UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

WILLIAN ANDRIGHETTO TREVIZAN

**Interação entre genes no modelo de spins e
bósons**

São Carlos
2011

WILLIAN ANDRIGHETTO TREVIZAN

Interação entre genes no modelo de spins e bósons

Dissertação apresentada ao Programa de Pós-graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para a obtenção do título de Mestre em Ciência.

Área de Concentração: Física Básica
Orientador: Prof. Dr. José Eduardo Martinho
Hornos

Versão Original

São Carlos

2011

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO, PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica elaborada pelo Serviço de Biblioteca e Informação IFSC/USP

Trevizan, Willian Andrighetto.

Interação entre genes no modelo de spins e bósons / Willian Andrighetto Trevizan; orientador José Eduardo Martinho Hornos.-- São Carlos, 2011.
79 p.

Dissertação (Mestrado em Ciências - Área de concentração: Física Básica) – Instituto de Física de São Carlos da Universidade de São Paulo.

1. Modelos estocásticos. 2. Expressão gênica. 3. Soluções exatas. 4. Analiticidade. 5. Método variacional. I. Título.

FOLHA DE APROVAÇÃO

Willian Andrighetto Trevizan

Dissertação apresentada ao Instituto de Física de São Carlos da Universidade de São Paulo para obtenção do título de Mestre em Ciências.
Área de Concentração: Física Básica

Aprovado(a) em: 24.02.2011

Comissão Julgadora

Prof(a). Dr(a). Alexandre Souto Martinez

Instituição: FFCLRP/USP

Assinatura

Prof(a). Dr(a). Luiz Agostinho Ferreira

Instituição: IFSC/USP

Assinatura

Prof(a). Dr(a). Jose Eduardo Martinho Hornos

Instituição: IFSC/USP

Assinatura

AGRADECIMENTOS

Em reconhecimento ao processo percorrido para chegar até aqui, rendo homenagem a todos os que permitiram que o objetivo fosse alcançado. Agradeço à família, pela oportunidade de estar aqui e por todo o amor e apoio concendidos. A todos que se apresentaram como grandes companheiros, dispostos a dividir os bons momentos e me aturar nos não tão bons. Aos bravos guerreiros (inclusive os desertores!) ao lado dos quais combati nestes anos de graduação e mestrado! Aos professores do Instituto de Física de São Carlos pela formação técnica e acadêmica, pela preocupação e dedicação na maioria dos casos. Aos funcionários por todo o suporte e galhos quebrados. À FAPESP pelo financiamento do projeto. Ao meu orientador pelo aprendizado, reconhecimento e paciência. Em especial, meu muito obrigado aos Amigos.

RESUMO

TREVIZAN, Willian Andrighetto *Interação entre genes no modelo de spins e bósons*. 2011. 79 p. Dissertação (Mestrado) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2011

Vários módulos funcionais de células ou bactérias são controlados por meio de genes que se regulam através da codificação de proteínas repressoras ou indutoras. A compreensão destas redes gênicas é um problema em aberto da biologia. Nesta dissertação procuramos aplicar o modelo estocástico de spins e bósons para o caso da rede mais simples, constituída de dois genes, o primeiro reprimindo o segundo. Apresentamos a solução exata para a distribuição de probabilidades sobre o número das proteínas dos dois genes no estado estacionário, e a probabilidade dependente do tempo sobre o estado do segundo gene (ativado ou desativado). Discutimos também maneiras de generalizar os resultados para redes maiores.

Palavras-chave: Modelos estocásticos. Expressão gênica. Soluções exatas. Analiticidade. Método variacional.

ABSTRACT

TREVIZAN, Willian Andrighetto *Gene interactions in the spin-boson model*. 2011. 79 p. Dissertação (Mestrado) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2011

Several functional modules in cells or bacteria are controlled by genes that regulate each other by coding repressive or inducer proteins. The understanding of these genetic networks is still an open problem in biology. In this work we apply the stochastic spin-boson model to the simplest interacting network, made of two genes, the first repressing the second. We present the exact solution for the stationary probability distribution over their both protein numbers, and the temporal evolution of the second gene's state (on or off). We also discuss ways of generalizing these results to larger networks.

Keywords: Stochastic models. Gene expression. Exact solutions. Analyticity. Variational method

LISTA DE FIGURAS

Figura 2.1 - Distribuição conjunta de probabilidades $\phi_{n,m}$ para os seguintes valores de parâmetros: $K = 5, \varrho = 0.5, \kappa = 40, \rho = 20, F = 5, \chi = 0.1$	26
Figura 2.2 - Redução da distribuição conjunta $\phi_{n,m}$ de 2.1 para ϕ_n	28
Figura 2.3 - Comportamento da probabilidade P_α em função do número médio $\langle n \rangle$ de proteínas repressoras, para $H = 0.5$ e diferentes valores de F	30
Figura 2.4 - Comportamento da probabilidade P_α em função do número médio $\langle n \rangle$ de proteínas repressoras, para $F = 1$ e diferentes valores de H	31
Figura 2.5 - Redução da distribuição conjunta $\phi_{n,m}$ da figura 2.1 para ϕ_m . $\langle n \rangle = 10, \frac{k_2}{\rho_2} = 20, P_\alpha \approx 0.5$	32
Figura 2.6 - Distribuição conjunta $\phi_{n,m}$ para $K = 9, \varrho = 0.9, \kappa = 40, \rho = 20, F = 5, \chi = 0.1$. Desta maneira, $\langle n \rangle = 10, \frac{k_2}{\rho_2} = 20$ e $P_\alpha \approx 0.8$	33
Figura 2.7 - Redução da distribuição conjunta $\phi_{n,m}$ da figura 2.6 para ϕ_m	33
Figura 2.8 - Distribuição conjunta $\phi_{n,m}$ para $K = 5, \varrho = 0.5, \kappa = 20, \rho = 1, F = 5, \chi = 0.1$. Desta maneira, $\langle n \rangle = 10, \frac{k_2}{\rho_2} = 20$ e $P_\alpha \approx 0.5$	34
Figura 2.9 - Curvas de nível da função $g(b_0, a)$. Na região amarela, alaranjada e vermelha temos respectivamente $g > 0.5, 0.1 < g < 0.5$ e $g < 0.1$. Para $b_0 \rightarrow 0$ e $a \neq 0, g \rightarrow \infty$	36
Figura 2.10 - Gráficos paramétricos das curvas dadas por $(a(H), b_0(H))$, com H indo de 0 a 1, para $F = 0.1$ e diferentes valores de $\langle n \rangle$	37
Figura 2.11 - Comparação entre os cálculos exatos e aproximados para a distribuição conjunta $\phi_{n,m}$. $\langle n \rangle = 10, \kappa = 40, \rho = 20, \chi = 0.1, P_\alpha = 0.43, a = 4.9, b_0 = 7.1$	41
Figura 2.12 - Comparação entre os cálculos exatos e aproximados para a distribuição conjunta $\phi_{n,m}$. $\langle n \rangle = 10, \kappa = 20, \rho = 1, \chi = 0.1, P_\alpha = 0.43, a = 4.9, b_0 = 7.1$	41

Figura 2.13 - <i>Comparação entre os cálculos exatos e aproximados para a distribuição conjunta $\phi_{n,m}$. $\langle n \rangle = 10, \kappa = 20, \rho = 1, \chi = 0.1, P_\alpha = 0.63, a = 0.9, b_0 = 7.1$</i>	42
Figura 2.14 - <i>Comparação entre os cálculos exatos e aproximados para a distribuição conjunta $\phi_{n,m}$. $\langle n \rangle = 10, \kappa = 40, \rho = 20, \chi = 0.1, P_\alpha = 0.63, a = 0.9, b_0 = 7.1$</i>	42
Figura 2.15 - <i>Comparação entre os cálculos exatos e aproximados para a distribuição conjunta $\phi_{n,m}$. $\langle n \rangle = 3, \kappa = 2, \rho = 0.1, \chi = 0.1, P_\alpha = 0.20, a = 0.75, b_0 = 1.05$</i>	43
Figura 2.16 - <i>Comparação entre os cálculos exatos e aproximados para a distribuição conjunta $\phi_{n,m}$. $\langle n \rangle = 3, \kappa = 100, \rho = 5, \chi = 0.1, P_\alpha = 0.20, a = 0.75, b_0 = 1.05$</i>	43
Figura 2.17 - <i>Comparação entre os cálculos exatos e aproximados para a distribuição conjunta $\phi_{n,m}$. $\langle n \rangle = 0.8, \kappa = 2, \rho = 0.1, \chi = 0.1, P_\alpha = 0.35, a = 0.07, b_0 = 0.27$</i>	44
Figura 2.18 - <i>Comparação entre os cálculos exatos e aproximados para a distribuição conjunta $\phi_{n,m}$. $\langle n \rangle = 0.8, \kappa = 100, \rho = 5, \chi = 0.1, P_\alpha = 0.35, a = 0.07, b_0 = 0.27$</i>	44
Figura 2.19 - <i>Efeito do parâmetro de adiabaticidade ε sobre a distribuição de proteínas do gene 2, na solução aproximada. Nas três curvas, $\frac{k_2}{\rho_2} = 30, \chi = 0.2$ e $P_\alpha^{ap} = 0.6$</i>	46
Figura 2.20 - <i>Comportamento da média no número de proteínas do tipo 1 como função da concentração de IPTG</i>	50
Figura 2.21 - <i>Comportamento do ruído no número de proteínas do tipo 1 como função da concentração de IPTG</i>	50
Figura 2.22 - <i>Comportamento da média no número de proteínas do tipo 2 como função da concentração de IPTG</i>	51
Figura 2.23 - <i>Comportamento do ruído no número de proteínas do tipo 2 como função da concentração de IPTG</i>	51
Figura 2.24 - <i>Coefficiente de correlação entre as proteínas 1 e 2 como função da concentração de IPTG</i>	52

SUMÁRIO

1	Introdução	15
1.1	O modelo de spins e bósons na interação gênica	18
2	Soluções estacionárias	23
2.1	Resultados exatos	23
2.1.1	Solução	23
2.1.2	Distribuição de probabilidades para as proteínas repressoras	27
2.1.3	Probabilidade de ocupação do gene 2	29
2.1.4	Número médio de proteínas do gene 2	31
2.1.5	Correlação	34
2.2	Resultados aproximados	38
2.2.1	Funções geradoras	38
2.2.2	Propriedades da solução aproximada e estimativa dos tempos de ida ao equilíbrio	40
2.3	Expressão em cascata	47
3	Método variacional	53
3.1	Descrição do método	53
3.2	Aplicação do método para o sistema de dois genes descrito no capítulo 2 .	56
4	Soluções exatas dependentes do tempo	59
4.1	Solução pelo método das características	60
4.1.1	Solução para $\phi(z, t)$	60

4.1.2	Solução para $\alpha(z, t)$	61
4.1.3	Solução explícita em termos das condições iniciais	63
4.2	Solução em termos das funções de Heun confluentes	65
5	Conclusão	69
	REFERÊNCIAS	71
	APÊNDICE A – Obtenção da equação 2.13	75
	APÊNDICE B – Funções geradoras aproximadas	77

1 *Introdução*

Toda a informação acumulada através da passagem de inúmeras gerações de uma determinada espécie ao longo da evolução está inscrita nos genes que compõe o DNA de cada indivíduo desta espécie. O processo através do qual essas informações são decodificadas, ou transformadas, em proteínas é chamado de expressão gênica. As proteínas são as executoras das ações inscritas no DNA.

Um ser vivo não expressa todo o seu genoma a todo o tempo, de modo que apenas alguns conjuntos de genes estão ativos em determinados estágios da vida do indivíduo. Como um exemplo, podemos citar o metabolismo da lactose pela bactéria *Escherichia coli*. Para a bactéria é energeticamente desfavorável produzir a enzima que digere a lactose (β -galactosidase) quando não há alimento disponível, ou quando há no meio fontes de energia mais diretas, como a glucose. Portanto, é necessário nos seres vivos um mecanismo de regulação gênica para controlar quando e quais partes do genoma devem ser ativados.

O mecanismo qualitativo fundamental que governa a expressão gênica em bactérias foi descoberto por Monod nos experimentos com *E. coli* que levaram ao conceito do operon lac (1-2) e também com os experimentos envolvendo o bacteriófago λ (3). No caso da digestão de lactose em *E. coli* a regulação se dá da seguinte maneira. Os três genes envolvidos nos processos de transporte e quebra da lactose (dentre os quais o codificador da β -galactosidase) encontram-se agrupados em uma unidade genômica básica do DNA chamada operon lac. O operon lac é regulado negativamente pelo gene *lacI* que constitutivamente codifica uma proteína repressora, que quando ligada ao sítio operador do operon lac impede ou dificulta a transcrição de seus genes. Entretanto, quando presente no meio, a lactose se liga às proteínas repressoras inutilizando-as, deixando desta forma o sítio operador do operon lac livre e os genes responsáveis pelo metabolismo do alimento sendo copiosamente transcritos.

Esta rede de regulação com um gene produzindo uma proteína que reprime um outro gene (*lacI*-operon lac) é uma das mais simples que podemos encontrar. Em outros módulos

funcionais de uma bactéria podemos nos deparar com dezenas de genes que se regulam por meio das proteínas por eles codificadas, formando uma cadeia de interações à semelhança dos circuitos elétricos. Em eucariotos a questão é ainda mais articulada mesmo em processos simples como a segmentação da *Drosophila* (4). A expressão do gene *even*, por exemplo, envolve sete proteínas reguladoras e pelo menos dezessete sítios de ligação numa região de 1.6 kilobases.

A maneira canônica de proceder no estudo das redes gênicas consiste em estudar as unidades mais simples da regulação gênica (como o exemplo *lacI*-operon *lac*), e através de seu entendimento teórico procurar maneiras de estender os resultados para redes mais gerais, através da elaboração de regras que permitam o acoplamento entre estas unidades.

Entretanto, a descrição deste tipo de sistema não envolve apenas o conhecimento do número de proteínas presentes na célula. Devido ao fato de os reagentes envolvidos nas reações de transcrição e tradução (RNA polimerases, fatores de transcrição, RNA mensageiros, etc) se apresentarem em geral em números pequenos, a presença da estocasticidade no número destes elementos é inevitável (5–7). Esta variabilidade é importante do ponto de vista biológico, pois permite a adaptabilidade das espécies frente a mudanças do ambiente, mas espera-se que haja um mecanismo de controle deste ruído de modo a não debilitar a confiabilidade da rede. A compreensão destes sistemas se dá portanto através do conhecimento das distribuições de probabilidade de encontrar um certo número de proteínas (ou outros fatores de interesse) na célula (no caso *lacI*-operon *lac*, as variáveis estocásticas podem ser o número de proteínas repressoras e o número de proteínas produzidas pelo gene reprimido).

No sentido de incorporar a estocasticidade na modelagem teórica das redes gênicas, tem-se observado na literatura de maneira geral três tipos de abordagem. Na primeira, investiga-se o comportamento das redes através de simulações de Monte Carlo (8–11), incorporando através de equações simplificadas as reações químicas envolvidas. Apesar desta abordagem poder ser utilizada em redes de tamanhos em princípio arbitrários (dentro da capacidade computacional permitida), a investigação por meio de um método puramente numérico não nos permite ter um entendimento preciso da relevância de cada um dos processos e efeitos envolvidos. Algumas simulações mais sofisticadas e de menor custo computacional que levam em conta o conhecimento de redes menores também têm sido desenvolvidas (12).

Em uma outra frente o número de proteínas é tratado como uma variável contínua que evolui no tempo segundo um sistema de equações diferenciais acopladas, que levam em

conta as interações de indução e repressão entre os componentes da rede. O ruído é introduzido no sistema por meio da técnica de Langevin (13–16), o que faz com que as análises normalmente se restrinjam aos momentos mais baixos da distribuição de probabilidades. Em geral algoritmos baseados em equações diferenciais acopladas (sem a introdução dos termos flutuantes) são utilizados na identificação das conexões entre os componentes das redes reais (17).

Finalmente, uma terceira abordagem foi sugerida por Kepler (18), Thattai (19) e Wolynes (20).

A proposta é modelar redes pequenas em que as reações envolvidas (criação e degradação de proteínas, RNA mensageiros, etc) são tratadas como processos de Markov que modificam as distribuições de probabilidades. Estas são portanto calculadas através de equações mestras, diferenciais no tempo e diferença no número de componentes (21–22). Como os processos envolvidos na transcrição e tradução são em geral bastante complicados, envolvendo muitas vezes vários elementos diferentes, uma abordagem através de uma equação de taxas como as equações mestras deve ser encarada como uma descrição mesoscópica do sistema. As taxas representariam portanto processos efetivos, que procuram incorporar todos os efeitos microscópicos intrínsecos. Por exemplo, alguns modelos markovianos (23) incorporam a transcrição e a tradução em uma única taxa de síntese proteica.

A vantagem desta terceira abordagem é a possibilidade de obtenção de soluções analíticas (devido ao fato de as redes serem pequenas). Um resultado analítico aliado ao nível de detalhe que um modelo mesoscópico pode atingir confere a capacidade de realização de aproximações poderosas para redes realistas.

O modelo introduzido por Peter Wolynes trata um gene como uma entidade binária, da seguinte maneira: quando o gene se encontra livre de proteínas repressoras ele é transcrito a uma certa taxa, enquanto que quando ocupado por um repressor ou indutor, é transcrito a uma taxa menor ou maior, respectivamente. O alvo da teoria são as probabilidades de encontrar o gene ligado ou desligado, e a transição entre estes estados depende do número de proteínas repressoras ou indutoras presentes no meio. Devido à analogia entre este sistema e um spin eletrônico (estados up e down) na presença de um campo bosônico (operadores de campo ligados à variação do número de fótons do meio), este modelo leva o nome de modelo de Spins e Bósons.

Uma vantagem desta analogia é a possibilidade de importar técnicas utilizadas em física do estado sólido, como a aproximação de Hartree-Fock, para os sistemas biológicos.

O modelo de Spins e Bósons foi resolvido por Innocentini e Hornos para o caso em que as taxas de ligamento e desligamento do gene independem do número de proteínas repressoras (24) e por Hornos *et al.* para o caso do gene auto-repressor (25), este último sendo de especial importância para a *E. coli* tendo em vista que cerca de 30 por cento do seu genoma é composto por genes auto-interagentes. As soluções para estado estacionário envolvem funções hipergeométricas confluentes (dois pontos singulares) e as soluções dependentes do tempo envolvem as funções de Heun confluentes (três pontos singulares). A incorporação de funções com mais de uma singularidade na biologia teórica demonstra um avanço natural no entendimento dos sistemas biológicos, uma vez que estas funções têm aparecido também em campos da física como relatividade geral (26).

Nesta dissertação analisamos o modelo de spins e bósons para o sistema de dois genes interagentes mais simples, como a rede *lacI-operon lac*. O gene repressor funciona em apenas um estado e as proteínas repressoras desligam um segundo gene, que pode funcionar em dois estados (ligado e desligado). No Capítulo 2 apresentamos a solução exata para a distribuição de probabilidades estacionária, e uma solução aproximada em termos de funções hipergeométricas confluentes. No mesmo capítulo mostramos como esta solução pode ser utilizada para a reprodução de dados experimentais para uma rede de três genes apresentada em (15). O Capítulo 3 explora um método variacional para este sistema, mostrando sua consistência com a solução aproximada do Capítulo 2. O Capítulo 4 traz a probabilidade de ocupação do sítio operador do gene reprimido (de ele estar ligado ou desligado) como função do tempo. Esta solução é dada em termos das funções confluentes de Heun. O último capítulo é dedicado às conclusões.

1.1 O modelo de spins e bósons na interação gênica

Consideremos como explicado anteriormente, um sistema constituído de dois genes, indexados por 1 e 2. O gene 1 é o repressor, que possui apenas um estado de funcionamento no qual produz as proteínas 1 (repressoras) a uma taxa (probabilidade por unidade de tempo) k_1 , que são degradadas cada uma a uma taxa ρ_1 (n proteínas repressoras são degradadas a uma taxa $n\rho_1$).

O gene 2 pode funcionar em dois estados. Quando seu sítio operador está livre, o gene está ligado e produz proteínas 2 a uma taxa k_2 . Se o sítio operador do gene 2 está ocupado por uma proteína repressora, produz proteínas a uma taxa menor, dada por χk_2 , onde $0 \leq \chi < 1$. Em ambos os estados, as proteínas do segundo gene são degradadas com taxa $m\rho_2$, onde m é o número de proteínas 2.

Assumiremos que a probabilidade por unidade de tempo de uma proteína repressora se ligar ao gene 2, desligando-o, é proporcional ao número n de proteínas repressoras (nh). A taxa com que uma proteína repressora se solta do gene 2, ligando-o, depende apenas da afinidade entre esta proteína e o sítio operador, e portanto é constante (f).

A probabilidade de em um instante t encontrar o gene 2 ligado, com n proteínas do gene 1 e m do gene 2 na célula é dada por $\alpha_{n,m}(t)$. Analogamente, a probabilidade de encontrar o gene 2 desligado, com n e m proteínas 1 e 2 é dada por $\beta_{n,m}(t)$. Desta maneira, a probabilidade conjunta de encontrar no meio n proteínas do tipo 1 e m do tipo 2 é $\phi_{n,m}(t) = \alpha_{n,m}(t) + \beta_{n,m}(t)$.

As equações mestras que governam estes processos de Markov são (21-22):

$$\begin{aligned} \frac{\partial \alpha_{n,m}}{\partial t} = & k_1 (\alpha_{n-1,m}(t) - \alpha_{n,m}(t)) + \rho_1 ((n+1)\alpha_{n+1,m}(t) - n\alpha_{n,m}(t)) \\ & k_2 (\alpha_{n,m-1}(t) - \alpha_{n,m}(t)) + \rho_2 ((m+1)\alpha_{n,m+1}(t) - m\alpha_{n,m}(t)) \\ & -nh\alpha_{n,m}(t) + f\beta_{n,m}(t) \end{aligned} \quad (1.1)$$

$$\begin{aligned} \frac{\partial \beta_{n,m}}{\partial t} = & k_1 (\beta_{n-1,m}(t) - \beta_{n,m}(t)) + \rho_1 ((n+1)\beta_{n+1,m}(t) - n\beta_{n,m}(t)) \\ & \chi k_2 (\beta_{n,m-1}(t) - \beta_{n,m}(t)) + \rho_2 ((m+1)\beta_{n,m+1}(t) - m\beta_{n,m}(t)) \\ & +nh\alpha_{n,m}(t) - f\beta_{n,m}(t) \end{aligned} \quad (1.2)$$

Como dito anteriormente, as equações mestras são equações diferenciais-diferença. Para transformá-las em equações diferenciais apenas, vamos introduzir as funções geradoras:

$$\begin{aligned} \alpha(z, y, t) &= \sum_{n,m=0}^{\infty} \alpha_{n,m}(t) z^n y^m \\ \beta(z, y, t) &= \sum_{n,m=0}^{\infty} \beta_{n,m}(t) z^n y^m \\ \phi(z, y, t) &= \sum_{n,m=0}^{\infty} \phi_{n,m}(t) z^n y^m \end{aligned} \quad (1.3)$$

Em termos das funções geradoras, as equações mestras ficam:

$$\begin{aligned} \frac{\partial \phi}{\partial t} = & (z-1) \left(k_1 \phi(z, y, t) - \rho_1 \frac{\partial \phi}{\partial z} \right) + \\ & + (y-1) \left(k_2 (1-\chi) \alpha(z, y, t) + \chi k_2 \phi(z, y, t) - \rho_2 \frac{\partial \phi}{\partial y} \right) \end{aligned} \quad (1.4)$$

$$\begin{aligned} \frac{\partial \alpha}{\partial t} = & (z-1) \left(k_1 \alpha(z, y, t) - \rho_1 \frac{\partial \alpha}{\partial z} \right) + \\ & + (y-1) \left(k_2 \alpha(z, y, t) - \rho_2 \frac{\partial \alpha}{\partial y} \right) - hz \frac{\partial \alpha}{\partial z} + f(\phi(z, y, t) - \alpha(z, y, t)) \end{aligned} \quad (1.5)$$

A equação 1.4 vem da soma das equações 1.1 e 1.2, e a equação 1.5 é a mesma que a 1.1

A distribuição de probabilidades $\phi_{n,m}(t)$ é reobtida da função geradora através da expressão:

$$\phi_{n,m}(t) = \frac{1}{n!m!} \left(\frac{\partial^{n+m} \phi}{\partial z^n \partial y^m} \right)_{z=0, y=0} \quad (1.6)$$

Os momentos da distribuição conjunta de probabilidades são dados por:

$$\langle n^i m^j \rangle(t) = \left(\left(z \frac{\partial}{\partial z} \right)^i \left(y \frac{\partial}{\partial y} \right)^j \phi(z, y, t) \right)_{z=1, y=1} \quad (1.7)$$

Devemos ter também:

$$\phi(z=1, y=1, t) = \sum_{n,m=0}^{\infty} \phi_{n,m}(t) = 1 \quad (1.8)$$

É importante notar que as funções geradoras, por definição (equação 1.3), são funções analíticas nas variáveis z e y . Como a obtenção das probabilidades e dos momentos se dão respectivamente nos pontos $(z, y) = (0, 0)$ e $(z, y) = (1, 1)$, precisamos impor que as soluções das equações diferenciais parciais 1.4 e 1.5 sejam analíticas no intervalo $0 \leq (z, y) \leq 1$. Tal requerimento é essencial para as soluções apresentadas nos próximos capítulos.

Outra propriedade importante que podemos obter é a probabilidade $P_\alpha(t)$ de encontrar o gene 2 ligado (sítio operador desocupado), e sua probabilidade complementar $P_\beta(t)$:

$$P_\alpha(t) = \alpha(z=1, y=1, t) = \sum_{n,m=0}^{\infty} \alpha_{n,m}(t) \quad (1.9)$$

$$P_\beta(t) = \beta(z = 1, y = 1, t) = \sum_{n,m=0}^{\infty} \beta_{n,m}(t) = 1 - P_\alpha(t) \quad (1.10)$$

2 *Soluções estacionárias*

Neste capítulo estamos interessados essencialmente no cálculo da distribuição conjunta de probabilidade $\phi_{n,m}$ de encontrar n proteínas do gene repressor (gene 1) e m proteínas do gene regulado (gene 2) na célula, no estado estacionário.

Primeiramente, apresentaremos a solução exata do problema através de uma prescrição para o cálculo de $\phi_{n,m}$. Em seguida, explorando resultados analíticos para os momentos de ordem mais baixa da distribuição, desenvolveremos uma solução aproximada e discutiremos seu regime de validade. Finalmente, apresentaremos uma proposta simples para a generalização destes resultados para a cadeia de três genes da referência (15).

2.1 Resultados exatos

2.1.1 Solução

A base para todas as discussões do capítulo são as equações diferenciais 1.4 e 1.5 para as funções geradoras das probabilidades de interesse (no estado estacionário $\phi(z, y, t \rightarrow \infty) \equiv \phi(z, y)$ e $\alpha(z, y, t \rightarrow \infty) \equiv \alpha(z, y)$).

O estado estacionário é obtido fazendo $\frac{\partial \phi}{\partial t} = 0$ e $\frac{\partial \alpha}{\partial t} = 0$ nas equações 1.4 e 1.5:

$$(z-1) \left(k_1 \phi(z, y) - \rho_1 \frac{\partial \phi}{\partial z} \right) + (y-1) \left(k_2(1-\chi)\alpha(z, y) + \chi k_2 \phi(z, y) - \rho_2 \frac{\partial \phi}{\partial y} \right) = 0 \quad (2.1)$$

$$(z-1) \left(k_1 \alpha(z, y) - \rho_1 \frac{\partial \alpha}{\partial z} \right) + (y-1) \left(k_2 \alpha(z, y) - \rho_2 \frac{\partial \alpha}{\partial y} \right) - h z \frac{\partial \alpha}{\partial z} + f(\phi(z, y) - \alpha(z, y)) = 0 \quad (2.2)$$

A equação 2.1 possui um ponto singular em $z = 1$ e outro em $y = 1$. Já a equação 2.2 possui singularidades em $z = \frac{\rho_1}{\rho_1+h}$ e $y = 1$. É mais conveniente utilizar uma variável $u = u(z)$ que desloque as singularidades para $u = 0$ e $u = 1$, pois assim colocamos as equações na forma canônica. Portanto, introduzindo os parâmetros $\varrho = \frac{\rho_1}{\rho_1+h}$ e $H = \frac{h}{\rho_1+h}$ (note que $\varrho + H = 1$), definimos $u = \frac{z-\varrho}{H}$.

A razão para esta mudança se baseia no fato de as funções geradoras serem definidas como funções analíticas em z e y (ambas variáveis entre 0 e 1). Como soluções regulares são mais facilmente obtidas e identificadas realizando-se expansões das equações diferenciais em torno dos pontos singulares, utilizamos a variável u , de modo que tais expansões expressem ϕ e α como séries de potências de u e v , ou $(u-1)$ e v .

Reescalando as taxas em unidades de $(\rho_1 + h)$ através da introdução dos parâmetros $K = \frac{k_1}{\rho_1+h}$, $\kappa = \frac{k_2}{\rho_1+h}$, $\rho = \frac{\rho_2}{\rho_1+h}$ e $F = \frac{f}{\rho_1+h}$, as equações 2.1 e 2.2 ficam:

$$\varrho(u-1) \frac{\partial \phi}{\partial u} + \rho v \frac{\partial \phi}{\partial v} - (KH(u-1) + \chi \kappa v) \phi(u, v) - \kappa v(1-\chi) \alpha(u, v) = 0 \quad (2.3)$$

$$u \frac{\partial \alpha}{\partial u} + \rho v \frac{\partial \alpha}{\partial v} - (KH(u-1) + \kappa v - F) \alpha(u, v) - F \phi(u, v) = 0 \quad (2.4)$$

Seguindo a explicação do parágrafo anterior, vamos começar expandindo as funções $\phi(u, v)$ e $\alpha(u, v)$ em potências de v :

$$\begin{aligned} \phi(u, v) &= \exp \left(KH(u-1) + \frac{\kappa}{\rho} v \right) \sum_{m=0}^{\infty} \phi_m(u) v^m \\ \alpha(u, v) &= \exp \left(KH(u-1) + \frac{\kappa}{\rho} v \right) \sum_{m=0}^{\infty} \alpha_m(u) v^m \end{aligned} \quad (2.5)$$

As equações 2.3 e 2.4 ficam:

$$\varrho(u-1) \frac{d\phi_m}{du} + (m\rho - KH^2(u-1)) \phi_m(u) + \kappa(1-\chi) (\phi_{m-1}(u) - \alpha_{m-1}(u)) = 0 \quad (2.6)$$

$$u \frac{d\alpha_m}{du} + (m\rho + F + KH) \alpha_m(u) - F\phi_m(u) = 0 \quad (2.7)$$

A exponencial em v nas transformações 2.5 foi colocada de modo que a equação 2.7 dependa apenas de $\alpha_m(u)$. Caso contrário, o termo $-\kappa v \alpha(u, v)$ de 2.4 produziria um termo proporcional a α_{m-1} na 2.7. A parte em u da exponencial assegura que a solução da equação 2.7 possa ser expressa de maneira simples como:

$$\alpha_m(u) = \frac{F}{u^{m\rho+F+KH}} \int \phi_m(u) u^{m\rho+F+KH-1} du + \frac{C_m}{u^{m\rho+F+KH}} \quad (2.8)$$

C_m são constantes arbitrárias advindas da solução da equação diferencial. Para que tenhamos analiticidade em $u = 0$, escolhemos $C_m = 0$.

Em princípio, a solução para o problema se dá da seguinte forma. Conhecendo ϕ_{m-1} e α_{m-1} , a equação 2.6 determina ϕ_m , que por sua vez dá α_m através da integral 2.8. Com estas funções em mãos, voltamos à equação 2.6 e determinamos ϕ_{m+1} . A inicialização destas relações de recorrência se dá com o cálculo de ϕ_0 , pois 2.6 com $m = 0$ é válida com a definição de que $\phi_{-1} = 0$ e $\alpha_{-1} = 0$.

Para efetivamente realizar os cálculos descritos acima, vamos introduzir na equação 2.6 (singular em $u = 1$) a expansão de $\phi_m(u)$ e $\alpha_m(u)$ em potências de $(u - 1)$, como sugerido nos dois parágrafos subsequentes à 2.2:

$$\begin{aligned} \phi_m(u) &= \sum_{n=0}^{\infty} T_{n,m} (u - 1)^n \\ \alpha_m(u) &= \sum_{n=0}^{\infty} A_{n,m} (u - 1)^n \end{aligned} \quad (2.9)$$

A equação 2.6 dá:

$$T_{n,m} = \frac{KH^2}{n\rho + m\rho} T_{n-1,m} - \frac{\kappa(1-\chi)}{n\rho + m\rho} (T_{n,m-1} - A_{n,m-1}) \quad (2.10)$$

A relação de recorrência 2.10 vale para n e m naturais, com a definição de que $T_{n,-1} = 0$ e $A_{n,-1} = 0$. Como $\phi(z = 1, y = 1) = 1$, temos que a 2.10 é iniciada com:

$$T_{0,0} = 1 \quad (2.11)$$

Para completar a solução resta encontrar uma expressão que relacione $A_{n,m}$ com os $T_{n,m}$. Como apontado anteriormente, a integral 2.8 dá esta prescrição. Substituindo a

expansão 2.9 em 2.8, e utilizando integral 15.3.1 da referência (27), temos:

$$\alpha_m(u) = \frac{F}{b_m} \sum_{n=0}^{\infty} (-1)^n T_{n,m} {}_2F_1(-n, b_m; 1 + b_m; u) \quad (2.12)$$

, com $b_m = m\rho + F + KH$, e ${}_2F_1$ a função hipergeométrica. Expandindo $\alpha_m(u)$ em potências de $(u - 1)$, pode-se mostrar que (demonstração feita no Apêndice A):

$$A_{n,m} = \sum_{k=n}^{\infty} T_{k,m} F \frac{(-1)^{k+n} k!}{n!} \frac{\Gamma(n + b_m)}{\Gamma(1 + k + b_m)} \quad (2.13)$$

Portanto, as relações de recorrência 2.10 e 2.13, juntamente com a condição 2.11, permitem a obtenção da distribuição completa de probabilidades $\phi_{n,m}$. Podemos também calcular o estado do sítio operador do gene 2 (a probabilidade de ele estar livre ou ocupado por uma proteína repressora).

Figura 2.1 – Distribuição conjunta de probabilidades $\phi_{n,m}$ para os seguintes valores de parâmetros: $K = 5, \varrho = 0.5, \kappa = 40, \rho = 20, F = 5, \chi = 0.1$

A figura 2.1 mostra um exemplo da distribuição conjunta $\phi_{n,m}$. Evidentemente, uma solução expressa na forma das equações 2.5, 2.9, 2.10 e 2.13 não nos dá nenhuma intuição de como o modelo se comporta em função dos parâmetros que o definem. Por esta razão, exploraremos nas seções seguintes resultados analíticos para os momentos de ordem

mais baixa da distribuição, e uma solução aproximada para o problema. Tais resultados nos permitirão extrair informações relevantes através de expressões relativamente simples sobre o comportamento de gráficos como da figura 2.1.

2.1.2 Distribuição de probabilidades para as proteínas repressoras

A distribuição de probabilidades de encontrar n proteínas repressoras na célula (gene 1), independente do número de proteínas do gene 2, é dada em termos de $\phi_{n,m}$ por:

$$\phi_n = \sum_{m=0}^{\infty} \phi_{n,m} \quad (2.14)$$

Em termos de $\phi(z, y)$, a função geradora da distribuição de probabilidades individual é dada por $\phi(z) = \phi(z, y = 1)$. Fazendo portanto $y = 1$ na 2.1 temos:

$$\frac{d\phi}{dz} = \frac{k_1}{\rho_1} \phi(z) \Rightarrow \phi(z) = \exp\left(\frac{k_1}{\rho_1}(z - 1)\right) \quad (2.15)$$

A constante arbitrária da solução da EDO foi escolhida de modo que $\phi(1) = 1$. Uma função geratriz exponencial gera uma distribuição de probabilidades Poissoniana:

$$\phi_n = \frac{(k_1/\rho_1)^n}{n!} \exp(-k_1/\rho_1) \quad (2.16)$$

A distribuição de Poisson possui apenas um pico, e portanto uma bimodalidade como a que aparece na figura 2.1 só é possível ao longo do eixo m . De fato, neste modelo o gene repressor apresenta apenas um nível de funcionamento, de modo que suas proteínas se estabilizam no estado estacionário próximas de um valor médio bem definido. Já o gene reprimido, por funcionar em dois níveis, pode apresentar uma distribuição estacionária sobre suas proteínas (m) que seja bimodal: um modo representando o estado em que o gene funciona parcialmente (ou inativo), e um modo de funcionamento pleno.

O valor médio dado pela distribuição de Poisson (em torno do qual todo gráfico de

$\phi_{n,m}$ de modo geral se distribui) é:

$$\langle n \rangle = \frac{k_1}{\rho_1} = \frac{K}{\varrho} \quad (2.17)$$

De fato, na figura 2.1 os dois picos estão aproximadamente centrados em torno de $n = 10$. A distribuição deste gráfico, quando reduzida apenas à variável n é apresentada na figura 2.2.

Figura 2.2 – Redução da distribuição conjunta $\phi_{n,m}$ de 2.1 para ϕ_n

O desvio padrão de σ_1 de ϕ_n escala com a média como:

$$\frac{\sigma_1}{\langle n \rangle} = \frac{1}{\sqrt{\langle n \rangle}} \quad (2.18)$$

2.1.3 Probabilidade de ocupação do gene 2

Além da informação sobre os números n e m das proteínas 1 e 2, a distribuição $\alpha_{n,m}$ contém informação sobre a probabilidade $P_\alpha [P_\beta]$ de encontrar o gene 2 ativo [inativo]. Como mostrado nas equações 1.9 e 1.10, temos:

$$P_\alpha = \alpha(z = 1, y = 1) \quad (2.19)$$

Ou, em termos das transformações 2.5:

$$P_\alpha = \alpha_0(u = 1) \quad (2.20)$$

Entretanto, a simples substituição de $u = 1$ e $m = 0$ nas equações 2.6 e 2.7 leva a:

$$\left(\frac{d\alpha_0}{du} \right)_{u=1} + (F + KH)P_\alpha = F \quad (2.21)$$

Através da equação acima, é impossível determinar P_α sem o conhecimento da derivada de $\alpha_0(u)$. Portanto, para calcular a probabilidade de ocupação é necessário resolver o seguinte sistema (equações 2.6 e 2.7):

$$\varrho \frac{d\phi_0}{du} - KH^2\phi_0(u) = 0 \quad (2.22)$$

$$u \frac{d\alpha_0}{du} + (F + KH)\alpha_0(u) - F\phi_0(u) = 0 \quad (2.23)$$

Isolando $\phi_0(u)$ na 2.23 e substituindo na 2.22, chegamos à seguinte equação de segunda ordem:

$$u \frac{d^2\alpha_0}{du^2} + \left(1 + F + KH - \frac{KH^2}{\varrho} u \right) \frac{d\alpha_0}{du} - \frac{KH^2}{\varrho} (F + KH)\alpha_0(u) = 0 \quad (2.24)$$

A equação 2.24 apresenta um ponto singular regular em $u = 0$ e um irregular em $u = \infty$, indicando que a solução pode ser expressa em termos de funções hipergeométricas confluentes. De fato, a solução é dada por:

$$\begin{aligned} \alpha_0(u) = & \exp\left(\frac{KH^2}{\varrho} u\right) \times \\ & \times \left(C_1 M\left(1, 1 + F + KH, -\frac{KH^2}{\varrho} u\right) + C_2 U\left(1, 1 + F + KH, -\frac{KH^2}{\varrho} u\right) \right) \end{aligned} \quad (2.25)$$

As funções M e U são as funções de Kummer (hipergeométricas confluentes) regular e irregular, respectivamente. Como procuramos por soluções analíticas, e a função U é irregular para $u = 0$, devemos ter $C_2 = 0$. A constante C_1 é determinada de modo que a equação 2.23 seja consistente com $\phi_0(u) = \exp\left(\frac{KH^2}{\varrho}(u-1)\right)$, levando a:

$$\alpha_0(u) = \frac{F}{F + KH} \exp\left(\frac{KH^2}{\varrho}(u-1)\right) M\left(1, 1 + F + KH, -\frac{KH^2}{\varrho}u\right) \quad (2.26)$$

Finalmente, a probabilidade de o sítio operador do gene 2 estar livre de proteínas repressoras é:

$$P_\alpha = \frac{F}{F + \langle n \rangle H \varrho} M\left(1, 1 + F + \langle n \rangle H \varrho, -\langle n \rangle H^2\right) \quad (2.27)$$

Portanto, o modelo é capaz de dizer como a concentração de proteínas repressoras presentes no meio age sobre o estado do sítio operador do gene reprimido.

Figura 2.3 – Comportamento da probabilidade P_α em função do número médio $\langle n \rangle$ de proteínas repressoras, para $H = 0.5$ e diferentes valores de F

Nas figuras 2.3 e 2.4, vemos como a repressão do gene 1 sobre o gene 2 depende dos dois parâmetros independentes F e H . F é o parâmetro ligado à taxa com que as proteínas repressoras se soltam do operador do segundo gene, e portanto vemos que conforme este parâmetro aumenta, a repressão tende a ser menos eficiente. O contrário ocorre com o parâmetro H , ligado à afinidade da proteína repressora com o gene 2. Conforme este parâmetro aumenta, o segundo gene é cada vez mais fortemente reprimido.

A expressão 2.27 para P_α consegue portanto reproduzir o chaveamento de um gene conforme a quantidade repressores no meio, usualmente modelado através de funções de

Figura 2.4 – Comportamento da probabilidade P_α em função do número médio $\langle n \rangle$ de proteínas repressoras, para $F = 1$ e diferentes valores de H

Hill.

2.1.4 Número médio de proteínas do gene 2

Devido à expressão 1.7, o número médio de proteínas do gene 2 é calculado por:

$$\langle m \rangle = \left(\frac{\partial \phi}{\partial y} \right)_{z=1, y=1} \quad (2.28)$$

Derivando a equação 2.1 em y e substituindo $z = 1$ e $y = 1$, temos:

$$\langle m \rangle = \frac{k_2}{\rho_2} (P_\alpha + \chi P_\beta) \quad (2.29)$$

ou:

$$\langle m \rangle = \frac{\kappa}{\rho} (P_\alpha + \chi P_\beta) \quad (2.30)$$

A expressão 2.29 para $\langle m \rangle$ é sob certo aspecto bastante intuitiva. Se o segundo gene funcionasse em apenas um estado (o estado ligado por exemplo), a distribuição de

probabilidades individual ϕ_m sobre seu número de proteínas seria de Poisson com média $\frac{k_2}{\rho_2}$, a exemplo do que foi discutido para ϕ_n na seção 2.1.2. Por outro lado, se o estado em que o gene funcionasse fosse apenas o reprimido, teríamos o mesmo tipo de distribuição mas com média $\chi \frac{k_2}{\rho_2}$. Portanto, o que a expressão 2.29 diz é que o número médio de proteínas produzidas pelo gene 2 é igual ao número médio de proteínas produzidas quando este gene se encontra ativado vezes a probabilidade de ele estar ativado, mais o número médio produzido quando o gene se encontra desligado vezes a probabilidade de ele estar desligado.

Esta discussão nos permite compreender mais um aspecto da distribuição conjunta $\phi_{n,m}$ apresentada na figura 2.1, ou na distribuição individual ϕ_m correspondente, apresentada na figura 2.5. Neste exemplo, temos $\frac{k_2}{\rho_2} = 20$, e $\chi = 0.1$, e os picos ao longo do eixo m estão localizados aproximadamente em torno de $m = 20$ e $m = 2$. O máximo ao redor de $m = 20$ é devido ao estado em que o gene 2 se encontra livre de proteínas repressoras, e o máximo ao redor de $m = 2$ representa o funcionamento reprimido deste gene.

Figura 2.5 – Redução da distribuição conjunta $\phi_{n,m}$ da figura 2.1 para ϕ_m . $\langle n \rangle = 10$, $\frac{k_2}{\rho_2} = 20$, $P_\alpha \approx 0.5$

A altura destes máximos depende dos valores de P_α e P_β . Para o caso das figuras 2.1 e 2.5 temos $P_\alpha = P_\beta = 0.5$. Já nas figuras 2.6 e 2.7 o máximo ao redor de $m = 20$ é bem mais intenso, pois $P_\alpha = 0.8$.

Figura 2.6 – Distribuição conjunta $\phi_{n,m}$ para $K = 9, \varrho = 0.9, \kappa = 40, \rho = 20, F = 5, \chi = 0.1$. Desta maneira, $\langle n \rangle = 10, \frac{k_2}{\rho_2} = 20$ e $P_\alpha \approx 0.8$

Figura 2.7 – Redução da distribuição conjunta $\phi_{n,m}$ da figura 2.6 para ϕ_m

Apesar do apelo intuitivo, a conclusão de que os números $\frac{k_2}{\rho_2}$ e $\chi \frac{k_2}{\rho_2}$ fixam as posições

dos picos não é válida sempre, pela seguinte razão. A produção de proteínas pelo gene reprimido envolve essencialmente dois processos estocásticos, com tempos característicos diferentes de chegada ao equilíbrio: o processo de 'liga/desliga' (switching) de proteínas repressoras no sítio operador (tempo característico T_{op}), e o processo de produção de proteínas em um determinado estado (tempo característico T_2). Se $T_2 \ll T_{op}$, o processo de switching é muito mais lento que a estabilização no número de proteínas. Desta maneira, se o gene se encontra no estado livre [reprimido], há tempo suficiente para que a distribuição de proteínas se estabilize em torno do valor $m = \frac{k_2}{\rho_2}$ $\left[m = \chi \frac{k_2}{\rho_2} \right]$ antes que o gene desligue [ligue]. Nestas condições veremos uma estrutura bimodal na distribuição de probabilidades.

Caso contrário, se $T_2 \gg T_{op}$, o processo de switching é muito mais rápido, de modo que a produção de proteínas se passa como se o gene funcionasse em um estado intermediário dado pela média dos dois estados originais. Neste caso, observamos uma estrutura na distribuição de probabilidades de apenas um pico, e a distinção entre dois estados de certa forma perde o sentido. Esta condição é denominada na literatura de limite adiabático, e um exemplo se encontra na figura 2.8.

Neste capítulo apresentaremos aproximações para T_2 e T_{op} , e no Capítulo 4 mostraremos uma estimativa mais rigorosa destes parâmetros.

2.1.5 Correlação

Através das equações 2.1 e 2.2, é possível obter uma expressão simples para a correlação entre os números n e m de proteínas dos genes 1 e 2, respectivamente. Definiremos aqui a correlação por:

$$Corr(n, m) = \langle nm \rangle - \langle n \rangle \langle m \rangle \quad (2.31)$$

A partir da função geradora (expressão 1.7), temos:

$$\langle nm \rangle = \left(\frac{\partial^2 \phi}{\partial z \partial y} \right)_{z=1, y=1} \quad (2.32)$$

Figura 2.8 – Distribuição conjunta $\phi_{n,m}$ para $K = 5, \varrho = 0.5, \kappa = 20, \rho = 1, F = 5, \chi = 0.1$. Desta maneira, $\langle n \rangle = 10, \frac{k_2}{\rho_2} = 20$ e $P_\alpha \approx 0.5$

Derivando a equação 2.1 em z e y e substituindo $z = 1$ e $y = 1$ nesta e na equação 2.2, obtém-se:

$$(\rho_1 + \rho_2) \langle nm \rangle = k_1 \langle m \rangle + k_2(1 - \chi) \left(\frac{\partial \alpha}{\partial z} \right)_{z=1, y=1} + \chi k_2 \langle n \rangle \quad (2.33)$$

$$\left(\frac{\partial \alpha}{\partial z} \right)_{z=1, y=1} = \frac{f}{h} P_\beta \quad (2.34)$$

Resolvendo as equações 2.33 e 2.34 para $\langle nm \rangle$, chegamos a:

$$Corr(n, m) = -(1 - \chi) \frac{\kappa}{\varrho + \rho} \left(\frac{K}{\varrho} P_\alpha - \frac{F}{H} P_\beta \right) \quad (2.35)$$

Quando duas variáveis estocásticas são independentes, a correlação entre elas é nula (a recíproca não é necessariamente verdadeira). Portanto, é interessante analisarmos a região do espaço de parâmetros que leva a $Corr(n, m) = 0$. Através da equação 2.35, vemos que isto ocorre quando:

$$P_\alpha = \frac{F}{F + \frac{KH}{\varrho}} \quad (2.36)$$

, ou equivalentemente:

$$\frac{1}{F + KH} M \left(1, 1 + F + KH, -\frac{KH^2}{\varrho} \right) = \frac{1}{F + KH + \frac{KH^2}{\varrho}} \quad (2.37)$$

A equação 2.37 depende apenas dos dois parâmetros $b_0 = F + KH$ e $a = \frac{KH^2}{\varrho}$, de modo que trabalharemos com a função:

$$g(b_0, a) = \frac{1}{b_0} M(1, 1 + b_0, -a) - \frac{1}{b_0 + a} \quad (2.38)$$

Figura 2.9 – Curvas de nível da função $g(b_0, a)$. Na região amarela, alaranjada e vermelha temos respectivamente $g > 0.5$, $0.1 < g < 0.5$ e $g < 0.1$. Para $b_0 \rightarrow 0$ e $a \neq 0$, $g \rightarrow \infty$

A função g se anula junto com a correlação, e ambas têm sinais opostos entre si. A figura 2.9 mostra a função $g(b_0, a)$ através de suas curvas de nível. Vemos da figura que a correlação se anula conforme b_0 ou a crescem, e que g cresce indefinidamente conforme b_0 se aproxima de zero, para $a \neq 0$. Nota-se que g é indefinidamente grande em uma região relativamente pequena do espaço de parâmetros, e portanto é importante estudar os valores de b_0 e a em termos dos parâmetros F , $\langle n \rangle$ e H (ou ϱ , uma vez que $H + \varrho = 1$), para entender o que as regiões da figura 2.9 representam, e quais regiões possuem interesse biológico.

Tendo em vista o comentário acima, a figura 2.10 mostra as curvas paramétricas $c : (a(H), b_0(H))$, para F fixo e diferentes valores de $\langle n \rangle$, sobrepostas ao gráfico da figura

Figura 2.10 – Gráficos paramétricos das curvas dadas por $(a(H), b_0(H))$, com H indo de 0 a 1, para $F = 0.1$ e diferentes valores de $\langle n \rangle$

2.9.

Vemos da figura 2.10 que a correlação entre n e m parece ser importante apenas quando o número médio de repressores é pequeno ($\langle n \rangle \approx 1$), e que, conforme este número cresce, a correlação entre as duas proteínas se aproxima rapidamente de zero. Aumentar o parâmetro F implica apenas em translações positivas das curvas ao longo do eixo b_0 . Para sistemas biológicos, em geral observamos números típicos de proteínas da ordem de centenas, ou até milhares. Portanto, se neste modelo n for tratado como uma proteína, e não como um RNA mensageiro ou algum outro fator de transcrição, pode ser válida para sistemas reais uma aproximação de independência estatística. Na seção seguinte desenvolveremos uma aproximação neste sentido baseada na equação 2.37.

Momentos de maior ordem nas proteínas do gene 2 não podem ser obtidos da mesma maneira simples com a qual foi obtida a média e a covariância, pois estes necessitam das soluções 2.10 e 2.13 das equações 2.6 e 2.7.

2.2 Resultados aproximados

2.2.1 Funções geradoras

Nas seções anteriores escrevemos as funções geradoras $\phi(z, y)$ e $\alpha(z, y)$ em termos das variáveis $u = \frac{z-\varrho}{H}$ e $v = y - 1$, na forma:

$$\begin{aligned}\phi(u, v) &= \exp\left(KH(u-1) + \frac{\kappa}{\rho}v\right) \sum_{m=0}^{\infty} \phi_m(u) v^m \\ \alpha(u, v) &= \exp\left(KH(u-1) + \frac{\kappa}{\rho}v\right) \sum_{m=0}^{\infty} \alpha_m(u) v^m\end{aligned}\quad (2.39)$$

, onde os $\phi_m(u)$ e $\alpha_m(u)$ são soluções de:

$$\varrho(u-1)\frac{d\phi_m}{du} + (m\rho - KH^2(u-1))\phi_m(u) + \kappa(1-\chi)(\phi_{m-1}(u) - \alpha_{m-1}(u)) = 0 \quad (2.40)$$

$$u\frac{d\alpha_m}{du} + (m\rho + F + KH)\alpha_m(u) - F\phi_m(u) = 0 \quad (2.41)$$

As equações 2.40 e 2.41 mais a condição de que as funções geradoras sejam analíticas levam a (como em 2.8):

$$\phi_m(u) = (\chi - 1)\frac{\kappa}{\varrho}\frac{\exp\left(\frac{KH^2}{\varrho}u\right)}{(u-1)^{m\frac{\rho}{\varrho}}} \int (u-1)^{(m\frac{\rho}{\varrho}-1)} \exp\left(-\frac{KH^2}{\varrho}u\right) (\phi_{m-1}(u) - \alpha_{m-1}(u)) du \quad (2.42)$$

$$\alpha_m(u) = \frac{F}{u^{m\rho+F+KH}} \int \phi_m(u) u^{m\rho+F+KH-1} du \quad (2.43)$$

Para $m = 0$, tínhamos (2.15 e 2.26):

$$\phi_0(u) = \exp\left(\frac{KH^2}{\varrho}(u-1)\right) \quad (2.44)$$

$$\alpha_0(u) = \frac{F}{F + KH} \exp\left(\frac{KH^2}{\varrho}(u - 1)\right) M\left(1, 1 + F + KH, -\frac{KH^2}{\varrho}u\right) \quad (2.45)$$

Na seção anterior vimos que na região do espaço de parâmetros biologicamente relevante, vale de maneira aproximada a equação 2.37, e a correlação é aproximadamente nula. Assim é razoável buscar por uma função geradora aproximada $\phi^{ap}(u, v) = U_\phi(u)V_\phi(v)$, separável nas variáveis u e v (ou z e y), de modo que n e m sejam variáveis estocásticas estatisticamente independentes. Tal objetivo é atingido aplicando-se a aproximação 2.37 à expressão 2.45 para $\alpha_0(u)$ da seguinte forma:

$$\alpha_0^{ap}(u) = \frac{F}{F + KH + \frac{KH^2}{\varrho}} \exp\left(\frac{KH^2}{\varrho}(u - 1)\right) \quad (2.46)$$

Comparando a aproximação 2.37 com $\alpha_0(u)$ de 2.45, vemos que o fator que acompanha a exponencial em α_0^{ap} deveria ser a rigor dado por $\left(\frac{F}{F + KH + \frac{KH^2}{\varrho}u}\right)$. Entretanto, esta forma não leva à independência estatística. Mesmo que o $\alpha_0^{ap}(u)$ dado em 2.46 possa a princípio ser válido apenas para $u \approx 1$, veremos na seção seguinte que as funções geradoras derivadas desta aproximação conseguem reproduzir os resultados exatos com bastante sucesso. No Capítulo 3 descrevemos uma maneira alternativa rigorosa de se obter as soluções aproximadas apresentadas aqui.

Substituindo o $\alpha_0^{ap}(u)$ de 2.46 em 2.42 (com $m = 1$), é fácil ver que:

$$\phi_1^{ap}(u) \propto \exp\left(\frac{KH^2}{\varrho}(u - 1)\right) \quad (2.47)$$

Substituindo este na 2.43 (para $m = 1$), vemos que (em analogia com a 2.45):

$$\alpha_1^{ap}(u) \propto \exp\left(\frac{KH^2}{\varrho}(u - 1)\right) \frac{F}{\rho + F + KH} M\left(1, 1 + \rho + F + KH, -\frac{KH^2}{\varrho}u\right) \quad (2.48)$$

Na discussão a respeito da figura 2.9, vimos que quanto maior for o valor de b_0 , mais próxima de zero está a função $g(b_0, a)$. Portanto, podemos utilizar outra vez a aproximação 2.37 fazendo:

$$\frac{F}{\rho + F + KH} M\left(1, 1 + \rho + F + KH, -\frac{KH^2}{\varrho}u\right) \rightarrow \frac{F}{\rho + F + KH + \frac{KH^2}{\varrho}}$$

Procedendo da mesma maneira para $m = 2, 3, \dots$ nas equações 2.42 e 2.43, obtemos as seguintes funções geradoras aproximadas, em termos dos parâmetros originais do modelo

(demonstração feita no Apêndice B):

$$\phi^{ap}(z, y) = e^{\left(\frac{k_1}{\rho_1}(z-1) + \frac{k_2}{\rho_2}(y-1)\right)} M\left(\frac{k_1}{\rho_1} \frac{h}{\rho_2}, \frac{f + h \frac{k_1}{\rho_1}}{\rho_2}; (\chi - 1) \frac{k_2}{\rho_2}(y - 1)\right) \quad (2.49)$$

$$\alpha^{ap}(z, y) = e^{\left(\frac{k_1}{\rho_1}(z-1) + \frac{k_2}{\rho_2}(y-1)\right)} \frac{f}{f + h \frac{k_1}{\rho_1}} M\left(\frac{k_1}{\rho_1} \frac{h}{\rho_2}, \frac{f + h \frac{k_1}{\rho_1}}{\rho_2} + 1; (\chi - 1) \frac{k_2}{\rho_2}(y - 1)\right) \quad (2.50)$$

$$\beta^{ap}(z, y) = e^{\left(\frac{k_1}{\rho_1}(z-1) + \frac{k_2}{\rho_2}(y-1)\right)} \frac{h \frac{k_1}{\rho_1}}{f + h \frac{k_1}{\rho_1}} M\left(\frac{k_1}{\rho_1} \frac{h}{\rho_2} + 1, \frac{f + h \frac{k_1}{\rho_1}}{\rho_2} + 1; (\chi - 1) \frac{k_2}{\rho_2}(y - 1)\right) \quad (2.51)$$

As figuras 2.11 a 2.18 comparam a distribuição conjunta de probabilidades $\phi_{n,m}$ obtida via os cálculos exatos da seção 2.1.1, com a obtida via a aproximação 2.49.

Nas figuras de 2.11 a 2.14, os valores de a e b_0 foram escolhidos de modo a estarmos na região do espaço de parâmetros em que a correlação se aproxima de zero (região vermelha da figura 2.9, onde espera-se que valha a equação 2.49). De fato, vemos nestes gráficos que a solução aproximada reproduz muito bem a exata.

As figuras 2.15 e 2.16 correspondem a valores de a e b_0 localizados na região alaranjada da figura 2.9, e nas figuras 2.17 e 2.18 a e b_0 se encontram na região amarela. Como era de se esperar, a solução aproximada começa a se desviar da exata nestas figuras, especialmente nas duas últimas. Ainda assim, observamos uma concordância qualitativa entre as soluções.

2.2.2 Propriedades da solução aproximada e estimativa dos tempos de ida ao equilíbrio

Como a função geradora 2.49 é separável nas variáveis z e y , a probabilidade conjunta $\phi_{n,m}$ é dada por um produto de probabilidades independentes em n e m , sendo que a distribuição sobre n é Poissoniana com média $\frac{k_1}{\rho_1}$, como discutido na seção 2.1.2. Por-

Figura 2.11 – Comparação entre os cálculos exatos e aproximados para a distribuição conjunta $\phi_{n,m}$. $\langle n \rangle = 10, \kappa = 40, \rho = 20, \chi = 0.1, P_\alpha = 0.43, a = 4.9, b_0 = 7.1$

Figura 2.12 – Comparação entre os cálculos exatos e aproximados para a distribuição conjunta $\phi_{n,m}$. $\langle n \rangle = 10, \kappa = 20, \rho = 1, \chi = 0.1, P_\alpha = 0.43, a = 4.9, b_0 = 7.1$

Figura 2.13 – Comparação entre os cálculos exatos e aproximados para a distribuição conjunta $\phi_{n,m}$. $\langle n \rangle = 10, \kappa = 20, \rho = 1, \chi = 0.1, P_\alpha = 0.63, a = 0.9, b_0 = 7.1$

Figura 2.14 – Comparação entre os cálculos exatos e aproximados para a distribuição conjunta $\phi_{n,m}$. $\langle n \rangle = 10, \kappa = 40, \rho = 20, \chi = 0.1, P_\alpha = 0.63, a = 0.9, b_0 = 7.1$

Figura 2.15 – Comparação entre os cálculos exatos e aproximados para a distribuição conjunta $\phi_{n,m}$. $\langle n \rangle = 3, \kappa = 2, \rho = 0.1, \chi = 0.1, P_\alpha = 0.20, a = 0.75, b_0 = 1.05$

Figura 2.16 – Comparação entre os cálculos exatos e aproximados para a distribuição conjunta $\phi_{n,m}$. $\langle n \rangle = 3, \kappa = 100, \rho = 5, \chi = 0.1, P_\alpha = 0.20, a = 0.75, b_0 = 1.05$

Figura 2.17 – Comparação entre os cálculos exatos e aproximados para a distribuição conjunta $\phi_{n,m}$. $\langle n \rangle = 0.8, \kappa = 2, \rho = 0.1, \chi = 0.1, P_\alpha = 0.35, a = 0.07, b_0 = 0.27$

Figura 2.18 – Comparação entre os cálculos exatos e aproximados para a distribuição conjunta $\phi_{n,m}$. $\langle n \rangle = 0.8, \kappa = 100, \rho = 5, \chi = 0.1, P_\alpha = 0.35, a = 0.07, b_0 = 0.27$

tanto, vamos discutir nesta seção apenas as funções geradoras das probabilidades sobre m , obtidas fazendo $z = 1$ nas 2.49, 2.50 e 2.51:

$$\phi^{ap}(y) = \exp\left(\frac{k_2}{\rho_2}(y-1)\right) M\left(\frac{k_1}{\rho_1} \frac{h}{\rho_2}, \frac{f + h \frac{k_1}{\rho_1}}{\rho_2}; (\chi-1) \frac{k_2}{\rho_2}(y-1)\right) \quad (2.52)$$

$$\alpha^{ap}(y) = \exp\left(\frac{k_2}{\rho_2}(y-1)\right) \frac{f}{f + h \frac{k_1}{\rho_1}} M\left(\frac{k_1}{\rho_1} \frac{h}{\rho_2}, \frac{f + h \frac{k_1}{\rho_1}}{\rho_2} + 1; (\chi-1) \frac{k_2}{\rho_2}(y-1)\right) \quad (2.53)$$

$$\beta^{ap}(y) = \exp\left(\frac{k_2}{\rho_2}(y-1)\right) \frac{h \frac{k_1}{\rho_1}}{f + h \frac{k_1}{\rho_1}} M\left(\frac{k_1}{\rho_1} \frac{h}{\rho_2} + 1, \frac{f + h \frac{k_1}{\rho_1}}{\rho_2} + 1; (\chi-1) \frac{k_2}{\rho_2}(y-1)\right) \quad (2.54)$$

Na referência (24), é apresentado um modelo estocástico para um único gene que pode funcionar em dois estados (como o gene 2 discutido aqui). A diferença é que lá a taxa com que uma proteína repressora se liga ao sítio operador, desligando o gene, é dada por uma constante h , ao invés de depender explicitamente do número de proteínas repressoras. Em outras palavras, em (24) a presença das proteínas repressoras é tratada como um efeito médio. As funções geradoras obtidas naquela referência são dadas por:

$$\bar{\phi}(y) = \exp\left(\frac{k_2}{\rho_2}(y-1)\right) M\left(\frac{h}{\rho_2}, \frac{f+h}{\rho_2}; (\chi-1) \frac{k_2}{\rho_2}(y-1)\right) \quad (2.55)$$

$$\bar{\alpha}(y) = \exp\left(\frac{k_2}{\rho_2}(y-1)\right) \frac{f}{f+h} M\left(\frac{h}{\rho_2}, \frac{f+h}{\rho_2} + 1; (\chi-1) \frac{k_2}{\rho_2}(y-1)\right) \quad (2.56)$$

$$\bar{\beta}(y) = \exp\left(\frac{k_2}{\rho_2}(y-1)\right) \frac{h}{f+h} M\left(\frac{h}{\rho_2} + 1, \frac{f+h}{\rho_2} + 1; (\chi-1) \frac{k_2}{\rho_2}(y-1)\right) \quad (2.57)$$

Vemos que a solução apresentada aqui é obtida de (24) fazendo na última $h \rightarrow h \langle n \rangle$. Portanto, uma vantagem obtida com a aproximação de independência estatística é que esta permitiu escrevermos as funções geradoras, e portanto as probabilidades, de maneira relativamente simples, e que conseguimos determinar exatamente como se dá a ação do gene repressor sobre o gene reprimido.

As probabilidades de ocupação do gene 2 são dadas pelas funções de Hill abaixo:

$$\begin{aligned} P_{\alpha}^{ap} &= \frac{f}{f + h \langle n \rangle} \\ P_{\beta}^{ap} &= \frac{h \langle n \rangle}{f + h \langle n \rangle} \end{aligned} \quad (2.58)$$

O número médio de proteínas do gene reprimido mantém a mesma forma que a da equação 2.29:

$$\langle m \rangle^{ap} = \frac{k_2}{\rho_2} (P_{\alpha}^{ap} + \chi P_{\beta}^{ap}) \quad (2.59)$$

Na seção 2.1.4 introduzimos o conceito do tempo de ida ao equilíbrio T_{op} do sítio operador do gene 2. Na referência (24) este tempo é calculado, e transpondo o resultado para a aproximação aqui apresentada (ou seja, fazendo naquela referência $h \rightarrow h \langle n \rangle$), temos:

$$T_{op} = \frac{1}{f + h \langle n \rangle} \quad (2.60)$$

Figura 2.19 – Efeito do parâmetro de adiabaticidade ε sobre a distribuição de proteínas do gene 2, na solução aproximada. Nas três curvas, $\frac{k_2}{\rho_2} = 30$, $\chi = 0.2$ e $P_{\alpha}^{ap} = 0.6$

Assumindo que o tempo de ida ao equilíbrio do processo de produção de proteínas em um determinado estado do gene 2 seja dado por $T_2 = \frac{1}{\rho_2}$, podemos definir o parâmetro de

adiabaticidade:

$$\varepsilon = \frac{T_2}{T_{op}} = \frac{f + h \langle n \rangle}{\rho_2} \quad (2.61)$$

Como discutido na seção 2.1.4, se $\varepsilon < 1$, há tempo para que o número de proteínas do gene 2 se estabilize enquanto este se encontra em um de seus dois estados. Nesta condição observamos uma distribuição de probabilidades bimodal, o que torna indispensável o conhecimento de toda a distribuição de probabilidades. Por outro lado, se $\varepsilon > 1$, observamos uma distribuição de apenas um pico. Este é o chamado limite adiabático, em que a distribuição de probabilidades pode ser estimada apenas através da sua média e variância. A figura 2.19 mostra o comportamento da solução aproximada em função do parâmetro de adiabaticidade ε .

Consistentemente com o discutido acima, os valores de ε das figuras 2.12, 2.14, 2.16 e 2.18 são respectivamente 0.6, 0.4, 0.36 e 0.068. Os valores de ε das figuras 2.11, 2.13, 2.15 e 2.17 são respectivamente 12, 8, 18 e 3.4.

A transição para o regime adiabático pode ser vista também através da expressão para a variância no número de proteínas do gene 2. Da equação 2.49, temos:

$$(\sigma_2^2)^{ap} = \langle m \rangle^{ap} + (1 - \chi)^2 \left(\frac{k_2}{\rho_2} \right)^2 \frac{P_\alpha^{ap} P_\beta^{ap}}{1 + \varepsilon} \quad (2.62)$$

Conforme ε cresce, $(\sigma_2^2)^{ap} \rightarrow \langle m \rangle^{ap}$, e a distribuição se aproxima de uma distribuição de Poisson.

2.3 Expressão em cascata

Na referência (15) é descrito um experimento realizado em uma rede genética sintética, montada em um plasmídeo de *Escherichia coli*. Esta rede é composta de quatro genes, que aqui indexaremos de 0 a 3. O gene 0 é transcrito constitutivamente (funciona em apenas um estado) produzindo a proteína LacI. Esta reprime o gene 1, que produz proteínas cujo número é reportado por proteínas fluorescentes de cor ciano (CFP - cyan fluorescent proteins). O produto do gene 1 reprime a produção de proteínas fluorescentes amarelas

(YFP - yellow fluorescent proteins) produzidas pelo gene 2. O gene 3 é independente do resto da cascata e serve como controle, produzindo proteínas fluorescentes vermelhas (RFP - red fluorescent proteins).

A quantidade de proteínas do gene 0 (LacI) é controlada através do indutor *isopropyl-β-D-thiogalactopyranoside* (IPTG). Quando o IPTG é adicionado ao meio, ele se liga às proteínas do gene 0, inutilizando-as. Portanto, o IPTG funciona como um indutor para o gene 1 e através de sua concentração no meio, controla-se as quantidades de proteínas dos genes 0, 1 e 2.

Estamos interessados aqui em aplicar as ideias desenvolvidas nas seções anteriores para reproduzir os valores médios e ruídos das proteínas dos genes 1 e 2, e o coeficiente de correlação entre as proteínas 1 e 2, obtidos experimentalmente em (15). Como cada um destes dois genes funciona em dois estados, precisaremos modificar as expressões 2.50 e 2.51 para a função geradora das probabilidades conjuntas das proteínas 1 e 2. Para tal, definimos:

$$\alpha_i(\lambda_i, x_i) = \frac{f_i}{f_i + \lambda_i} \exp\left(\frac{k_i}{\rho_i}(x_i - 1)\right) M\left(\frac{\lambda_i}{\rho_i}, \frac{f_i + \lambda_i}{\rho_i} + 1, \frac{k_i}{\rho_i}(\chi_i - 1)(x_i - 1)\right) \quad (2.63)$$

$$\beta_i(\lambda_i, x_i) = \frac{\lambda_i}{f_i + \lambda_i} \exp\left(\frac{k_i}{\rho_i}(x_i - 1)\right) M\left(\frac{\lambda_i}{\rho_i} + 1, \frac{f_i + \lambda_i}{\rho_i} + 1, \frac{k_i}{\rho_i}(\chi_i - 1)(x_i - 1)\right) \quad (2.64)$$

As taxas indexadas por i nas expressões 2.63 e 2.64 podem se referir às proteínas do tipo 1 ou 2, conforme tenhamos $i = 1$ ou $i = 2$.

O sistema em questão possui quatro estados possíveis:

- 1- os genes 1 e 2 se encontram ativos $\rightarrow$ função geradora Ψ_1
- 2- o gene 1 se encontra ativo e o gene 2 reprimido $\rightarrow$ função geradora Ψ_2
- 3- o gene 1 se encontra reprimido e o gene 2 ativo $\rightarrow$ função geradora Ψ_3
- 4- os genes 1 e 2 se encontram reprimidos $\rightarrow$ função geradora Ψ_4

Vimos na seção 2.2.1 que se o gene 1 funcionasse apenas no estado ligado, a função geradora do gene 2 teria $\lambda_2 = h_2 \frac{k_1}{\rho_1}$ (equações 2.50 e 2.51). Analogamente, se o gene 1 funcionasse apenas no estado desligado, teríamos para o gene 2 $\lambda_2 = h_2 \chi_1 \frac{k_1}{\rho_1}$. Portanto,

um *Ansatz* razoável para as funções geradoras deste sistema é:

$$\begin{aligned}
 \Psi_1(x_0, x_1, x_2) &= \exp\left(\frac{k_0}{\rho_0}(x_0 - 1)\right) \alpha_1\left(h_1 \frac{k_0}{\rho_0}, x_1\right) \alpha_2\left(h_2 \frac{k_1}{\rho_1}, x_2\right) \\
 \Psi_2(x_0, x_1, x_2) &= \exp\left(\frac{k_0}{\rho_0}(x_0 - 1)\right) \alpha_1\left(h_1 \frac{k_0}{\rho_0}, x_1\right) \beta_2\left(h_2 \frac{k_1}{\rho_1}, x_2\right) \\
 \Psi_3(x_0, x_1, x_2) &= \exp\left(\frac{k_0}{\rho_0}(x_0 - 1)\right) \beta_1\left(h_1 \frac{k_0}{\rho_0}, x_1\right) \alpha_2\left(h_2 \chi_1 \frac{k_1}{\rho_1}, x_2\right) \\
 \Psi_4(x_0, x_1, x_2) &= \exp\left(\frac{k_0}{\rho_0}(x_0 - 1)\right) \beta_1\left(h_1 \frac{k_0}{\rho_0}, x_1\right) \beta_2\left(h_2 \chi_1 \frac{k_1}{\rho_1}, x_2\right)
 \end{aligned} \tag{2.65}$$

As exponenciais em x_0 nas equações 2.65 se referem ao gene 0. Como este funciona em apenas um estado, apresenta a distribuição de probabilidades poissoniana (seção 2.1.2).

Na figuras 1 e 2 da referência (15) estão apresentados os valores médios e ruídos das proteínas 1 e 2, e o coeficiente de correlação entre elas como função da concentração de IPTG. O ruído da proteína i é definido como a razão entre o desvio padrão e o valor médio ($\eta_i = \sigma_i / \langle n_i \rangle$), e o coeficiente de correlação como $C_{12} = \frac{\langle n_1 n_2 \rangle - \langle n_1 \rangle \langle n_2 \rangle}{\langle n_1 \rangle \langle n_2 \rangle}$.

Quando o IPTG está presente no meio, ele inutiliza as proteínas do gene 0, de modo que seu número médio efetivo $\langle n_0 \rangle = \frac{k_0}{\rho_0}$ que participam da cascata é em princípio função da concentração de IPTG. Vamos assumir que a dependência funcional seja da forma (similarmente ao que é sugerido na referência (15)):

$$\langle n_0 \rangle = \frac{100}{1 + \left(\frac{[IPTG]}{0.01}\right)^4} \tag{2.66}$$

A relação 2.66 explicita o fato do número efetivo de proteínas do tipo 0 participantes da cadeia ser reduzido conforme adição de IPTG na cultura.

Calculando os momentos necessários através da expressão 1.7 aplicada à função geradora $\Phi = \Psi_1 + \Psi_2 + \Psi_3 + \Psi_4$, obtemos as expressões para as médias, ruídos e coeficiente de correlação graficadas nas figuras de 2.20 a 2.24. Os parâmetros utilizados no cálculo foram: $k_1 = 4000, \rho_1 = 1, h_1 = 1, f_1 = 1, k_2 = 2500, \rho_2 = 1, h_2 = 0.01, f_2 = 5, \chi_1 = \chi_2 = 0$.

As figuras de 2.20 a 2.24 devem ser comparadas com as figuras 1 e 2 da referência (15). Todos os gráficos concordam de maneira qualitativa com o apresentado no experimento. Os ruídos calculados aqui diferem dos experimentais apresentando em algumas regiões valores menores. De fato, nosso modelo levou em conta apenas flutuações no número de proteínas. Outras fontes de ruído entretanto também estão presentes, como flutuações na disponibilidade de RNA polimerases no início da transcrição (exemplo de ruído extrínscico

Figura 2.20 – Comportamento da média no número de proteínas do tipo 1 como função da concentração de IPTG

Figura 2.21 – Comportamento do ruído no número de proteínas do tipo 1 como função da concentração de IPTG

(28)). O coeficiente de correlação apresenta um bom acordo com o experimento.

É importante notar que a forma para o número médio de proteínas do gene 0 escolhida na equação 2.66 determina diretamente o valor médio de proteínas do gene 1. Desta maneira, $\langle n_1 \rangle$ foi escolhido de modo a concordar com o experimento. Os outros momentos entretanto, são consequência da forma da função geradora, e portanto predições do

Figura 2.22 – Comportamento da média no número de proteínas do tipo 2 como função da concentração de IPTG

Figura 2.23 – Comportamento do ruído no número de proteínas do tipo 2 como função da concentração de IPTG

modelo.

As funções geradoras Ψ_1, Ψ_2, Ψ_3 e Ψ_4 foram propostas a partir do conhecimento adquirido no estudo da rede mais simples com apenas dois genes das seções anteriores, e obteve grande sucesso na reprodução de experimentos. Isto demonstra o poder das soluções analíticas, aproximadas ou não, para a construção de redes mais realistas, e dá

Figura 2.24 – Coeficiente de correlação entre as proteínas 1 e 2 como função da concentração de IPTG

indícios de como esta generalização pode ser feita. O Capítulo 3 pretende introduzir a um método variacional que pode contribuir para esta tarefa de generalização de maneira rigorosa de soluções simples para redes mais complicadas.

3 *Método variacional*

No capítulo 2 introduzimos uma maneira de generalizar resultados de uma rede simples para uma cascata de genes. Entretanto, utilizamos para a obtenção das funções geradoras 2.65 apenas a intuição advinda da física e biologia do sistema. Neste capítulo descreveremos brevemente um método aproximado (método variacional) (20, 29–31) para a solução das equações 2.1 e 2.2, que recupera as equações 2.49, 2.50 e 2.51 de uma forma mais rigorosa e mais simples do que a utilizada no Capítulo 2.

3.1 Descrição do método

Para fixar as ideias, vamos introduzir o método variacional através de um exemplo. Suponha um processo markoviano de nascimento e morte, onde o sistema vai de um estado n para um estado $n + 1$ com taxa k , e de n para $n - 1$ com taxa $n\rho$. n pode ser o número de proteínas em uma célula, o número de bactérias em um meio de cultura, a posição de um bêbado em algum tipo de passeio aleatório, etc. A equação mestra que rege a probabilidade $p_n(t)$ de encontrar o sistema em n é:

$$\frac{dp_n}{dt} = kp_{n-1}(t) - (k + n\rho)p_n(t) + (n + 1)\rho p_{n+1}(t) \quad (3.1)$$

Procedendo como feito nos capítulos anteriores, introduzimos a função geradora:

$$P(z, t) = \sum_{n=0}^{\infty} p_n(t) z^n \quad (3.2)$$

, e a equação 3.1 fica:

$$\frac{\partial P}{\partial t} = (z - 1) \left(kP(z, t) - \rho \frac{\partial P}{\partial z} \right) \quad (3.3)$$

A solução para o estado estacionário é, como feito na seção 2.1.2:

$$P(z) = \exp \left(\frac{k}{\rho} (z - 1) \right) \quad (3.4)$$

Para desenvolver o método variacional, vamos introduzir uma representação mais geral que a de funções geradoras, escrevendo a expressão 3.2 como:

$$| P(t) \rangle = \sum_{n=0}^{\infty} p_n(t) | n \rangle \quad (3.5)$$

A expressão 3.5 representa o estado do sistema de uma forma abstrata, mas essencialmente pode ser encarada como uma mudança de notação da equação 3.2. Para escrever a equação mestra, precisamos definir operadores que levem em conta os termos p_{n-1} e p_{n+1} de 3.1. Para tal, definimos a e $a^\dagger$ de modo que:

$$\begin{aligned} a | n \rangle &= n | n - 1 \rangle \\ a^\dagger | n \rangle &= | n + 1 \rangle \end{aligned} \quad (3.6)$$

Como consequência das 3.6, temos (20, 29):

$$\begin{aligned} [a, a^\dagger] &= 1 \\ \langle m | n \rangle &= n! \delta_{n,m} \end{aligned} \quad (3.7)$$

Portanto, a e $a^\dagger$ são análogos aos operadores bosônicos de criação e destruição da mecânica quântica.

A equação mestra de qualquer processo markoviano pode ser escrita em termos de 3.5 por:

$$\frac{\partial | P \rangle}{\partial t} = H | P(t) \rangle \quad (3.8)$$

Para a equação mestra 3.1 do processo de nascimento e morte, temos em particular:

$$H = (a^\dagger - 1) (k - \rho a) \quad (3.9)$$

Dada a similaridade da equação 3.8 com a equação de Schrodinger, a ideia é tentar

importar da mecânica quântica métodos aproximados para sua solução. No caso, buscaremos por um funcional cuja extremalização leve à equação 3.8 (para o estado estacionário apenas), a partir do qual podemos tornar estados testes o mais próximo possíveis da solução exata. Descreveremos o método seguindo a referência (30):

Consideremos a equação que desejamos resolver:

$$H | \Psi \rangle = 0 \quad (3.10)$$

Como H em geral não é um operador hermitiano, à equação 3.10 está associado um autovetor à esquerda $\langle \Psi^L |$, diferente de $| \Psi \rangle$, tal que:

$$\langle \Psi^L | \Psi \rangle = 1 \quad (3.11)$$

Dada a natureza probabilística do problema, é conveniente definir o produto escalar de modo que a normalização 3.11 corresponda à normalização da probabilidade total:

$$\sum_{n=0}^{\infty} p_n = 1 \quad (3.12)$$

Desta maneira, o autovetor à esquerda do estado estacionário é (20, 29):

$$\langle \Psi^L | = \langle 0 | e^a \quad (3.13)$$

É fácil de ver que a substituição de 3.13 em $\langle \Psi^L | P(t) \rangle$, lembrando de 3.7, leva a 3.12.

É possível mostrar (30) que a equação 3.10 advém da seguinte extremalização:

$$\delta \langle \Psi^L | H | \Psi \rangle = 0 \quad (3.14)$$

Escolhendo uma função teste $| \Psi(\vec{\lambda}) \rangle$ dependente de parâmetros livres $\vec{\lambda} = (\lambda_1 \dots \lambda_N)$ e modificando o autovetor à esquerda através de parâmetros $\vec{\mu} = (\mu_1 \dots \mu_N)$, a equação 3.14 deixa de ser a variação de um funcional nas funções $\langle \Psi^L |$ e $| \Psi \rangle$, e passa a ser a extremalização de uma função dos parâmetros $(\vec{\lambda}, \vec{\mu})$. Eyink (30) mostra que $\langle \Psi^L(\vec{\mu}) |$ deve ser escolhido de forma que $\langle \Psi^L(0) | = \langle \Psi^L |$, de modo que a equação 3.14 leve a:

$$\left(\frac{\partial \langle \Psi^L(\vec{\mu}) |}{\partial \mu_i} \right)_{\vec{\mu}=0} H | \Psi(\vec{\lambda}) \rangle_{i=1, \dots, N} = 0 \quad (3.15)$$

Vamos aplicar o método variacional (equações 3.15) para o caso de H dado pela 3.9.

Comparando as equações 3.2 e 3.5, podemos traçar a seguinte analogia:

$$\begin{aligned}
 |n\rangle &\rightarrow z^n \\
 |P\rangle &\rightarrow P(z) \\
 a &\rightarrow \frac{\partial}{\partial z} \\
 a^\dagger &\rightarrow z \\
 \langle 0 | e^a | \Psi \rangle &\rightarrow \Psi(z)_{z=1}
 \end{aligned} \tag{3.16}$$

As expressões 3.16 são análogas às realizações de operações em estados abstratos em mecânica quântica, em termos de alguma representação escolhida. Por exemplo, na representação de posição $|\vec{r}\rangle$ o estado $|\Psi\rangle$ em mecânica quântica é dado por uma função de onda $\Psi(\vec{r})$, e os operadores $\vec{R}$ e $\vec{P}$ são dados por $\vec{r}$ e $\frac{\hbar}{i}\vec{\nabla}$.

Escolhendo como funções testes $\langle \Psi^L(\mu) | = \langle 0 | e^{a+\mu a}$ e $\Psi(z) = \exp(\lambda(z-1))$, a equação 3.15 dá:

$$\left(\frac{\partial}{\partial z}(z-1) \left(k - \rho \frac{\partial}{\partial z} \right) \exp(\lambda(z-1)) \right)_{z=1} = 0 \tag{3.17}$$

, o que leva neste caso à solução exata $\lambda = \frac{k}{\rho}$.

3.2 Aplicação do método para o sistema de dois genes descrito no capítulo 2

Para o caso do modelo de dois genes introduzidos no capítulo 2, as expressões 3.5, 3.6 e 3.13 devem ser generalizadas para o caso de duas variáveis estocásticas n (gene 1) e m (gene 2):

$$|\Psi\rangle = \sum_{n,m=0}^{\infty} \begin{pmatrix} \alpha_{n,m} \\ \beta_{n,m} \end{pmatrix} |n, m\rangle \tag{3.18}$$

$$\begin{aligned}
a_z |n, m\rangle &= n |n-1, m\rangle \\
a_z^\dagger |n, m\rangle &= |n+1, m\rangle \\
a_y |n, m\rangle &= m |n, m-1\rangle \\
a_y^\dagger |n, m\rangle &= |n, m+1\rangle
\end{aligned} \tag{3.19}$$

$$\langle \Psi^L | = \langle 0, 0 | e^{a_z + a_y} \begin{pmatrix} 1 & 1 \end{pmatrix} \tag{3.20}$$

Desta maneira, as equações 2.1 e 2.2 são colocadas na forma da equação 3.8, com $H =$

$$\begin{pmatrix} k_1(a_z^\dagger - 1) - \rho_1 a_z + k_2(a_y^\dagger - 1) - \rho_2 a_y - h a_z^\dagger a_z & f \\ h a_z^\dagger a_z & k_1(a_z^\dagger - 1) - \rho_1 a_z + \chi k_2(a_y^\dagger - 1) - \rho_2 a_y - f \end{pmatrix}$$

A generalização das relações 3.16 é dada por:

$$\begin{aligned}
|n, m\rangle &\rightarrow z^n y^m \\
|\Psi\rangle &\rightarrow \Psi(z, y) \\
a_z, a_y &\rightarrow \frac{\partial}{\partial z}, \frac{\partial}{\partial y} \\
a_z^\dagger, a_y^\dagger &\rightarrow z, y \\
\langle \Psi^L | \Psi \rangle &\rightarrow \Psi(z, y)_{z=1, y=1}
\end{aligned} \tag{3.21}$$

Para aplicar as equações 3.15 à solução deste problema, precisamos propor uma função geradora teste $\Psi(z, y)$. Como dito no Capítulo 2, a referência (24) dá a solução para o problema de um gene que possa funcionar em 2 estados, sendo desativado a uma taxa constante λ ($\alpha_2(\lambda, y)$ e $\beta_2(\lambda, y)$ das equações 2.63 e 2.64). A solução exata para as proteínas repressoras é dada por 2.15. Portanto, uma boa proposição para a função geradora é:

$$\Psi(z, y) = \exp\left(\frac{k_1}{\rho_1}(z-1)\right) \begin{pmatrix} \alpha_2(\lambda, y) \\ \beta_2(\lambda, y) \end{pmatrix} \tag{3.22}$$

Escrevendo o autovetor à esquerda como:

$$\langle \Psi^L | = \langle 0, 0 | e^{a_z + a_y + \mu a_y} \begin{pmatrix} 1 & 1 \end{pmatrix} \tag{3.23}$$

, e aplicando a equação do método variacional 3.15, obtém-se $\lambda = h \frac{k_1}{\rho_1}$.

Portanto, o método variacional é capaz de reproduzir a solução aproximada apresentada no capítulo 2 (equações 2.49, 2.50 e 2.51), obtendo-a de uma maneira mais rigorosa e mais simples. Estas características permitem que o método possa ser aplicado com sucesso para redes mais complicadas.

A abordagem de redes mais realistas sob a luz do método variacional se daria da seguinte forma. Constrói-se funções geradoras testes através do conhecimento da solução exata de partes mais simples da rede, utilizando o máximo possível a física e biologia do problema, como feito na seção 2.3. Utiliza-se as equações variacionais 3.15 com o 'Hamiltoniano' do sistema completo, aproximando a solução proposta o melhor possível da solução exata. A vantagem é que desta forma obtém-se a distribuição de probabilidades completa considerando-se todos os componentes envolvidos, de uma maneira relativamente simples.

4 *Soluções exatas dependentes do tempo*

Este capítulo apresenta resultados preliminares quanto ao estudo da dinâmica de interação entre as proteínas repressoras (gene 1) e o sítio operador do gene reprimido (gene 2). Aqui está calculada a probabilidade dependente do tempo $\phi_n(t)$ de encontrar n proteínas repressoras na célula, e também a probabilidade $\alpha_n(t)$ de encontrar n proteínas do tipo 1 com o gene 2 ligado. Estas soluções poderão futuramente ser usadas como base para a compreensão de problemas mais complexos, a exemplo do que foi apresentado no Capítulo 3.

As equações diferenciais que descrevem as funções geradoras destas probabilidades, $\phi(z, t) = \sum_{n=0}^{\infty} \phi_n(t) z^n$ e $\alpha(z, t) = \sum_{n=0}^{\infty} \alpha_n(t) z^n$, são obtidas fazendo $y = 1$ nas equações 1.4 e 1.5 (ou seja, somando sobre todas as proteínas do segundo gene).

$$\frac{\partial \phi}{\partial t} = (z - 1) \left(k_1 \phi(z, t) - \rho_1 \frac{\partial \phi}{\partial z} \right) \quad (4.1)$$

$$\frac{\partial \alpha}{\partial t} = (z - 1) \left(k_1 \alpha(z, t) - \rho_1 \frac{\partial \alpha}{\partial z} \right) - h z \frac{\partial \alpha}{\partial z} + f(\phi(z, t) - \alpha(z, t)) \quad (4.2)$$

O capítulo será dividido em duas partes. Na primeira, resolveremos as equações 4.1 e 4.2 utilizando o método das funções características obtendo as soluções $\phi(z, t)$ e $\alpha(z, t)$ em termos das condições iniciais $\phi(z, 0)$ e $\alpha(z, 0)$. Na segunda, reobteremos as soluções em termos de funções de Heun confluentes, expressando os resultados em uma forma mais compacta. Como consequência, obteremos também uma expansão de uma classe de funções de Heun confluentes em termos de uma soma finita de funções hypergeométricas confluentes.

4.1 Solução pelo método das características

4.1.1 Solução para $\phi(z, t)$

O método das características consiste em encontrar uma família de curvas no espaço das variáveis de uma equação diferencial parcial, sobre as quais a EDP em questão se transforma em uma equação diferencial ordinária. A solução desta EDO juntamente com o conhecimento da família de curvas permitem a obtenção da solução da EDP original.

Consideremos portanto a curva parametrizada por s : $(z(s), t(s))$. Sobre esta curva, $\phi(z, t) = \phi(z(s), t(s)) \equiv \phi(s)$. A derivada de ϕ com respeito a s é:

$$\frac{d\phi}{ds} = \frac{d\phi}{dt} \frac{dt}{ds} + \frac{d\phi}{dz} \frac{dz}{ds} \quad (4.3)$$

Escolhendo,

$$\frac{dt}{ds} = 1 \quad (4.4)$$

$$\frac{dz}{ds} = \rho_1(z - 1) \quad (4.5)$$

a equação 4.1 passa a ser:

$$\frac{d\phi}{ds} = k_1 (z(s) - 1) \phi(s) \quad (4.6)$$

Ou seja, as equações 4.4 e 4.5 definem uma família de curvas sobre as quais a EDP 4.1 se transforma na EDO 4.6. A solução de 4.4 e 4.5 é:

$$\begin{aligned} t(s) &= s \\ z(s) &= 1 + Ae^{\rho_1 s} \end{aligned} \quad (4.7)$$

, onde A é uma constante que define qual curva característica estamos olhando.

Com as expressões acima para $t(s)$ e $z(s)$ obtém-se a solução para $\phi(s)$:

$$\phi(s) = B \exp \left(\frac{k_1}{\rho_1} (z(s) - 1) \right) \quad (4.8)$$

Na expressão acima B é uma constante sobre a característica $(z(s), t(s))$. Entretanto, pelas equações 4.7, qualquer função de $A = (z - 1)e^{-\rho_1 t}$ é constante sobre a curva dada. Portanto, a solução mais geral para a equação 4.1 é

$$\phi(z, t) = g \left((z - 1)e^{-\rho_1 t} \right) \exp \left(\frac{k_1}{\rho_1} (z(s) - 1) \right) \quad (4.9)$$

, em termos de uma função arbitrária g , ou, em termos de uma dada condição inicial $\phi(z, 0)$:

$$\phi(z, t) = \phi \left(1 + (z - 1)e^{-\rho_1 t}, 0 \right) \exp \left(\frac{k_1}{\rho_1} (z - 1) (1 - e^{-\rho_1 t}) \right) \quad (4.10)$$

Como podemos ver da solução 4.10, o gene repressor vai para o estado estacionário na forma $\phi(z, t) = \sum_{n=0}^{\infty} \phi_n(z) e^{-n\rho_1 t}$, de modo que seu tempo característico de chegada ao equilíbrio é dado por $T_1 = \frac{1}{\rho_1}$.

4.1.2 Solução para $\alpha(z, t)$

Procedemos de maneira análoga à seção anterior, escolhendo uma curva característica no plano (z, t) , parametrizada por s de modo que $\alpha(z, t) = \alpha(z(s), t(s)) \equiv \alpha(s)$.

As equações que definem as curvas características são:

$$\begin{aligned} \frac{dt}{ds} &= 1 \\ \frac{dz}{ds} &= (\rho_1 + h)z - \rho_1 \end{aligned} \quad (4.11)$$

As soluções para 4.11 são:

$$\begin{aligned} t(s) &= s \\ z(s) &= \frac{\rho_1}{\rho_1 + h} + \frac{\bar{A}}{\rho_1 + h} e^{(\rho_1 + h)s} \end{aligned} \quad (4.12)$$

Como na seção anterior, $\bar{A}$ é uma constante que seleciona uma curva da família, sobre a qual a EDP 4.2 se transforma na EDO abaixo:

$$\frac{d\alpha}{ds} + (f - k_1(z(s) - 1)) \alpha(z(s), s) = f\phi(z(s), s) \quad (4.13)$$

A equação 4.13 é uma EDO inhomogênea, e portanto sua solução particular dependerá de uma integral do tipo $\int_t^s G(s)\phi(z(s), s)ds$. O limite t sobre o sinal de integração indica que a expressão para a primitiva de $G(s)\phi(z(s), s)$ deve ser calculada para $s = t$. O outro limite da integral que contribuiria com uma constante aditiva é embutido na solução homogênea da equação. Entretanto, $\phi(z, t)$ é em princípio uma função arbitrária e, portanto, expressar a solução particular desta maneira torna obscuro o modo como o estado do sítio operador do segundo gene vai para o equilíbrio. Portanto, faremos na equação 4.13 a mudança de variáveis $x = \frac{\rho_1}{\rho_1+h} + \frac{A}{\rho_1+h}e^{(\rho_1+h)s}$, de modo que a solução particular possa ser escrita da forma $\int^z \bar{G}(x)\phi(x, t(x))dx$, com $t(x) = t + \frac{1}{\rho_1+h} \ln \frac{(\rho_1+h)x - \rho_1}{(\rho_1+h)z - \rho_1}$, consistente com as equações 4.12. Desta maneira, a dependência da solução particular com o tempo fica mais clara, saindo do limite de integração e facilitando a identificação dos tempos de relaxação.

Após a mudança de variável a equação 4.13 assume a forma:

$$\frac{d}{dx} ((x - \varrho)^{F+KH} e^{-Kx} \alpha) = F(x - \varrho)^{F+KH-1} e^{-Kx} \phi(x, t(x)) \quad (4.14)$$

Os novos parâmetros que aparecem na equação 4.14 são os mesmos definidos no Capítulo 2.

Da equação 4.14, $\alpha(z, t)$ é dado por

$$\alpha(z, t) = \frac{F e^{Kz}}{(z - \varrho)^{F+KH}} \int^z (z - \varrho)^{F+KH-1} e^{-Kx} \phi(x, t(x)) dx + \bar{B} \frac{e^{Kz}}{(z - \varrho)^{F+KH}} \quad (4.15)$$

, onde $\bar{B}$ é uma constante para a curva característica em questão.

Pelas equações 4.12, qualquer função de $\bar{A} = (z - \varrho)e^{-(\rho_1+h)t}$ é uma constante sobre as curvas características, de modo que podemos escrever $\bar{B} = \bar{g}((z - \varrho)e^{-(\rho_1+h)t})$, com $\bar{g}$ uma função arbitrária.

Chamando a solução particular (dependente da integral de ϕ) de $\alpha_p(z, t)$, e expres-

sando $\bar{g}$ em termos da condição inicial dada $\alpha(z, 0)$, temos finalmente:

$$\alpha(z, t) = \alpha_p(z, t) + [\alpha((z - \varrho)e^{-(\rho_1+h)t} + \varrho, 0) - \alpha_p((z - \varrho)e^{-(\rho_1+h)t} + \varrho, 0)] \times \exp[K(z - \varrho)(1 - e^{-(\rho_1+h)t})]e^{-\delta t} \quad (4.16)$$

Na expressão 4.16, temos:

$$\delta = f + \frac{k_1 h}{\rho_1 + h} \quad (4.17)$$

Substituindo a solução geral para $\phi(z, t)$ (equação 4.10), em $\alpha_p(z, t)$ encontramos:

$$\begin{aligned} \alpha_p(z, t) &= \frac{F e^{Kz}}{(z - \varrho)^{F+KH}} \int_0^z (x - \varrho)^{F+KH-1} e^{-Kx} \phi\left(1 + \left(\frac{z - \varrho}{x - \varrho}\right)^{\varrho} (x - 1)e^{-\rho_1 t}, 0\right) \times \\ &\times \exp\left[\frac{K}{\varrho}(x - 1)\left(1 - \left(\frac{z - \varrho}{x - \varrho}\right)^{\varrho} e^{-\rho_1 t}\right)\right] dx \end{aligned} \quad (4.18)$$

Vemos da expressão para $\alpha(z, t)$ (4.16) que o estado do sítio operador do segundo gene vai para o equilíbrio com tempo característico $T_{op} = \frac{1}{\delta}$. Portanto, assumindo que o processo de produção de proteínas pelo gene 2 se estabilize com tempo $T_2 = \frac{1}{\rho_2}$ (em analogia com T_1), temos aqui uma estimativa mais precisa para o parâmetro de adiabaticidade ε introduzido no Capítulo 2.

4.1.3 Solução explícita em termos das condições iniciais

Para a solução completa do problema, precisamos calcular $\alpha_p(z, t)$ para qualquer $\phi(z, 0)$ dado. Consideremos então uma condição inicial genérica para ϕ dada por:

$$\phi(z, 0) = \sum_{m=0}^{\infty} T_m (z - 1)^m \quad (4.19)$$

Pela expressão 1.7 vemos que os T_m estão relacionados com os m-ésimos momentos da distribuição, e que $T_0 = 1$. Substituindo a 4.19 na 4.18 vemos que a determinação de

$\alpha_p(z, t)$ depende do cálculo da integral

$$I_m = T_m \int_0^z (x - \varrho)^{F+KH-1} e^{-Kx} \left(\frac{z - \varrho}{x - \varrho} \right)^{m\varrho} (x - 1)^m e^{-m\rho_1 t} \times \\ \times \exp \left[\frac{K}{\varrho} (x - 1) \left(1 - \left(\frac{z - \varrho}{x - \varrho} \right)^\varrho e^{-\rho_1 t} \right) \right] dx \quad (4.20)$$

através de:

$$\alpha_p(z, t) = \frac{F e^{Kz}}{(z - \varrho)^{F+KH}} \sum_{m=0}^{\infty} I_m \quad (4.21)$$

Expandindo $\exp \left(-\frac{K}{\varrho} (x - 1) \left(\frac{z - \varrho}{x - \varrho} \right)^\varrho e^{-\rho_1 t} \right)$ em série de potências, temos:

$$I_m = \sum_{n=0}^{\infty} (-1)^n a_{n,m} \int_0^z (x - 1)^{n+m} (x - \varrho)^{F+KH-1-(n+m)\varrho} e^{\frac{K}{\varrho}(Hx-1)} dx \\ a_{n,m} = \frac{1}{n!} \left(\frac{K}{\varrho} \right)^n (z - \varrho)^{(n+m)\varrho} e^{-(n+m)\rho_1 t} T_m \quad (4.22)$$

Fazendo $j = n - m$, $(x - 1)^j = \sum_{q=0}^j \frac{j!}{q!(j-q)!} (-H)^{j-q} (x - \varrho)^q$, e mudando a variável de integração para $y = x - \varrho$, temos:

$$I_m = \sum_{j=m}^{\infty} \sum_{q=0}^j \frac{(-1)^{m+q} j!}{q!(j-q)!} H^{j-q} a_{j,m} e^{-\frac{K}{\varrho}(H+\varrho^2)} \int_0^{z-\varrho} y^{F+KH+q-j\varrho-1} e^{\frac{K}{\varrho}Hy} dy \quad (4.23)$$

A integral acima é tabelada, e apresenta solução em termos da função hipergeométrica confluyente regular M , ou função de Kummer (equação 13.2.1 da referência (27)):

$$\int_0^{z-\varrho} y^{F+KH+q-j\varrho-1} e^{\frac{K}{\varrho}Hy} dy = \frac{y^{F+KH+q-j\varrho} \exp \left(\frac{K}{\varrho}Hy \right)}{F + KH + q - j\varrho} \times \\ \times M \left(1, 1 + F + KH + q - j\varrho, -\frac{K}{\varrho}Hy \right) \quad (4.24)$$

Portanto, combinando o resultado acima com 4.23 e 4.22 temos:

$$\frac{F e^{Kz}}{(z - \varrho)^{F+KH}} I_m = e^{\frac{K}{\varrho}(z-1)} \sum_{j=m}^{\infty} \frac{(-1)^m}{(j-m)!} \frac{T_m}{(K/\varrho)^m j!} \left(\frac{KH}{\varrho} \right)^j e^{-j\rho_1 t} \times \\ \times \sum_{q=0}^j \frac{(-1)^q F}{(F + KH + q - j\varrho) q! (j-q)!} \left(\frac{z - \varrho}{H} \right)^q \times \\ \times M \left(1, 1 + F + KH + q - j\varrho, -\frac{KH}{\varrho}(z - \varrho) \right) \quad (4.25)$$

Finalmente, partindo da expressão 4.21, $\alpha_p(z, t)$ é escrita como

$$\begin{aligned} \alpha_p(z, t) = & e^{\frac{K}{\varrho}(z-1)} \sum_{j=0}^{\infty} C_{j,\phi} e^{-j\rho_1 t} \times \\ & \times \sum_{q=0}^j \frac{(-1)^q F}{(F + KH + q - j\varrho)q!(j-q)!} \left(\frac{z - \varrho}{H} \right)^q \times \\ & \times M \left(1, 1 + F + KH + q - j\varrho, -\frac{KH}{\varrho}(z - \varrho) \right) \end{aligned} \quad (4.26)$$

, com:

$$C_{j,\phi} = \sum_{m=0}^j \frac{(-1)^m j!}{(j-m)!} \frac{T_m}{(K/\varrho)^m} \left(\frac{KH}{\varrho} \right)^j \quad (4.27)$$

Portanto, dadas as distribuições iniciais $\phi(z, 0)$ e $\alpha(z, 0)$, as expressões 4.10, 4.16 e 4.26 fecham completamente a solução do problema.

4.2 Solução em termos das funções de Heun confluentes

Vamos agora reobter a solução particular $\alpha_p(z, t)$ de uma maneira mais direta em termos das funções de Heun confluentes.

As equações 4.1 e 4.2 apresentam pontos singulares em $z = 1$ e $z = \varrho$, respectivamente. Para escrevê-las numa forma canônica colocaremos as singularidades em $u = 1$ e $u = 0$ através da mudança de variáveis $u = \frac{z-\varrho}{H}$ como feito no Capítulo 2. Substituiremos também a variável temporal por $\tau = \exp(-\rho_1 t)$, uma vez que vimos que as soluções $\alpha_p(z, t)$ e $\phi(z, t)$ (equação 4.26 e comentário subsequente à 4.10) são naturalmente expandidas em 'modos normais' de decaimento com taxa ρ_1 . Baseado na solução 4.26 para $\alpha_p(z, t)$ e notando que $(z - 1) = H(u - 1)$, vamos escrever ϕ e α_p como:

$$\begin{aligned} \phi(z, t) = & \exp \left(\frac{K}{\varrho}(z - 1) \right) \sum_{j=0}^{\infty} \tilde{\phi}_j(u) e^{-j\rho_1 t} = \exp \left(\frac{KH}{\varrho}(u - 1) \right) \sum_{j=0}^{\infty} \tilde{\phi}_j(u) \tau^j \\ \alpha_p(z, t) = & \exp \left(\frac{K}{\varrho}(z - 1) \right) \sum_{j=0}^{\infty} \tilde{\alpha}_j(u) e^{-j\rho_1 t} = \exp \left(\frac{KH}{\varrho}(u - 1) \right) \sum_{j=0}^{\infty} \tilde{\alpha}_j(u) \tau^j \end{aligned} \quad (4.28)$$

Inserindo as transformações 4.28 nas equações 4.1 e 4.2, temos:

$$(u-1)\frac{d\tilde{\phi}_j}{du} - j\tilde{\phi}_j(u) = 0 \quad (4.29)$$

$$u\frac{d\tilde{\alpha}_j}{du} + \left(\frac{KH^2}{\varrho}u + F + KH - j\varrho\right)\tilde{\alpha}_j(u) = F\tilde{\phi}_j(u) \quad (4.30)$$

A solução de 4.29 é de imediato dada por $(u-1)^j$, a menos de uma constante multiplicativa. Este resultado se inserido na 4.28, é idêntico à 4.10, obtida pelo método das características.

A seguinte equação diferencial homogênea de segunda ordem para $\tilde{\alpha}_j$ é obtida isolando-se o $\tilde{\phi}_j$ na 4.30 e substituindo na 4.29

$$u(u-1)\frac{d^2\tilde{\alpha}_j}{du^2} + (u-1)(au + b_j + 1 - j)\frac{d\tilde{\alpha}_j}{du} - (a(j-1)u + a + jb_j)\tilde{\alpha}_j(u) = 0 \quad (4.31)$$

, com:

$$\begin{aligned} a &= \frac{KH^2}{\varrho} \\ b_j &= F + KH - j\varrho \end{aligned} \quad (4.32)$$

A equação 4.31 apresenta dois pontos singulares regulares, em $u = 0$ e $u = 1$, e um irregular no infinito. Portanto, a solução pode ser expressa em termos das funções de Heun confluentes (*HeunC*). De fato a solução é dada por (com A_j e $\bar{A}_j$ constantes):

$$\begin{aligned} \tilde{\alpha}_j(u) = & A_j(u-1)^{j+1}\text{HeunC}\left(a, b_j, 1+j, -\frac{1}{2}a(b_j+j-1), \frac{1}{2} + \frac{1}{2}(a-j)(b_j-1), u\right) + \\ & + \bar{A}_j u^{-b_j}(u-1)^{j+1}\text{HeunC}\left(a, -b_j, 1+j, -\frac{1}{2}a(b_j+j-1), \frac{1}{2} + \frac{1}{2}(a-j)(b_j-1), u\right) \end{aligned} \quad (4.33)$$

A solução particular que estamos procurando ($\alpha_p(z, t)$ dada pela 4.26) não possui uma estrutura singular do tipo $(z - \varrho)^{-(F+KH)}$. De fato, vemos que esta estrutura aparece na parte homogênea da solução, dada pela equação 4.15. Portanto, para reobtermos a solução particular, devemos ter $\bar{A}_j = 0$.

Portanto, como as equações 4.26 e 4.33 (levando-se em conta a transformação 4.28)

representam a mesma solução, temos:

$$\begin{aligned} A_j(u-1)^{j+1}HeunC\left(a, b_j, 1+j, -\frac{1}{2}a(b_j+j-1), \frac{1}{2} + \frac{1}{2}(a-j)(b_j-1), u\right) = \\ = C_{j,\phi} \sum_{q=0}^j \frac{(-1)^q F}{(b_j+q)q!(j-q)!} u^q M(1, 1+b_j+q, -au) \end{aligned} \quad (4.34)$$

A relação acima tem que valer qualquer que seja o valor de u . Em particular, tem que valer para $u = 0$. Para este valor de argumento tanto a função de Kummer quanto a função de Heun confluyente valem 1, e a igualdade 4.34 dá para a constante A_j :

$$A_j = \frac{F}{b_j} \frac{(-1)^{j+1}}{j!} C_{j,\phi} \quad (4.35)$$

A solução particular $\alpha_p(z, t)$ para a função geradora da probabilidade de haver n proteínas repressoras na célula, com o gene 2 ligado é, portanto, de uma maneira mais compacta:

$$\begin{aligned} \alpha_p(u, \tau) = e^{\frac{KH}{\epsilon}(u-1)} \sum_{j=0}^{\infty} \frac{F}{j!b_j} C_{j,\phi} (1-u)^{j+1} \times \\ \times HeunC\left(a, b_j, 1+j, -\frac{1}{2}a(b_j+j-1), \frac{1}{2} + \frac{1}{2}(a-j)(b_j-1), u\right) \tau^j \end{aligned} \quad (4.36)$$

Como consequência das expressões 4.34 e 4.35, obtivemos também a seguinte expansão para uma classe de funções de Heun confluentes em termos das funções de Kummer:

$$\begin{aligned} (1-u)^{j+1}HeunC\left(a, b, 1+j, -\frac{1}{2}a(b+j-1), \frac{1}{2} + \frac{1}{2}(a-j)(b-1), u\right) = \\ = \sum_{q=0}^j (-1)^q \frac{b}{(b+q)} \frac{j!}{q!(j-q)!} u^q M(1, 1+b+q, -au) \end{aligned} \quad (4.37)$$

A igualdade matemática acima vale, nas condições em que foi demonstrado aqui, para a e b reais positivos e j um número natural.

5 *Conclusão*

A postura adotada para o desenvolvimento deste trabalho consiste na ideia de que uma melhor compreensão das redes regulatórias deve advir de um entendimento teórico das sub-redes que as compõe. Portanto, a atenção foi focada no sistema de dois genes interagentes mais simples, em que o primeiro atua como repressor, e o segundo alterna entre os estados ligado ou desligado, conforme a ação das proteínas repressoras.

No capítulo 2 a solução exata para as funções geradoras das distribuições de probabilidade no estado estacionário foram obtidas, através do requerimento de que as funções geradoras fossem analíticas nas suas variáveis. Esta solução, aliada à solução da dependência temporal apresentada no Capítulo 4, expande o limite de resultados exatos inicialmente restritos ao caso de um gene isolado ou de um gene puramente constitutivo (que funciona em apenas um estado) (19, 24, 25).

Devido ao desenvolvimento atual na área de computação simbólica, com softwares como Maple e Mathematica, o fato de as soluções se apresentarem em termos de funções hipergeométricas e de Heun não debilitam sua utilidade. Pelo contrário, a introdução destas funções passam a representar um avanço na área de biologia teórica.

A figura 2.10 traz um resultado interessante obtido da solução exata para o estado estacionário. Para valores típicos de quantidade de repressores, o esquema de controle estudado neste trabalho (equações 1.1 e 1.2) ocorre de modo que a correlação entre os números de proteínas 1 e 2 seja muito próxima de zero. Isto não significa que as concentrações de proteínas do tipo 2 não dependam do número de proteínas repressoras, mas sim que o número médio de proteínas produzidas pelo gene 2 é pouco sensível a flutuações no número de proteínas produzidas pelo gene 1. Este comportamento está de acordo com a intuição de que apesar da estocasticidade inerente à expressão gênica, as redes têm que apresentar uma certa confiabilidade na transmissão de informação.

Podemos portanto especular sobre uma possível regra de design de redes gênicas, de que dado um certo modelo para a expressão gênica, o coeficiente de correlação entre os

componentes interagentes deva ser minimizado. Apesar do caráter especulativo, foi esta ideia que possibilitou a obtenção no Capítulo 2 de uma excelente aproximação para a solução exata (equações 2.49, 2.50 e 2.51).

A solução aproximada das equações de 2.49 a 2.51 constituem uma generalização da solução apresentada em (24) para o caso de um único gene, e nos dá expressões simples para a maneira pela qual a concentração de repressores altera os momentos (probabilidade de ocupação, média, variância e assim por diante) da distribuição das proteínas 2 (equações 2.58, 2.59 e 2.62).

Um aspecto extremamente relevante é o fato da solução aproximada permitir de maneira bem sucedida a generalização dos resultados para redes reais, como apresentado na seção 2.3. Os resultados experimentais da referência (15) são reproduzidos a partir de funções com a forma das 2.63 e 2.64.

Portanto, funções geradoras como as das equações 2.63 e 2.64 se apresentam ideais para serem utilizadas como base para generalizações a outras redes. Portanto, um possível caminho para a modelagem de redes mais completas é a utilização destas funções como base para o método variacional apresentado no Capítulo 3, em contraposição à utilização da função geradora puramente exponencial (distribuição de Poisson) utilizada na literatura (20, 31).

REFERÊNCIAS

- 1 MONOD, J.; JACOBI, F. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, v. 3, n. 3, p. 318–356, 1961.
- 2 MÜLLER-HILL, B. *The lac operon: a short history of a genetic paradigm*. Berlin-New York: Walter de Gruyter, 1996.
- 3 PTASHNE, M. *A genetic switch: phage lambda and higher order organisms*. 2nd. ed. Cambridge: Cell Press/Blackwell, 1992.
- 4 JANSSENS, H.; HOU, S.; JAEGGER, J.; KIM, A.-R.; MYASNIKOVA, E.; SHARP, D.; REINITZ, J. Quantitative and predictive model of transcriptional control of the drosophila melanogaster even skipped gene. *Nature Genetics*, v. 38, n. 10, p. 1159–1164, 2006.
- 5 RAJ, A.; VAN OUDENAARDEN, A. Single-molecule approaches to stochastic gene expression. *Annual Review of Biophysics*, v. 38, n. 1, p. 255–270, 2009.
- 6 ELOWITZ, M. B.; LEVINE, A. J.; SIGGIA, E. D.; SWAIN, P. Stochastic gene expression in a single cell. *Science*, Washington, v. 297, n. 5584, p. 1183–1186, 2002.
- 7 OZBUDAK, E. M.; THATTAI, M.; KURTSEER, I.; GROSSMAN, A. D.; VAN OUDENAARDEN, A. Regulation of noise in the expression of a single gene. *Nature Genetics*, v. 31, n. 1, p. 69–73, 2002.
- 8 GILLESPIE, D. T. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, Washington, v. 81, n. 25, p. 2340–2361, 1977.
- 9 MCADAMS, H. H.; ARKIN, A. Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, v. 94, n. 3, p. 814–819, 1997.
- 10 MCADAMS, H. H.; ARKIN, A. It's a noisy business! Genetic regulation at the nanomolar scale. *Trends in Genetics*, v. 15, n. 2, p. 65–69, 1999.
- 11 SCHULTZ, D.; ONUCHIC, J. N.; WOLYNES, P. G. Understanding stochastic simulations of the smallest genetic networks. *The Journal of Chemical Physics*, v. 126, n. 24, p. 245102 1 –245102 11, 2007.

- 12 WALCZAK, A. M.; MUGLER, A.; WIGGINS, C. A stochastic spectral analysis of transcriptional regulatory cascades. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, v. 106, n. 16, p. 6529–6534, 2009.
- 13 HASTY, J.; PRADINES, J.; DOLNIK, M.; COLLINS, J. J. Noise based switches and amplifiers for gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, v. 97, n. 5, p. 2075–2080, 2000.
- 14 THATTAI, M.; VAN OUDENAARDEN, A. Attenuation of noise in ultrasensitive signaling cascades. *Biophysical Journal*, v. 82, n. 6, p. 2943–2950, 2002.
- 15 PEDRAZA, J. M.; VAN OUDENAARDEN, A. Noise propagation in gene networks. *Science*, Washington, v. 307, n. 5717, p. 1965–1969, 2005.
- 16 PAULSSON, J. Summing up the noise in gene networks. *Nature*, London, v. 427, n. 6973, p. 415–418, 2004.
- 17 BONNEAU, R.; REISS, D. J.; SHANNON, P.; FACCIOTTI, M.; HOOD, L.; BALIGA, N. S.; THORSSON, V. The inferelator: an algorithm for learning parsimonious regulatory networks from systems-biology data sets de novo. *Genome Biology*, v. 7, n. 5, p. 1–16, 2006.
- 18 KEPLER, T. B.; ELSTON, T. C. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophysical Journal*, v. 81, n. 6, p. 3116–3136, 2001.
- 19 THATTAI, M.; VAN OUDENAARDEN, A. Intrinsic noise in gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, v. 98, n. 15, p. 8168–8172, 2001.
- 20 SASAI, M.; WOLYNES, P. G. Stochastic gene expression as a many-body problem. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, v. 100, n. 5, p. 2374–2379, 2003.
- 21 REICHL, L. E. *A modern course in statistical physics*. 3rd. ed. New York: John Wiley & Sons, Inc, 1998.
- 22 VAN KAMPEN, N. G. *Stochastic processes in physics and chemistry*. 2nd. ed. Amsterdam: North-Holland, 1992.
- 23 WALCZAK, A. M.; SASAI, M.; WOLYNES, P. G. Self-consistent proteomic field theory of stochastic gene switches. *Biophysical Journal*, v. 88, n. 2, p. 828–850, 2005.
- 24 INNOCENTINI, G. C. P.; HORNOS, J. E. M. Modeling stochastic gene expression under repression. *Journal of Mathematical Biology*, v. 55, n. 3, p. 413–431, 2007.

- 25 HORNOS, J. E. M.; SCHULTZ, D.; INNOCENTINI, G. C. P.; WANG, J.; WALCZAK, A. M.; ONUCHIC, J. N.; ONUCHIC, P. G. Self-regulating gene: an exact solution. *Physical Review E*, v. 72, n. 5, p. 051907 1–051907 5, 2005.
- 26 FIZIEV, P. P. Exact solutions of regge-wheeler equation and quasi-normal modes of compact objects. *Classical and Quantum Gravity*, v. 23, n. 7, p. 2447–2468, 2006.
- 27 ABRAMOWITZ, A.; STEGUN, I. A. *Handbook of mathematical functions with formulas, graphs and mathematical tables*. 10th. ed. Washington, D.C.: Nat. Bur. Standards Appl. Series, U.S. Government Printing Office, 1972.
- 28 MAHESHRI, N.; O'SHEA, E. K. Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annual Review of Biophysics and Biomolecular Structure*, v. 36, n. 1, p. 413–434, 2007.
- 29 DOI, M. Second quantization representation for classical many-particle system. *Journal of Physics A: mathematical and general*, v. 9, n. 9, p. 1465–1477, 1976.
- 30 EYINK, G. L. Action principle in nonequilibrium statistical dynamics. *Physical Review E*, v. 54, n. 4, p. 3419–3435, 1996.
- 31 LAN, Y.; WOLYNES, P. G.; PAPOIAN, G. A. A variational approach to the stochastic aspects of cellular signal transduction. *The Journal of Chemical Physics*, v. 125, n. 12, p. 124106 1–124106 11, 2006.

APÊNDICE A – Obtenção da equação 2.13

Na seção 2.1.1 do capítulo 2 introduzimos a função $\alpha_m(u)$ e vimos que para a obtenção da distribuição de probabilidades conjunta era necessário exprimir esta função na forma $\sum_{n=0}^{\infty} A_{n,m}(u-1)^n$. Obtivemos naquela seção que (equação 2.12):

$$\alpha_m(u) = \frac{F}{b_m} \sum_{n=0}^{\infty} (-1)^n T_{n,m} {}_2F1(-n, b_m; 1 + b_m; u) \quad (\text{A.1})$$

, onde ${}_2F1$ é a função hipergeométrica, definida através da expressão 15.1.1 da referência (27):

$${}_2F1(a, b; c; x) = \sum_{n=0}^{\infty} \frac{(a)_n (b)_n}{(c)_n} \frac{x^n}{n!} \quad (\text{A.2})$$

$(\)_n$ denota o símbolo de Pochhammer, definido por:

$$(a)_n = a(a+1)(a+2)\dots(a+n-1) \quad (\text{A.3})$$

As derivadas da função hipergeométrica podem ser obtidas através das expressões A.2 e A.3 e são dadas por:

$$\frac{d^n}{dx^n} {}_2F1(a, b; c; x) = \frac{(a)_n (b)_n}{(c)_n} {}_2F1(a+n, b+n; c+n; x) \quad (\text{A.4})$$

Expandindo a expressão A.1 em série de Taylor em torno de $u = 1$, temos:

$$\alpha_m(u) = \sum_{n=0}^{\infty} \sum_{k=0}^n \Delta_{n,k,m} (u-1)^k \quad (\text{A.5})$$

, onde pelas expressões A.1 e A.4:

$$\begin{aligned} \Delta_{n,k,m} &= \frac{(-1)^n F T_{n,m}}{b_m} \frac{(-n)_k (b_m+1)_k}{k! (b_m)_k} {}_2F1(-n+k, b_m+k; 1+b_m+k; 1) = \\ &= \frac{(-1)^{n+k} F T_{n,m}}{b_m+k} \frac{n!}{k! (n-k)!} {}_2F1(-n+k, b_m+k; 1+b_m+k; 1) \end{aligned} \quad (\text{A.6})$$

A soma em k da equação A.5 vai apenas até n pois $(-n)_k = 0$ se $k \geq (n + 1)$. Utilizando a relação 15.1.20 da referência (27) podemos simplificar a expressão A.6 para:

$$\Delta_{n,k,m} = FT_{n,m} \frac{(-1)^{n+k} n!}{k!} \frac{\Gamma(b_m + k)}{\Gamma(1 + b_m + n)} \quad (\text{A.7})$$

A expressão A.5 pode ser reescrita como:

$$\alpha_m(u) = \sum_{n=0}^{\infty} \left(\sum_{k=n}^{\infty} \Delta_{k,n,m} \right) (u - 1)^n \quad (\text{A.8})$$

Portanto, $A_{n,m}$ é dado por (equação 2.13):

$$A_{n,m} = \sum_{k=n}^{\infty} T_{k,m} F \frac{(-1)^{k+n} k!}{n!} \frac{\Gamma(n + b_m)}{\Gamma(1 + k + b_m)} \quad (\text{A.9})$$

APÊNDICE B – Funções geradoras aproximadas

No capítulo 2 reescrevemos as funções geradoras como:

$$\begin{aligned}\phi(u, v) &= \exp \left(KH(u-1) + \frac{\kappa}{\rho}v \right) \sum_{m=0}^{\infty} \phi_m(u) v^m \\ \alpha(u, v) &= \exp \left(KH(u-1) + \frac{\kappa}{\rho}v \right) \sum_{m=0}^{\infty} \alpha_m(u) v^m\end{aligned}\tag{B.1}$$

Os ϕ_m e α_m são dados por:

$$\phi_m(u) = (\chi - 1) \frac{\kappa}{\varrho} \frac{\exp \left(\frac{KH^2}{\varrho} u \right)}{(u-1)^{m\frac{\varrho}{e}}} \int (u-1)^{(m\frac{\varrho}{e}-1)} \exp \left(-\frac{KH^2}{\varrho} u \right) (\phi_{m-1}(u) - \alpha_{m-1}(u)) du\tag{B.2}$$

$$\alpha_m(u) = \frac{F}{u^{m\rho+F+KH}} \int \phi_m(u) u^{m\rho+F+KH-1} du\tag{B.3}$$

, com:

$$\phi_0(u) = \exp \left(\frac{KH^2}{\varrho} (u-1) \right)\tag{B.4}$$

Na seção 2.2.1 buscamos por soluções que fossem separáveis nas variáveis u e v , e que fossem consistentes com as equações B.2, B.3 e B.4. Tal requerimento como discutido naquela seção é atingido através da aproximação:

$$\frac{F}{m\rho + F + KH} M \left(1, 1 + m\rho + F + KH, -\frac{KH^2}{\varrho} u \right) \approx \frac{F}{m\rho + F + KH + \frac{KH^2}{\varrho}}\tag{B.5}$$

Desta maneira, $\phi_m(u)$ e $\alpha_m(u)$ assumem a forma (como nas equações B.4, 2.46, 2.47

e 2.48):

$$\begin{aligned}\phi_m^{ap}(u) &= \varphi_m \exp\left(\frac{KH^2}{\varrho}(u-1)\right) \\ \alpha_m^{ap}(u) &= a_m \exp\left(\frac{KH^2}{\varrho}(u-1)\right) \\ \varphi_0 &= 1\end{aligned}\tag{B.6}$$

Substituindo as expressões B.6 em B.2 e B.3 temos:

$$\phi_m^{ap}(u) = -(1-\chi)\frac{\kappa}{m\rho}(\varphi_{m-1} - a_{m-1}) \exp\left(\frac{KH^2}{\varrho}(u-1)\right)\tag{B.7}$$

$$\alpha_m^{ap}(u) = \varphi_m \frac{F}{m\rho + F + KH} M\left(1, 1 + m\rho + F + KH, -\frac{KH^2}{\varrho}u\right) \exp\left(\frac{KH^2}{\varrho}(u-1)\right)\tag{B.8}$$

Através da aproximação B.5 as expressões B.7 e B.8 levam às seguintes relações de recorrência:

$$m\rho\varphi_m + \kappa(1-\chi)(\varphi_{m-1} - a_{m-1}) = 0\tag{B.9}$$

$$\left(m\rho + F + \frac{KH}{\varrho}\right)a_m - F\varphi_m = 0\tag{B.10}$$

Substituindo as expressões B.6 nas funções geradoras B.1, temos:

$$\begin{aligned}\phi^{ap}(u, v) &= \exp\left(\frac{k_1}{\rho_1}(z-1) + \frac{k_2}{\rho_2}(y-1)\right) \sum_{m=0}^{\infty} \varphi_m v^m \equiv \\ &\equiv \exp\left(\frac{k_1}{\rho_1}(z-1) + \frac{k_2}{\rho_2}(y-1)\right) \tilde{\phi}(v)\end{aligned}\tag{B.11}$$

$$\begin{aligned}\alpha^{ap}(u, v) &= \exp\left(\frac{k_1}{\rho_1}(z-1) + \frac{k_2}{\rho_2}(y-1)\right) \sum_{m=0}^{\infty} a_m v^m \equiv \\ &\equiv \exp\left(\frac{k_1}{\rho_1}(z-1) + \frac{k_2}{\rho_2}(y-1)\right) \tilde{\alpha}(v)\end{aligned}\tag{B.12}$$

Em termos das novas funções geradoras $\tilde{\phi}(v)$ e $\tilde{\alpha}(v)$ definidas em B.11 e B.12, as relações de recorrência B.9 e B.10 ficam:

$$\rho v \frac{d\tilde{\phi}}{dv} + \kappa(1-\chi)(\tilde{\phi}(v) - \tilde{\alpha}(v)) = 0\tag{B.13}$$

$$\rho v \frac{d\tilde{\alpha}}{dv} + \left(F + \frac{KH}{\varrho}\right)\tilde{\alpha}(v) - F\tilde{\phi} = 0\tag{B.14}$$

A solução do sistema de equações acima é dado por (como feito em (24)):

$$\tilde{\phi}(v) = M \left(\frac{k_1}{\rho_1} \frac{h}{\rho_2}, \frac{f + h \frac{k_1}{\rho_1}}{\rho_2}; (\chi - 1) \frac{k_2}{\rho_2} v \right) \quad (\text{B.15})$$

$$\tilde{\alpha}(v) = M \left(\frac{k_1}{\rho_1} \frac{h}{\rho_2}, \frac{f + h \frac{k_1}{\rho_1}}{\rho_2} + 1; (\chi - 1) \frac{k_2}{\rho_2} v \right) \quad (\text{B.16})$$

As expressões B.15 e B.16 junto com as B.11 e B.12 correspondem exatamente às soluções aproximadas apresentadas em 2.49, 2.50 e 2.51.