

Universidad de La Coruña



MODELIZACIÓN DE LA EXPRESIÓN
GÉNICA MEDIANTE ECUACIONES
DIFERENCIALES

Modelización en Biomedicina

Miriam Gutiérrez Serrano

Índice

1. Introducción	2
2. Modelo matemático de expresión génica	2
2.1. Ejemplo adicional: comportamiento estable del sistema	4
3. Reconstrucción del modelo a partir de datos temporales	5
3.1. Transformada de Fourier para sistemas estables (FTSS)	5
3.2. Solución de peso mínimo de las ecuaciones lineales (MWSLE)	5
3.3. Ejemplo	6
4. Modelos extendidos y soluciones alternativas	6
4.1. Modelo 2: RNA Model	7
4.2. Modelo 3: Protein Model	7
4.3. Modelo 4: Time-Delay Model	7
5. Limitaciones del modelo	8
6. Conclusión	8

1. Introducción

La modelización matemática en biomedicina ofrece herramientas clave para comprender procesos biológicos complejos como la expresión génica. A través de ecuaciones diferenciales ordinarias (EDOs) o estocásticas como se ha visto en clase, es posible representar de manera precisa los mecanismos dinámicos subyacentes a la regulación genética. En particular, los sistemas de ecuaciones diferenciales permiten describir cómo varían las concentraciones de ARN mensajero (mRNA) y proteínas a lo largo del tiempo, considerando tanto procesos de síntesis como de degradación.

Se nos han introducido diversos métodos de modelización mediante EDOs, así como técnicas numéricas para su resolución. Estos modelos se han aplicado, por ejemplo, a ecuaciones tipo Langevin con ruido aditivo, o a procesos de Markov, todos ellos fundamentalmente para la comprensión de fenómenos aleatorios en biomedicina. En este contexto, el artículo proporcionado de Chen et al. (1999) es un ejemplo representativo y aplicado de este enfoque. Propone un modelo determinista basado en ecuaciones diferenciales para capturar la dinámica de la expresión génica con retroalimentación entre mRNA y proteínas. Además, introduce dos métodos para ajustar los parámetros del modelo a partir de datos temporales: uno basado en soluciones de sistemas lineales y otro en transformadas de Fourier.

En este proyecto resumimos los principales contenidos del artículo y presentamos una simulación numérica del modelo utilizando Python. Todo ello con el fin de mostrar la dinámica propuesta en un caso simplificado.

2. Modelo matemático de expresión génica

El artículo propone un modelo determinista de la expresión génica basado en un sistema de ecuaciones diferenciales ordinarias de tipo lineal. En este modelo se describen dos procesos acoplados: la transcripción, que genera mRNA a partir del ADN, y la traducción, que produce proteínas a partir del mRNA. Ambos tipos de moléculas se degradan a lo largo del tiempo, y las proteínas actúan como reguladores de la transcripción, introduciendo retroalimentación en el sistema.

El modelo básico se expresa mediante el siguiente sistema de ecuaciones:

$$\begin{cases} \frac{dr}{dt} = f(p) - Vr = Cp - Vr \\ \frac{dp}{dt} = Lr - Up \end{cases}$$

donde:

- $r(t) \in \mathbb{R}^n$: vector de concentraciones de mRNA,
- $p(t) \in \mathbb{R}^n$: vector de concentraciones de proteínas,
- $C \in \mathbb{R}^{n \times n}$: matriz de regulación (transcripción dependiente de proteínas),
- $V, U, L \in \mathbb{R}^{n \times n}$: matrices diagonales que representan, respectivamente, las tasas de degradación del mRNA, degradación de proteínas y eficiencia de traducción.

Este sistema se puede escribir en forma matricial como:

$$\frac{dx}{dt} = Mx, \quad \text{con } x = \begin{pmatrix} r \\ p \end{pmatrix}, \quad M = \begin{pmatrix} -V & C \\ L & -U \end{pmatrix}$$

Solución general del sistema lineal

Este sistema corresponde a una ecuación diferencial lineal con coeficientes constantes, cuya solución general puede escribirse mediante la exponencial de matrices:

$$x(t) = e^{Mt}x(0),$$

donde $x(0)$ es el estado inicial y M es la matriz de coeficientes del sistema. Esta solución se basa en la teoría general de ecuaciones diferenciales lineales homogéneas, vista en clase.

En el caso de que M sea diagonalizable, es decir, si existe una matriz Q invertible tal que $M = Q\Lambda Q^{-1}$ con Λ diagonal, entonces:

$$e^{Mt} = Qe^{\Lambda t}Q^{-1},$$

donde $e^{\Lambda t}$ se calcula de forma sencilla como una matriz diagonal con las exponenciales de los autovalores de M .

Esta formulación permite estudiar cualitativamente la estabilidad del sistema: si todos los autovalores de M tienen parte real negativa, la solución tenderá a cero con el tiempo (sistema estable). En cambio, si alguno tiene parte real positiva, las soluciones crecerán exponencialmente, como se observa en el ejemplo de simulación que mostramos a continuación.

Este modelo se conoce en el artículo como Modelo Lineal de Transcripción (Linear Transcription Model), y constituye la base sobre la que se construyen las versiones más complejas, incluyendo retardos o modelos reducidos a solo mRNA o proteínas.

Para mostrar el comportamiento del modelo, hemos implementado una simulación para un sistema con $n = 2$ genes. Hemos utilizado los siguientes parámetros:

- Matriz de regulación

$$C = \begin{pmatrix} 0,5 & 0,1 \\ 0,3 & 0,7 \end{pmatrix}$$

- Matrices de degradación y traducción

$$V = \text{diag}(0,1, 0,3), \quad U = \text{diag}(0,2, 0,1), \quad L = \text{diag}(0,8, 0,9)$$

- Condiciones iniciales

$$r(0) = (1, 0,5), \quad p(0) = (0,2, 0,1)$$

Como hemos visto en clase, el sistema se ha resuelto con el método de integración `solve_ivp`, evaluando en el intervalo $t \in [0, 50]$. La siguiente figura muestra la evolución temporal de las concentraciones de ARNm y proteínas.

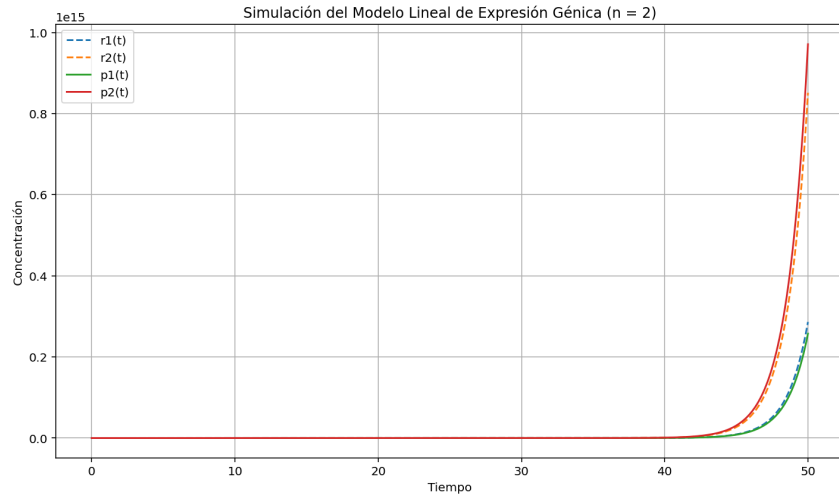


Figura 1: Simulación del Modelo Lineal de Transcripción con 2 genes.

Como podemos observar, las concentraciones crecen rápidamente: al menos uno de los autovalores de M tiene parte real positiva. Por lo que, coincide con la teoría expuesta en el artículo: una mala elección de parámetros puede producir soluciones no realistas.

2.1. Ejemplo adicional: comportamiento estable del sistema

Para ilustrar la influencia de los parámetros en la dinámica del modelo, simulamos nuevamente el sistema lineal, esta vez con valores que garantizan estabilidad. En concreto, se eligieron matrices con valores menores en C y mayores en V y U :

$$C = \begin{pmatrix} 0,1 & 0,05 \\ 0,02 & 0,1 \end{pmatrix}, \quad V = \text{diag}(0,5, 0,4), \quad U = \text{diag}(0,3, 0,6), \quad L = \text{diag}(0,2, 0,3)$$

Con estas condiciones, se observa que todas las variables del sistema tienden a un equilibrio, como corresponde a un sistema lineal estable. La siguiente figura muestra la evolución temporal del sistema:

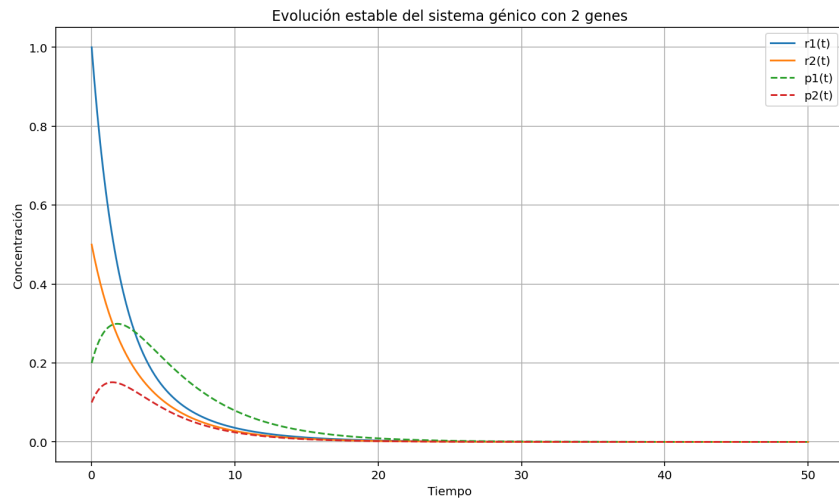


Figura 2: Simulación del modelo lineal con parámetros estables.

Este resultado confirma que, ajustando adecuadamente las tasas de transcripción, traducción y degradación, el modelo puede representar un comportamiento biológico realista con homeostasis de las concentraciones.

3. Reconstrucción del modelo a partir de datos temporales

Como hemos visto, el modelo lineal propuesto en el artículo describe dinámicas de mRNA y proteínas en términos de un sistema de ecuaciones. Sin embargo, para poder usarlo en la práctica es necesario estimar las matrices del sistema (C, V, L, U, \dots) a partir de datos experimentales. El artículo propone dos métodos distintos para realizar esta reconstrucción:

3.1. Transformada de Fourier para sistemas estables (FTSS)

Este método se basa en asumir que las soluciones del sistema son estables y periódicas. En tal caso, cada variable del modelo puede aproximarse mediante una serie de Fourier:

$$x_j(t) = \sum_{k=-N}^N c_{jk} e^{i w_k t}, \text{ con } w_k = \frac{2\pi k}{T}$$

donde

- $x_j(t)$ es la j -ésima variable (mRNA o proteína),
- c_{jk} son los coeficientes de Fourier, y
- T periodo.

La derivada temporal $\frac{dx_j}{dt}$ se obtiene derivando la serie término a término, y el modelo se reescribe como un sistema lineal algebraico para cada t . De esta forma, se puede calcular la matriz M del sistema.

Este enfoque es eficiente computacionalmente, pero depende fuertemente de la estabilidad del sistema y la calidad de la aproximación periódica.

3.2. Solución de peso mínimo de las ecuaciones lineales (MWS-LE)

El segundo método se basa en un sistema de la forma

$$Ax = b$$

con A una matriz de las mediciones de las variables del sistema y b a las derivadas temporales estimadas numéricamente.

En vez de resolver directamente el sistema, se busca una solución dispersa (sparse). Suponemos que cada gen está regulado por un número pequeño de genes, es decir,

$$\min \|x\|_0, \quad Ax = b.$$

El artículo propone resolver una relajación con soluciones de peso mínimo restringido a h entradas no nulas y evaluar los subconjuntos posibles $\left(\binom{n}{k}\right)$.

Así, no es necesario asumir periodicidad ni estabilidad, y puede aplicarse a datos reales no acotados. A cambio, su coste computacional crece rápidamente a medida que incrementan el número de genes.

Ambos métodos permiten estimar los parámetros del modelo a partir de datos experimentales, y así se puede usar una estructura matemática en tareas como predicción de dinámicas o identificación de genes reguladores clave.

3.3. Ejemplo

Dado un sistema dinámico como el del modelo lineal

$$\frac{dr}{dt} = Cp - Vr, \quad (1)$$

queremos reconstruir las matrices C y V a partir de datos simulados de $r(t)$, $p(t)$ y $\frac{dr}{dt}$. Supongamos que tenemos muestras para cada instante de tiempo t_i con $i = 1, \dots, m$. La ecuación queda de la forma

$$\frac{dr}{dt}(t_i) = Cp(t_i) - Vr(t_i) \quad (2)$$

y, como en el modelo comentado, agrupamos en un sistema lineal $Ax = b$, donde x contiene las entradas de C y V .

El método permite recuperar las matrices originales con buena precisión. La salida del código proporciona las matrices estimadas y el error:

```
C estimada:
[[-0.32071331 -0.1657487 ]
 [ 0.15823528 -0.01543166]]
V estimada:
[[-0.11895218  0.          ]
 [ 0.          0.13989327]]
Error en C: 1.150033158117958
Error en V: 0.4208472389832928
```

Este ejemplo muestra cómo puede plantearse la reconstrucción del modelo usando mínimos cuadrados, tal como se ha trabajado. A diferencia de otros métodos más complejos o con supuestos adicionales (como estabilidad o periodicidad), esta técnica puede aplicarse de forma directa a datos temporales con ruido, sin requerir regularización.

4. Modelos extendidos y soluciones alternativas

Además del modelo lineal básico de expresión génica, el artículo presenta varias extensiones que permiten adaptar el sistema a distintos contextos biológicos o simplificar su tratamiento matemático. Estas variantes incluyen modelos con retardos temporales, modelos reducidos y soluciones aproximadas.

4.1. Modelo 2: RNA Model

Este modelo elimina las proteínas como variables explícitas para centrarse únicamente en la dinámica del mRNA $r(t)$. A partir del sistema:

$$\frac{dr}{dt} = Cp - Vr, \quad \frac{dp}{dt} = Lr - Up,$$

se sustituye la solución general de $p(t)$ usando variación de constantes:

$$p(t) = e^{-Ut} \int e^{Ut} Lr(t) dt,$$

y al introducir esta expresión en la ecuación de r , se obtiene finalmente una ecuación diferencial de segundo orden para r :

$$\frac{d^2r}{dt^2} = (-CUC^{-1} - V)\frac{dr}{dt} + (-CUC^{-1}V + CL)r.$$

Este modelo es útil cuando solo se dispone de datos de mRNA. Sin embargo, la degeneración potencial de la matriz C implica que r no está completamente determinada sin conocer el estado inicial de p , lo cual refleja la dependencia funcional de la transcripción respecto a las proteínas.

4.2. Modelo 3: Protein Model

Análogamente, se puede obtener un modelo solo en términos de proteínas $p(t)$, eliminando las variables $r(t)$. El resultado es:

$$\frac{d^2p}{dt^2} = (-LVL^{-1} - U)\frac{dp}{dt} + (-LVL^{-1}U + LC)p.$$

Este modelo permite trabajar directamente con datos de proteínas. Como V , L y U son matrices diagonales, los términos LVL^{-1} y $LVL^{-1}U$ también lo son, y LC hereda la esparsidad de C . Por tanto, el modelo puede reconstruirse usando el mismo esquema de mínimos cuadrados dispersos descrito anteriormente.

4.3. Modelo 4: Time-Delay Model

Para reflejar mecanismos más realistas, el artículo introduce un modelo con retardos en la transcripción y traducción. Sean τ y δ los retardos de transcripción y traducción, respectivamente:

$$\begin{cases} \frac{dr}{dt}(t) = Cp(t - \tau) - Vr(t), \\ \frac{dp}{dt}(t) = Lr(t - \delta) - Up(t). \end{cases}$$

Aplicando la transformada de Fourier al sistema se obtiene una condición sobre los valores propios ω de la forma:

$$\det(i\omega I + V - Ce^{-i\omega\tau}(i\omega I + U)^{-1}Le^{-i\omega\delta}) = 0.$$

Este conjunto de frecuencias define el espectro del sistema. Aunque no se resuelve de forma explícita, el artículo señala que el análisis de estabilidad del sistema se puede estudiar a través de esta condición espectral.

5. Limitaciones del modelo

El artículo reconoce que el Modelo Lineal de Transcripción (Modelo 1) presenta varias limitaciones importantes:

- No incluye retardos de forma natural, salvo en el Modelo 4.
- No contempla otros reguladores relevantes como metabolitos u otras señales celulares.
- El método de reconstrucción por transformada de Fourier supone periodicidad, lo cual no siempre se verifica en datos reales.
- El método MWSLE asume que cada gen tiene pocos reguladores (modelo esparso), lo cual puede no cumplirse en sistemas altamente conectados.

A pesar de estas limitaciones, el modelo propuesto permite capturar dinámicas fundamentales de la expresión génica con una formulación matemática clara y eficiencia computacional, siendo una herramienta útil para el análisis preliminar y la reconstrucción de redes génicas.

6. Conclusión

El artículo de Chen et al. (1999) propone un marco matemático claro y versátil para modelar la expresión génica mediante ecuaciones diferenciales lineales. A partir de un sistema acoplado que describe la transcripción y traducción de genes, se construyen diversas extensiones que permiten adaptarse a distintos tipos de datos experimentales y fenómenos biológicos, como los retardos o los modelos reducidos a mRNA o proteínas.

A lo largo de este trabajo se han reproducido los elementos principales del modelo y se han implementado simulaciones numéricas para estudiar su comportamiento en distintos regímenes. Se ha verificado que una elección adecuada de parámetros puede conducir a soluciones estables y biológicamente coherentes, mientras que otros valores generan dinámicas no realistas.

Además, se ha descrito el proceso de reconstrucción de los parámetros del modelo a partir de datos temporales, tanto mediante la transformada de Fourier como a través de mínimos cuadrados esparsos. Estas metodologías son relevantes para la inferencia de redes génicas en aplicaciones reales de la biología computacional.

El código utilizado en Python para la resolución numérica y reconstrucción del modelo se adjunta como material complementario para su consulta.