

# Отчет?

Климов Иван

В дальнейших рассуждениях будем считать число популяций равное единице.  
Функция, которую мы хотим оптимизировать:

$$h(\Theta|S) = \prod_{d=0 \dots n} \frac{e^{-M[d]} M[d]^{S[d]}}{S[d]!}$$

Посмотрим что вообще хочется знать в общем случае:

$$\frac{dh}{d\theta} = \frac{\partial h}{\partial \phi} \frac{\partial \phi}{\partial \theta} + \frac{\partial h}{\partial \theta}$$

Заметим, что в  $h$  не используется  $\theta$  в явном виде. Поэтому  $\frac{\partial h}{\partial \theta} = 0$ . Попробуем избежать подсчета  $\frac{\partial \phi}{\partial \theta}$  при помощи Adjoint-State Method. Пусть:

$$\mathcal{L}(\phi(\theta), \theta) = h(\phi(\theta), \theta) + \lambda \cdot F(\phi(\theta), \theta)$$

Что такое  $F$ ? Разумеется условие на то, что  $\phi(\theta)$  является ровно корнем нашего дифференциального уравнения. Продифференцируем Лагранжиан:

$$\frac{d\mathcal{L}}{d\theta} = \frac{\partial h}{\partial \phi} \frac{\partial \phi}{\partial \theta} + \lambda \cdot \left( \frac{\partial F}{\partial \phi} \frac{\partial \phi}{\partial \theta} + \frac{\partial F}{\partial \theta} \right) = \left( \frac{\partial h}{\partial \phi} + \lambda \frac{\partial F}{\partial \phi} \right) \cdot \frac{\partial \phi}{\partial \theta} + \lambda \frac{\partial F}{\partial \theta}$$

Распишем размерности функций чтобы понимать что вообще происходит (в одномерном случае):

- $h$  выдает в качестве результата конкретный likelihood, поэтому она действует в  $\mathbb{R}$
- $\theta$  можно воспринимать как "константную" функцию, она всегда выдает  $\mathbb{R}^M$ , где  $M$  — количество параметров.
- $\phi$  выдает в качестве результата все ответы на константной сетке, то есть вектор  $\mathbb{R}^{N^P}$  или  $\mathbb{R}^N$  в случае одной популяции, где  $N$  — количество точек на сетке.
- $F$  выдает то же самое что и  $\phi$ , поэтому результат это вектор  $\mathbb{R}^{N^P}$  или  $\mathbb{R}^N$  в случае одной популяции.

Тогда можем сделать следующие выводы по размерностям производных:

- $\frac{dh}{d\theta}$  размерность  $\mathbb{R}^{1 \times M}$
- $\frac{\partial h}{\partial \phi}$  размерность  $\mathbb{R}^{1 \times N}$
- $\frac{\partial F}{\partial \phi}$  размерность  $\mathbb{R}^{N \times N}$
- $\frac{\partial \phi}{\partial \theta}$  размерность  $\mathbb{R}^{N \times M}$
- $\frac{\partial F}{\partial \theta}$  размерность  $\mathbb{R}^{N \times M}$

Легко видеть что тогда для получения корректного результата  $\lambda$  должна иметь размерность  $\mathbb{R}^{1 \times N}$ . Давайте выберем  $\lambda$ , так чтобы множитель занулился и нам не пришлось бы считать  $\frac{\partial \phi}{\partial \theta}$ . Как именно его выбрать? Пусть  $A := \frac{\partial F}{\partial \phi}$ ,  $b = \frac{\partial h}{\partial \phi}$ . Тогда:

$$\lambda A = -b \implies A^T \lambda^T = -b^T$$

Таким образом наш вектор  $\lambda$  это просто решение понятной системы уравнений. Теперь нам нужно посчитать три производные:

- $\boxed{\frac{\partial h}{\partial \phi}}$

Считать градиент от  $n+1$  произведения не очень хочется, так как производная произведения это не произведение производных и там все очень страшно разрастается.

Тогда можем заметить, что у функции и логарифма одинаковые промежутки монотонности, следовательно можно найти градиент логарифма и это будет то что нам нужно. Не забываем о том, что весь аллель-частотный спектр нормируется. Обозначим:  $\mathcal{A} = \sum_{i=0}^N S[i]$ ,  $\mathcal{B} = \sum_{i=0}^N M[i]$ . Тогда:

$$\begin{aligned} \frac{d \log h}{d\phi} &= \sum_{i=0}^N -\mathcal{A} \cdot \frac{d}{d\phi} \frac{M[i]}{\mathcal{B}} + S[i] \cdot \frac{d \log M[i]}{d\phi} - S[i] \cdot \frac{d \log \mathcal{B}}{d\phi} = \\ &= \sum_{i=0}^N -\frac{\mathcal{A}}{\mathcal{B}^2} \left( \mathcal{B} \frac{dM[i]}{d\phi} - M[i] \frac{d\mathcal{B}}{d\phi} \right) + \frac{S[i]}{M[i]} \frac{dM[i]}{d\phi} - S[i] \frac{d\mathcal{B}}{d\phi} \end{aligned}$$

Свели задачу к нахождению  $\frac{dM[i]}{d\phi}$  при фиксированном  $i$ . Следующим шагом проталкиваем градиент внутрь  $M$ :

$$\frac{dM[i]}{d\phi} = \int_0^1 \binom{n}{i} x^i (1-x)^{n-i} dx$$

Тогда значение АЧС в клетке можно вычислить приближенно через метод трапеций:

$$\frac{dM[i]}{d\phi} = \binom{n}{i} \sum_{k=0}^{N-1} \frac{x_{k+1} - x_k}{2} \cdot (x_k^i (1-x_k)^{n-i} + x_{k+1}^i (1-x_{k+1})^{n-i})$$

То есть эта часть решения свелась к нахождению какой-то фиксированной константы, равной численному значению такого интеграла.

- $\boxed{\frac{\partial F}{\partial \phi}}$  Посмотрим на определение  $F$  из основной статьи:

$$F = \frac{1}{2} \sum_{i=1}^P \frac{\partial^2}{\partial^2 x_i} \frac{x_i(1-x_i)}{\nu_i} \phi - \sum_{i=1}^P \frac{\partial}{\partial x_i} \left( \gamma_i x_i (1-x_i) + \sum_{j=1}^P M_{i \leftarrow j} (x_j - x_i) \right) \phi - \frac{\partial}{\partial \tau} \phi$$

Тогда, сразу воспользовавшись тем что популяция у нас пока одна:

$$\frac{\partial F}{\partial \phi} = -\frac{1}{\nu} - \gamma(1-2x)$$

- $\boxed{\frac{\partial F}{\partial \theta}}$

Если рассмотреть численную схему вычисления  $\varphi$ , то получим следующее тождество:

$$A\varphi^{t+1} = \varphi^t + C$$

Где  $C$  — какая-то константа домноженная на базисный вектор, отвечающая за мутации,  $A$  — наша трехдиагональная матрица перехода, а  $\varphi$  — это вектора, отвечающие за значения функции на сетке в моментах времени  $t+1$  и  $t$ . Распишем подробнее что же такое  $A$  и  $C$  в случае одной популяции:

—

$$C = \frac{\Delta t}{x_1} \cdot \frac{\theta}{x_2 - x_0} \cdot \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

— Для подсчета матрицы  $A$  посмотрим на вывод трехдиагональной системы:

$$\begin{aligned} \frac{\phi_j^{t+1} - \phi_j^t}{\Delta t} &= \frac{1}{\Delta_j} \left( F_{j+1/2}^{t+1} - F_{j-1/2}^{t+1} \right) \\ \phi_j^{t+1} - \frac{\Delta t}{\Delta_j} \left( F_{j+1/2}^{t+1} - F_{j-1/2}^{t+1} \right) &= \phi_j^t \end{aligned}$$

При этом:

$$F_{j+1/2}^{t+1} = \frac{1}{x_{j+1} - x_j} \frac{1}{2\nu} (x_{j+1}(1 - x_{j+1})\phi_{j+1}^{t+1} - x_j(1 - x_j)\phi_j^{t+1}) - \frac{\gamma}{2} \frac{(x_j + x_{j+1})}{2} \left(1 - \frac{x_j + x_{j+1}}{2}\right) (\phi_j^{t+1} + \phi_{j+1}^{t+1})$$

$$F_{j-1/2}^{t+1} = \frac{1}{x_j - x_{j-1}} \frac{1}{2\nu} (x_j(1 - x_j)\phi_j^{t+1} - x_{j-1}(1 - x_{j-1})\phi_{j-1}^{t+1}) - \frac{\gamma}{2} \frac{(x_{j-1} + x_j)}{2} \left(1 - \frac{x_{j-1} + x_j}{2}\right) (\phi_{j-1}^{t+1} + \phi_j^{t+1})$$

Распишем их разницу в удобном формате:

$$\begin{aligned} F_{j+1/2}^{t+1} - F_{j-1/2}^{t+1} &= \phi_{j-1}^{t+1} \left( \frac{x_{j-1}(1 - x_{j-1})}{2\nu(x_j - x_{j-1})} + \frac{\gamma}{2} \frac{x_{j-1} + x_j}{2} \left(1 - \frac{x_{j-1} + x_j}{2}\right) \right) \\ &\quad + \phi_j^{t+1} \left( -\frac{x_j(1 - x_j)}{2\nu} \left( \frac{1}{x_{j+1} - x_j} + \frac{1}{x_j - x_{j-1}} \right) \right) \\ &\quad + \phi_j^{t+1} \left( \frac{\gamma}{2} \frac{x_{j-1} + x_j}{2} \left(1 - \frac{x_{j-1} + x_j}{2}\right) - \frac{\gamma}{2} \frac{x_j + x_{j+1}}{2} \left(1 - \frac{x_j + x_{j+1}}{2}\right) \right) \\ &\quad + \phi_{j+1}^{t+1} \left( \frac{x_{j+1}(1 - x_{j+1})}{2\nu(x_{j+1} - x_j)} - \frac{\gamma}{2} \frac{x_j + x_{j+1}}{2} \left(1 - \frac{x_j + x_{j+1}}{2}\right) \right) \end{aligned}$$

Тогда, домножив это на  $-\frac{\Delta t}{\Delta_j}$  и добавив 1 к коэффициенту при  $\phi_j^{t+1}$  мы получаем следующие коэффициенты в трехдиагональной матрице:

$$\begin{aligned} [a_j] &= -\frac{\Delta t}{\Delta_j} \left( \frac{x_{j-1}(1 - x_{j-1})}{2\nu(x_j - x_{j-1})} + \frac{\gamma}{2} \frac{x_{j-1} + x_j}{2} \left(1 - \frac{x_{j-1} + x_j}{2}\right) \right) \\ [b_j] &= 1 + \frac{\Delta t}{\Delta_j} \left( \frac{x_j(1 - x_j)}{2\nu} \frac{\Delta_j}{(x_{j+1} - x_j)(x_j - x_{j-1})} - \frac{\gamma}{2} \frac{x_{j-1} + x_j}{2} \left(1 - \frac{x_{j-1} + x_j}{2}\right) + \frac{\gamma}{2} \frac{x_j + x_{j+1}}{2} \left(1 - \frac{x_j + x_{j+1}}{2}\right) \right) \\ [c_j] &= \frac{\Delta t}{\Delta_j} \left( \frac{\gamma}{2} \frac{x_j + x_{j+1}}{2} \left(1 - \frac{x_j + x_{j+1}}{2}\right) - \frac{x_{j+1}(1 - x_{j+1})}{2\nu(x_{j+1} - x_j)} \right) \end{aligned}$$

Производные коэффициентов по  $\theta$  очевидны. Все формулы совпадают с исходным кодом *dad*.

В граничных точках получается следующая ситуация:  $a_0 = c_G = 0$ . При это  $c_0$  и  $a_G$  совпадают с формулами выше. Осталось посмотреть на код и понять что происходит с коэффициентами  $b$ :

$$\begin{aligned} b_0 &= 1 + \frac{\Delta t}{\Delta_j} \left( \frac{x_0(1 - x_0)}{2\nu(x_1 - x_0)} + \frac{\gamma}{2} \frac{x_0 + x_1}{2} \left(1 - \frac{x_0 + x_1}{2}\right) \right) \\ b_G &= 1 + \frac{\Delta t}{\Delta_j} \left( \frac{x_G(1 - x_G)}{2\nu(x_G - x_{G-1})} - \frac{\gamma}{2} \frac{x_{G-1} + x_G}{2} \left(1 - \frac{x_{G-1} + x_G}{2}\right) \right) + \frac{\Delta t}{\Delta_j} \left( \frac{1}{\nu} + \gamma \cdot x_G \cdot (1 - x_G) \right) \end{aligned}$$

Производная будет считаться следующим образом:

$$\frac{d\phi^{t+1}}{d\theta} = \frac{dA^{-1}}{d\theta} (\phi^t + C) + A^{-1} \left( \frac{d\phi^t}{d\theta} + \frac{dC}{d\theta} \right)$$