eq.(1).

$$P(x^{(r)} \mid z_r, \sigma_x) = \prod_{n=s(r)}^{e(r)} \text{Normal}\Big(x_n \ \Big| \ \text{mean} = x_{r,0} + f_r(t_n)\mu_{r,1} + g_r(t_n)\mu_{r,2}, \ \text{variance} = \sigma_x^2\Big)$$

Now, we can expand pdf of the normal distribution and write

$$P(x^{(r)} \mid z_r, \sigma_x) = \exp\left(-\frac{N_r}{2}\log(2\pi\sigma_x^2)\right) \ \exp\left(-\frac{1}{2\sigma_x^2}\sum_{n=s(r)}^{e(r)}\Big[x_{r,0} + f_r(t_n)\mu_{r,1} + g_r(t_n)\mu_{r,2} - x_n\Big]^2\right)$$

where $N_r = e(r) - s(r) + 1$. We expand the square bracket and distribute the summation on the terms. The result is a quadratic form of the vector z_r,

$$
\begin{aligned}
-\frac{1}{2}\sum_{n=s(r)}^{e(r)}\Big[\dots\Big]^2 &= \gamma_r + b_r^\top z_r - \frac{1}{2} z_r^\top A_r z_r, \qquad \text{where} \\
\gamma_r &= -\frac{1}{2}\sum_{n=s(r)}^{e(r)} (x_n)^2 \\
b_r &= \sum_{n=s(r)}^{e(r)} \begin{bmatrix} x_n \\ x_n f_r(t_n) \\ x_n g_r(t_n) \end{bmatrix} \\
A_r &= \sum_{n=s(r)}^{e(r)} \begin{bmatrix} 1 & f_r(t_n) & g_r(t_n) \\ f_r(t_n) & (f_r(t_n))^2 & f_r(t_n) g_r(t_n) \\ g_r(t_n) & f_r(t_n) g_r(t_n) & (g_r(t_n))^2 \end{bmatrix}
\end{aligned}
$$

When we consider all regions, their generating distribution can be written as

$$
\begin{aligned}
P(x \mid z, \sigma_x) &= \prod_{r=1}^{R} P(x^{(r)} \mid z_r, \sigma_x) \\
&= \exp\left(-\frac{N}{2}\log(2\pi\sigma_x^2)\right) \exp\left(\frac{1}{\sigma_x^2}\sum_{r=1}^{R}\Big[\gamma_r + b_r^\top z_r - \frac{1}{2} z_r^\top A_r z_r\Big]\right) \\
&= \exp\left(-\frac{N}{2}\log(2\pi\sigma_x^2)\right) \exp\left(\frac{1}{\sigma_x^2}\Big[\gamma + b^\top \tilde{z} - \frac{1}{2}\tilde{z}^\top A \tilde{z}\Big]\right) \qquad (5)
\end{aligned}
$$

where the new variables γ, b, A are (direct) sums of the individual γ_r, b_r, A_r variables:

$$
\gamma = \sum_{r=1}^{R} \gamma_r, \qquad b = \bigoplus_{r=1}^{R} b_r, \qquad A = \bigoplus_{r=1}^{R} A_r,
$$

where we define the direct sum of the matrices as the operation of concatenating them in a block-diagonal fashion. (Note: Since γ, b, A do not depend on model parameters, we can compute them once and store their values to improve efficiency.)

3.2 Expanding Eq.(2)

Now, we expand the formula describing how the hyperparameters $\mu_0, \nu_0, D, \sigma_\mu, \tau$ affect the region-specific parameters $z_r = (x_{r,0}, \mu_{r,1}, \mu_{r,2})$. While Eq.(2) describes the joint distribution of all elements of μ_1 and μ_2, here we incorporate x_0 into the formula, and express the joint distribution of all elements of z ($= S\tilde{z}$).

$$
\begin{aligned}
\mu = \bigoplus_{r=1}^{R}(\mu_{r,1}, \mu_{r,2}) &\sim \text{Multi-Normal}\Big(\mu \;\Big|\; \text{mean} = \underbrace{m^{\text{int.BM}} + m^{\text{sq.exp}}}_{m^{(\mu)}},\ \text{cov} = \underbrace{\Sigma^{\text{int.BM}} + \Sigma^{\text{sq.exp}}}_{\Sigma^{(\mu)}}\Big) \\
x_0 = \bigoplus_{r=1}^{R} x_{r,0} &\sim \text{Multi-Normal}\Big(x_0 \;\Big|\; \text{mean} = \underbrace{\mathbf{1}_R \tilde{x}}_{m^{(x_0)}},\ \text{cov} = \underbrace{\mathbb{I}_{R\times R}\lambda^2(\Delta x)^2}_{\Sigma^{(x_0)}}\Big) \\
z = x_0 \oplus \mu &\sim \text{Multi-Normal}\Big(z \;\Big|\; \text{mean} = m,\ \text{cov} = \Sigma\Big) \\
&= \exp\left(-\frac{1}{2}\log\Big(\det(2\pi\Sigma)\Big)\right) \exp\left(-\frac{1}{2}(z-m)^\top \Sigma^{-1}(z-m)\right) \qquad (6)
\end{aligned}
$$

where $\mathbf{1}_R$ is the all-1 vector, and $\mathbb{I}_{R\times R}$ is the identity matrix, and the mean and covariance of z can be written as

$$m = m^{(x_0)} \oplus m^{(\mu)}, \qquad \Sigma = \Sigma^{(x_0)} \oplus \Sigma^{(\mu)}$$

where the direct sum ($\oplus$) is defined as concatenation between vectors and block-diagonal composition between matrices. Separating the determinant and the inverse operations on the x_0 and the μ spaces,

$$\begin{aligned} \Sigma^{-1} &= \left(\Sigma^{(x_0)}\right)^{-1} \oplus \left(\Sigma^{(\mu)}\right)^{-1} \quad , \\ \log\Big(\det(2\pi\Sigma)\Big) &= 3R\log(2\pi) + \log\Big(\det(\Sigma^{(x_0)})\Big) + \log\Big(\det(\Sigma^{(\mu)})\Big) \quad , \end{aligned}$$

will lead more efficient calculations because $m^{(x_0)}$ and $\Sigma^{(x_0)}$ are known but $m^{(\mu)}$ and $\Sigma^{(\mu)}$ are unknown, and need to be evaluated at every iteration step during fitting.

3.3 Eliminating z

Multiplying Eq.(6) and Eq.(5), and converting $\tilde{z}$ to z using the definition of the permutation matrix S from Eq.(4) ($\tilde{z} = S^\top z$) yields the likelihood in the following quadratic exponential form

$$\begin{aligned} P(x \mid z)\, P(z) &= \exp\left(-\frac{N}{2}\log(2\pi\sigma_x^2)\right) \exp\left(\frac{1}{\sigma_x^2}\Big[\gamma + (Sb)^\top z - \frac{1}{2} z^\top (SAS^\top) z\Big]\right) \times \\ &\quad \exp\left(-\frac{1}{2}\log\Big(\det(2\pi\Sigma)\Big)\right) \exp\left(-\frac{1}{2}(z-m)^\top \Sigma^{-1}(z-m)\right) \\ &= \exp\left(-\frac{1}{2} z^\top \mathcal{A} z + \mathcal{B}^\top z + \mathcal{C}\right) \end{aligned}$$

where

$$\begin{aligned} \mathcal{A} &= \frac{1}{\sigma_x^2} SAS^\top + \Sigma^{-1} \\ \mathcal{B} &= \frac{1}{\sigma_x^2} Sb + \Sigma^{-1} m \\ \mathcal{C} &= -\frac{N}{2}\log(2\pi\sigma_x^2) - \frac{1}{2}\log\Big(\det(2\pi\Sigma)\Big) + \frac{\gamma}{\sigma_x^2} - \frac{1}{2} m^\top \Sigma^{-1} m \quad . \end{aligned}$$

To eliminate z from the likelihood, we integrate with respect to z. Using the result for such a Gaussian integral (see Appendix A.2), we obtain

$$P(x) = \int dz\; P(x \mid z) P(z) = \int dz\; \exp\left(-\frac{1}{2} z^\top \mathcal{A} z + \mathcal{B}^\top z + \mathcal{C}\right) = \sqrt{\frac{(2\pi)^{3R}}{\det(\mathcal{A})}} \exp\left(\frac{1}{2}\mathcal{B}^\top \mathcal{A}^{-1}\mathcal{B} + \mathcal{C}\right) \quad .$$

Taking the logarithm yields the log likelihood as a function of the data x and the hyperparameters μ_0, ν_0, D, σ_μ, τ, σ_x,

$$L\big(\mu_0, \nu_0, D, \sigma_\mu, \tau, \sigma_x\big) = \log P(x) = \frac{3R}{2}\log(2\pi) - \frac{1}{2}\log\Big(\det(\mathcal{A})\Big) + \frac{1}{2}\mathcal{B}^\top \mathcal{A}^{-1}\mathcal{B} + \mathcal{C} \quad . \tag{7}$$

We can start from realistic values for the hyperparameters, use gradient-based optimization methods to find maximum of L. This yields the maximum likelihood estimates μ_0^*, ν_0^*, D^*, σ_μ^*, τ^*, σ_x^*.

3.4 Posterior mean and variance of μ

We can use the maximum likelihood estimates of μ_0^*, ν_0^*, D^*, σ_μ^*, τ^*, σ_x^* to calculate the mean and variance of the growth rate from the joint distribution

$$P(x, z) = P(x \mid z) P(z) = \exp\left(-\frac{1}{2} z^\top \mathcal{A} z + \mathcal{B}^\top z + \mathcal{C}\right) \quad ,$$

which, after dividing it with $P(x)$ (which is just a normalization constant from the point of view of z) yields the conditional probability

$$\begin{aligned} P(z \mid x) &= \frac{P(x,z)}{P(x)} \propto \exp\left(-\frac{1}{2} z^\top \mathcal{A} z + \mathcal{B}^\top z\right) \propto \exp\left(-\frac{1}{2}\Big(z - \mathcal{A}^{-1}\mathcal{B}\Big)^\top \mathcal{A}\Big(z - \mathcal{A}^{-1}\mathcal{B}\Big)\right) \\ &= \text{Multi-Normal}\Big(z \;\Big|\; \text{mean} = \mathcal{A}^{-1}\mathcal{B},\ \text{cov} = \mathcal{A}^{-1}\Big) \quad , \end{aligned}$$

which means that each component of z ($x_{r,0}$ and $\mu_{r,1}, \mu_{r,2}$) is distributed as a normal distribution. Remembering that $z = x_0 \oplus \mu$, we can express the mean and variance of each element of $\mu \in \mathbb{R}^{2R}$ as

$$\mathbb{E}(\mu_i \mid x, \mu_0^*, \nu_0^*, D^*, \sigma_\mu^*, \tau^*, \sigma_x^*) = \Big[\mathcal{A}^{-1}\mathcal{B}\Big]_{R+i} \quad , \qquad \text{var}(\mu_i \mid x, \mu_0^*, \nu_0^*, D^*, \sigma_\mu^*, \tau^*, \sigma_x^*) = \Big[\mathcal{A}^{-1}\Big]_{R+i,\,R+i} \quad . \tag{8}$$

A Derivations

A.1 Mean and covariance of integral of Brownian motion

Shifted Brownian motion is defined as a Gaussian process $\nu(t)$ on $t \geq 0$, with mean and covariance functions defined as

$$m^{(\nu)}(t) = \nu_0 \quad , \qquad K^{(\nu)}(t,t') = D\min(t,t') \quad ,$$

where ν_0 is the starting point at $t = 0$, and D is the diffusion coefficient of the Brownian motion.

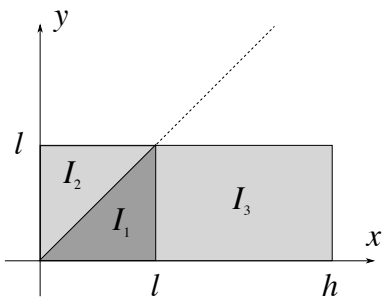
We define the integral of the Brownian motion as the process $\mu(t) = \mu_0 + \int_0^t dx\ \nu(x)$ on $t \geq 0$, with mean and covariance functions defined as

$$m^{(\mu)}(t) = \mu_0 + \int_0^t dx\ m^{(\nu)}(x) = \mu_0 + \nu_0 t \quad , \qquad K^{(\mu)}(t,t') = \int_0^t dx \int_0^{t'} dy\ K^{(\nu)}(x,y)$$

The double integral can be separated into 3 non-overlapping parts. With the notation $l = \min(t,t')$ and $h = \max(t,t')$, we can write

$$K^{(\mu)}(t,t') = D \int_0^h dx \int_0^l dy\ \min(x,y)$$

since $K^{(\nu)}(x,y) = D\min(x,y)$ is symmetric in its arguments. The integration domain can be partitioned into three non-overlapping domains,



$$
\begin{aligned}
I_1 &= \int_0^l dx \int_0^x dy \underbrace{\min(x,y)}_{y} = \int_0^l dx \frac{x^2}{2} = \frac{l^3}{6} \\
I_2 &= \int_0^l dy \int_0^y dx \underbrace{\min(x,y)}_{x} = \int_0^l dy \frac{y^2}{2} = \frac{l^3}{6} \\
I_3 &= \int_l^h dx \int_0^l dy \underbrace{\min(x,y)}_{y} = \int_l^h dx \frac{l^2}{2} = (h-l)\frac{l^2}{2}
\end{aligned}
$$

the sum of which yields

$$
K^{(\mu)}(t,t') = D(I_1+I_2+I_3) = D\frac{l^2}{2}\left(h-\frac{l}{3}\right) = D\frac{\min(t,t')^2}{2}\left(\max(t,t')-\frac{\min(t,t')}{3}\right)
$$

A.2 Gaussian integral

Here we calculate the following multi-dimensional Gaussian integral, where $z, \mathcal{B} \in \mathbb{R}^n$, $\mathcal{A} \in \mathbb{R}^{n\times n}$ (positive definite) and $\mathcal{C} \in \mathbb{R}$,

$$
\begin{aligned}
I &:= \int dz\ \exp\left(-\frac{1}{2}z^\top \mathcal{A} z + \mathcal{B}^\top z + \mathcal{C}\right) \\
&= \int dz\ \exp\left(-\frac{1}{2}z^\top \mathcal{A} z + \frac{1}{2}z^\top \mathcal{B} + \frac{1}{2}\mathcal{B}^\top z + \mathcal{C}\right) \\
&= \int dz\ \exp\left(-\frac{1}{2}\Big(z-\mathcal{A}^{-1}\mathcal{B}\Big)^\top \mathcal{A}\Big(z-\mathcal{A}^{-1}\mathcal{B}\Big) + \frac{1}{2}\mathcal{B}^\top\mathcal{A}^{-1}\mathcal{B} + \mathcal{C}\right) \\
&= \exp\left(\frac{1}{2}\mathcal{B}^\top\mathcal{A}^{-1}\mathcal{B} + \mathcal{C}\right)\int dz'\ \exp\left(-\frac{1}{2}(z')^\top \mathcal{A} z'\right).
\end{aligned}
$$

Since $\mathcal{A}$ is positive definite there exists an orthogonal matrix $\mathcal{O}$ that diagonalizes it,

$$
\mathcal{O}\mathcal{A}\mathcal{O}^\top = \Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \dots \lambda_n)
$$

where λ_i are the eigenvalues of $\mathcal{A}$. In terms of the transformed coordinates $z'' = \mathcal{O}z'$ we get a separable integral:

$$
\int dz'\ \exp\left(-\frac{1}{2}(z')^\top \mathcal{A} z'\right)\Big) = \int dz''\ \exp\left(-\frac{1}{2}(z'')^\top \Lambda z''\right)\Big) = \prod_{i=1}^n \int dx_i\ \exp\left(-\frac{1}{2}x_i\lambda_i x_i\right)\Big)
$$

Each factor can be simplified using the result for a 1-dimensional Gaussian integral $\int dx\ \exp(-\frac{1}{2}ax^2) = \sqrt{2\pi/a}$, giving

$$
\prod_{i=1}^n \sqrt{2\pi/\lambda_i} = \sqrt{\frac{(2\pi)^n}{\prod_i \lambda_i}} = \sqrt{\frac{(2\pi)^n}{\det(\Lambda)}} = \sqrt{\frac{(2\pi)^n}{\det(\mathcal{A})}},
$$

where $\det(\Lambda) = \det(\mathcal{A})$, because the transformation matrix $\mathcal{O}$ is orthogonal, i.e. $\det(\mathcal{O}) = 1$. This results in

$$
I = \exp\left(\frac{1}{2}\mathcal{B}^\top\mathcal{A}^{-1}\mathcal{B} + \mathcal{C}\right) \times \sqrt{\frac{(2\pi)^n}{\det(\mathcal{A})}} \quad .
$$

B Implementation

B.1 Stan code for Eq.(7)