

Notas de implementación para el Gillespie con estructura de edad

September 4, 2023

1 Consideraciones generales

1.1 Directiva de compilación

Se crea el ejecutable con una directiva como la que sigue:

```
gcc Main.c Read.c PDESolver.c Utilities.c Gillespie.c -o EXE
```

Flags:

- Si se compila en Linux, añadir

`-lm`

- Para cuestiones de rendimiento, añadir

`-O2`

(el uso de

`-O3`

parece arriesgado).

1.2 Unidades

Por ahora, en los ficheros de entrada se entiende que las magnitudes están dadas en unidades del sistema internacional (metros y segundos y, en principio, moles). Internamente se trabaja de forma adimensional. Las salidas que dependen de espacio están dadas de forma adimensional. Otras salidas, según se especifique.

1.3 Ecuaciones

La ecuación para el oxígeno que trabajamos en la actualidad es

$$\frac{\partial c}{\partial t} = D_c \frac{\partial^2 c}{\partial x^2} - k_1 n(t, x) c - k_2 c + S, \quad (1.1)$$

donde entendimos que la n es una densidad de individuos. Se resuelve en un intervalo espacial, puede tener condiciones de flujo cero o entrante. Podría no tener un término fuente extra en algunas aplicaciones. La ratio de proliferación coarse-grained $\lambda = \lambda_n(c)$ (*que en esta rutina se usa para determinar la distribución de edades en equilibrio*) se calcula mediante

$$\frac{2F_S}{\tau_p} \frac{e^{-(\lambda+\nu)a_{G1/S}}}{\lambda + \nu + 1/\tau_p} = 1. \quad (1.2)$$

1.4 Formato de entrada

La simulación se ejecuta con una directiva de tipo

```
./EXE host_necrotic oxy_necrotic pars_pop pars_sim a
```

Por orden,

1. El primer argumento es el nombre del ejecutable.
2. El segundo y el tercero son las distribuciones iniciales de población y oxígeno.
3. El cuarto es un fichero de parámetros relativos a características de la población.
4. El quinto es un fichero de parámetros que indica información sobre la simulación.
5. El sexto es un literal de caracteres que identifica a los ficheros y carpetas de salida.

Formato para las distribuciones Ambos son ficheros dados por una única columna de valores. Leídos de arriba a abajo, se entiende que son los valores sobre una malla de paso uniforme (cuyo valor se especifica en otro fichero de entrada, ver más abajo), especificados de izquierda a derecha. Cada valor va separado del siguiente por un salto de línea. No se incluye un salto de línea tras el último valor (cuidado porque si se hace se procesa mal el tamaño de la malla espacial). El número de slots de la malla espacial se deduce de forma indirecta a partir del número de líneas de estos ficheros.

El fichero de entrada para la población se interpreta como una cantidad de partículas (en cada localización espacial -slots de tamaño Δx , que *no tiene por qué* ser exactamente 1 metro). El fichero de entrada para el oxígeno, de forma similar con valores de concentración, dados en el SI.

Formato para el fichero de parámetros de la población Se incluyen una serie de números en la misma línea, separados por espacios. No hay ningún caracter tras el último valor numérico. Por orden, se trata de:

1. The death rate, ν .
2. The proliferation-related parameter, τ_p .
3. The value of a_+ , proliferation rate at zero oxygen. *De momento no vamos a considerar el régimen de $p6/p3$ para el cuál esta cantidad es relevante. Así que podríamos suprimirla.*
4. The value of $p6/p3$
5. The diffusion coefficient, D_n .
6. The survival rate (take $F_S = 1$ when we do not include therapy).

Formato para el fichero de parámetros de la simulación Se incluyen una serie de números en la misma línea, separados por espacios. No hay ningún caracter tras el último valor numérico. Por orden, se trata de:

1. El decay rate para el oxígeno, k_2 .
2. La tasa de consumo de oxígeno, k_1 .
3. El coeficiente de difusión del oxígeno, D_c .
4. El término fuente de la ecuación para el oxígeno, S .
5. El tiempo de parada de la simulación (suponiendo que empezamos a tiempo cero).
6. El número de ficheros de salida deseados para las distribuciones espaciales, $nfiles$ (aparte los de las condiciones iniciales). Equidistribuidos en tiempo entre el instante cero y el tiempo final especificado.
7. El paso de la malla, Δx .
8. El Δt sugerido para el integrador en tiempo (internamente se refina si fuera necesario para cumplir las condiciones de estabilidad pertinentes). Recomendable que divida de forma exacta al tiempo final.

Este formato puede variar (por el final) si hay que anhadir un flujo entrante, etc. Consultar el fichero de cabecera en cada caso. Todos los parámetros especificados van en unidades del SI siempre que sea aplicable.

1.5 Formato de salida

Se obtienen dos ficheros de salida y dos carpetas.

- Un fichero messages con información sobre los parámetros utilizados en la simulación asociada, expresados en las unidades del SI. También incluye los valores del equilibrio homogéneo no trivial asociado a los parámetros de la simulación proporcionados, expresado en las unidades del SI.
- Un fichero de dos columnas con el número total de individuos de la población (segunda columna) en función del tiempo (primera columna, en segundos), en los instantes de tiempo determinados conforme al tiempo final y el número de ficheros de salida especificados. Tales instantes están relacionados con un equiespaciado en tiempo (primer instante que el tiempo del proceso de Markov pasa por encima de cada una de las marcas) y van correlativos a los ficheros en espacio que la rutina devuelve, ver siguiente punto.
- Una carpeta con $nfiles+1$ ficheros que representan el número de individuos por casilla a cada instante de tiempo especificado (incluida la distribución inicial). El formato es a columna única, como la del fichero de entrada de la población.
- Otra carpeta similar para los valores de la concentración de oxígeno en función del tiempo.

1.6 Sobre integradores en tiempo

De momento tenemos un Euler explícito, el más sencillo y de peores prestaciones. RK4 puede ser muy lento, ver si hay discrepancias que justifiquen su uso. Específicos de la GSL por ver si merece la pena, e implícitos lo mismo, ya que la ventaja de tomar pasos en tiempo grandes se puede ver anulada si esperamos variación temporal rápida del λ , o cuando los mezclamos con el híbrido ya que estamos subordinados al τ del Gillespie -otra cosa es que usemos algún tipo de tau-leaping, por investigar. Otro problema de los implícitos es que el modelo no es lineal, de forma que a cada paso hay que resolver un sistema grande de ecuaciones via Newton o similar, lo que puede ser bastante caro (al menos si se plantea de forma inocente) y anular la supuesta ventaja de tomar Δt grande.

2 Secuencia de pasos del código

2.1 Procesamiento de la entrada

Tras leer todos los ficheros de entrada de la simulación, se calculan los equilibrios no triviales (c_∞, n_∞) conforme a¹

$$c_\infty = c_{cr} + c_{cr} \left(\frac{1}{a_- \nu} \log \left(\frac{2F_S}{1 + \nu \tau_p} \right) \right)^{-1/\beta}, \quad n_\infty = \frac{S - k_2 c_\infty}{k_1 c_\infty}.$$

Esta información se escribe en el fichero de messages. Se tomarán además un tiempo y un espacio adimensionales dados por

$$\hat{t} = t/\tau_p, \quad \hat{x} = x/l,$$

donde la escala característica en longitud viene dada por

$$l = \sqrt{D_n \tau_p}.$$

2.2 Cálculos iniciales

Se sobreentiende que de aquí en adelante trabajamos internamente con $\hat{c}(\hat{t}, \hat{x}) = c(t, x)/c_\infty$, donde c es la solución de (1.1). Se lleva a cabo esta operación con las lecturas iniciales de oxígeno.

Se definen los parámetros adimensionales

$$\hat{D}_c = D_c/D_n, \quad \hat{k}_1 = k_1 n_\infty \tau_p, \quad \hat{k}_2 = k_2 \tau_p, \quad \hat{S} = S \tau_p / c_\infty, \quad \hat{\nu} = \nu \tau_p.$$

Se definen además

$$\hat{a}_{G_1 \setminus S}(\hat{c}) = a_{G_1 \setminus S}(c)/\tau_p, \quad \hat{\lambda}(\hat{c}) = \lambda(c).$$

Para realizar el cálculo inicial de las edades de división, se utiliza el modelo (dimensional)

$$a_{G_1 \setminus S}(c) = a_- \left(\frac{c}{c_{cr}(p^*)} - 1 \right)^{-\beta}.$$

Aquí $p^* = p6/p3$ que de momento se está tomando como igual a uno. Para realizar los cálculos adimensionales, se introduce $\hat{a} = a/\tau_p$ y por tanto se reescribe todo como

$$\hat{a}_{G_1 \setminus S}(\hat{c}) = \frac{a_{G_1 \setminus S}(c)}{\tau_p} = \hat{a}_- \left(\frac{\hat{c}}{\hat{c}_{cr}(p^*)} - 1 \right)^{-\beta}$$

El valor a_- está fijado a 8.5 horas (30600 segundos), al menos por ahora. Nótese también que ahora mismo el valor $c_{cr}(p^*)$ está dado por la expresión

$$c_{cr}(p^*) = \frac{c_{typ} c_{cr}(p^*)}{100 c_{cr}(1)}$$

¹No sé si hay alguna condición cómoda para poder asegurar a priori que $n_\infty, c_\infty \geq 0$.

donde c_{typ} es un valor típico para la concentración del oxígeno², y c_{cr} es la fórmula de Tomás para el oxígeno crítico.

2.3 Determinación de la distribución inicial de edades

Se trata de muestrear la distribución de edades de equilibrio (adimensional),

$$P(\hat{a}) = \exp(-\hat{\nu}\hat{a} - \hat{\lambda}(\hat{c})\hat{a} - (\hat{a} - \hat{a}_{G_1 \setminus S})H(\hat{a} - \hat{a}_{G_1 \setminus S}))/Z, \quad (2.1)$$

(donde H es la función de Heaviside) el número adecuado de veces en cada localización espacial, conforme a la población inicial dada.

El hipotético caso en el que $\hat{\lambda}(\hat{c}) \simeq -\hat{\nu}$ puede ser problemático, de manera que lo tratamos aparte, justo como si fuera $\hat{\lambda}(\hat{c}) = -\hat{\nu}$. En esta situación,

$$Z = 1 + \hat{a}_{G_1 \setminus S}, \quad P_1(\hat{a}) := \int_0^{\hat{a}} ZP(\hat{a}) d\hat{a} = \hat{a} \quad \text{para } \hat{a} < \hat{a}_{G_1 \setminus S},$$

$$P_1^* := P_1(\hat{a}_{G_1 \setminus S}), \quad P_2(\hat{a}) := \int_{\hat{a}_{G_1 \setminus S}}^{\hat{a}} ZP(\hat{a}) d\hat{a} = 1 - \exp(\hat{a}_{G_1 \setminus S} - \hat{a}) \quad \text{para } \hat{a} > \hat{a}_{G_1 \setminus S}.$$

Para muestrear la distribución en este caso, se obtiene $r \in (0, 1)$ mediante una uniforme y se aplica (**por volver a comprobar**):

- Si $r \leq P_1^*/Z$, entonces $\hat{a} = rZ$.
- Si $r > P_1^*/Z$, entonces $\hat{a} = \hat{a}_{G_1 \setminus S} - \log(Z(1 - r))$.

En el resto de casos, se obtiene que

$$Z = \frac{1}{\hat{\nu} + \hat{\lambda}} \left(1 - \frac{\exp(-(\hat{\nu} + \hat{\lambda})\hat{a}_{G_1 \setminus S})}{1 + \hat{\nu} + \hat{\lambda}} \right), \quad P_1(\hat{a}) := \frac{1 - \exp(-(\hat{\nu} + \hat{\lambda})\hat{a})}{\hat{\nu} + \hat{\lambda}} \quad \text{para } \hat{a} < \hat{a}_{G_1 \setminus S},$$

$$P_2(\hat{a}) := \int_{\hat{a}_{G_1 \setminus S}}^{\hat{a}} ZP(a) d\hat{a} = \frac{e^{\hat{a}_{G_1 \setminus S}}}{1 + \hat{\nu} + \hat{\lambda}} [\exp(\hat{a}_{G_1 \setminus S}(1 + \hat{\nu} + \hat{\lambda})) - \exp(\hat{a}(1 + \hat{\nu} + \hat{\lambda}))] \quad \text{para } \hat{a} > \hat{a}_{G_1 \setminus S}.$$

Para muestrear la distribución ahora, se obtiene $r \in (0, 1)$ mediante una uniforme y se aplica (**por volver a comprobar**):

- Si $r \leq P_1^*/Z$, entonces $\hat{a} = -\frac{1}{\hat{\nu} + \hat{\lambda}} \log(1 - rZ(\hat{\nu} + \hat{\lambda}))$.
- Si $r > P_1^*/Z$, entonces $\hat{a} = -\frac{1}{1 + \hat{\nu} + \hat{\lambda}} [\log((1 - r)Z(1 + \hat{\nu} + \hat{\lambda})) - \hat{a}_{G_1 \setminus S}]$.

²Se aplican las mismas consideraciones delicadas que con la rutina coarse-grained.

Para realizar el cálculo de la tasa neta de proliferación (adimensionalizada) que va en (2.1) se reescribe (1.2) como

$$\frac{1}{2F_S} = \frac{\exp\left(-\hat{a}_{G_1 \setminus S}(\hat{c})(\hat{\nu} + \hat{\lambda}(\hat{c}))\right)}{1 + \hat{\nu} + \hat{\lambda}(\hat{c})}.$$

Que a su vez se resuelve tomando logaritmos, es decir, se le aplica Newton-Rhapson a

$$\log(1 + \hat{\nu} + \hat{\lambda}(\hat{c})) - \log(2F_S) + \hat{a}_{G_1 \setminus S}(\hat{c})(\hat{\nu} + \hat{\lambda}(\hat{c})) = 0.$$

2.4 Bucle principal

Se simula una versión adimensional de la ecuación para el oxígeno. Conforme a todo lo descrito anteriormente, se trata de

$$\frac{\partial \hat{c}}{\partial \hat{t}} = \hat{D}_c \frac{\partial^2 \hat{c}}{\partial \hat{x}^2} - \hat{k}_1 \hat{n} \hat{c} - \hat{k}_2 \hat{c} + \hat{S}. \quad (2.2)$$

Para discretizar las derivadas espaciales se usa la fórmula centrada de orden 2, con respecto al paso de malla adimensionalizado $\Delta \hat{x} = \Delta x/l$. A este fin, internamente se consideran coeficientes efectivos dados por

$$\bar{D}_c = \frac{\hat{D}_c}{(\Delta \hat{x})^2}, \quad \bar{k}_1 = \frac{\hat{k}_1 \Delta x}{n_\infty} = k_1 \tau_p \Delta x.$$

(El motivo del último es que la población se representa en el código como un vector de número de individuos por caja, así que la conversión a densidad normalizada es cómodo hacerla vía la constante multiplicativa; recordar que $\hat{n}(\hat{t}, \hat{x}) = n(t, x)/n_\infty$, estando n_∞ medido en el SI.) A cada paso del bucle principal se intenta avanzar el sistema un incremento de tiempo de $\Delta \hat{t}$ segundos; en el modelo adimensionalizado (2.2), esto se traduce en realizar un avance de $\Delta \hat{t} = \Delta t/\tau_p$. Si este incremento de tiempo no cumple las condiciones de estabilidad para hacer un Euler explícito con (2.2), entonces se refina internamente de forma que las cumpla.

Las ratios de transición para la población se adimensionalizan conforme al siguiente criterio **(no he podido repasar los cálculos aún:)** Doy las propensidades, todo son cinéticas de primer orden:

- Muerte: $\hat{\nu}$.
- Nacimiento: $H(\hat{a} - \hat{a}_{G_1 \setminus S})$
- Difusión (equiprobable en cada dirección): $1/(\Delta \hat{x})^2$.

2.5 Datos de salida y terminación

Se divide el intervalo temporal $[0, tstop]$ en $nfiles$ subintervalos de igual longitud. Con referencias en cada nodo temporal así obtenido se mandan datos a la salida, conforme se explicó en la subsección 1.5.

Nótese que si la simulación aborta antes de terminar, sí se generan las salidas espaciales previas al instante de fallo, pero no el fichero de mensajes ni el de número total de células vs tiempo. Creemos que el mecanismo de extensión de memoria genera violaciones de segmento en algunas ocasiones. Esto no está investigado en detalle, de momento lo sorteamos por fuerza bruta, asignando de salida $5 * 10^4$ casilla de memoria para edades en cada localización espacial.