

UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

ALEXANDRE FERREIRA RAMOS

**Simetrias de Lie e modelagem
estocástica da regulação da expressão gênica**

São Carlos

2008

ALEXANDRE FERREIRA RAMOS

**Simetrias de Lie e modelagem
estocástica da regulação da expressão gênica**

Tese apresentada ao Programa de Pós-Graduação do Instituto de Física de São Carlos, da Universidade de São Paulo, para obtenção do título de Doutor em Ciências.

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

São Carlos

2008

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

Ramos, Alexandre Ferreira

Simetrias de Lie e modelagem estocástica da regulação da expressão gênica / Alexandre Ferreira Ramos; orientador José Eduardo Martinho Hornos./.- São Carlos, 2008.

83 p.

Tese (Doutorado em Ciências - Área de concentração: Física Básica)- Instituto de Física de São Carlos da Universidade de São Paulo.

1. Simetrias de Lie. 2. Regulação gênica. 3. Métodos estocásticos. 4. Funções especiais. I. Título.

folha de aprovacao

Esta tese é dedicada aos meus pais:
Humberto e Judith.

Este trabalho teve o apoio financeiro do CNPq.

Agradecimentos

Ess'é pra acabar Goias!!! Amadeu

É com muita felicidade que concluo estes quatro anos de doutorado. Como não poderia deixar de ser, foi com emoção. Há muitos aos quais devoto meus sinceros agradecimentos, pois sozinho eu não conseguiria chegar ao final com sucesso. Agradeço ao meu orientador, Zé Edu, por incentivar-me a buscar a desordem conceitual e pela paciência no decorrer deste processo. Agradeço ao CNPq pela bolsa e o auxílio, com que participei de duas conferências internacionais (numa delas houve auxílio da CPG-IFSC). Ao IFSC e seu corpo de professores e funcionários. Destes, cito os nomes do Silvio, Ivone e Mariana. Agradeço principalmente aos professores e pesquisadores que direta ou indiretamente, consciente ou inconscientemente, contribuíram para a formação de minha personalidade científica e/ou realização deste trabalho: S. Ragusa, F. Fontanari, T. Bonagamba, J. Egues, J. Peko, F. Forger, J. Reinitz, E. Lima, J. Teles, A. Colato, L. Castelano, N. Pasqua, R. Auccaise, M. Novaes. Agradeço, muito especialmente, à minha esposa Fabiana pela paciência, compreensão, incentivos e bom humor. Agradeço aos meus familiares, especialmente minhas avós figuras Joana e Cândida e minhas irmãs, Paula e Maíra. Agradeço aos meus amigos Gil, Tiagão, Edilson, Graça, Bianca, Gustavo, Letícia, Marcel.

“Eu sou é eu mesmo. Diverjo de todo o mundo... Eu quase que nada não sei. Mas desconfio de muita coisa. O senhor concedendo, eu digo: para pensar longe, sou cão mestre - o senhor solte em minha frente uma idéia ligeira, e eu rastreio essa por fundo de todos os matos, amém”

Riobaldo - Grande sertão: veredas

Resumo

RAMOS, A. F. **Simetrias de Lie e modelagem estocástica da regulação da expressão gênica.** 2008. 82 f. Tese (Doutorado) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2008.

Nesta tese, mostramos que o modelo estocástico binário para expressão gênica, por um gene auto-regulado, possui solução completa. A solução dependente do tempo é escrita via expansão em termos das funções de Heun confluentes. Apresentamos um exemplo de dinâmica estocástica desse gene. Para tal, desenvolvemos uma relação de recorrência entre derivadas arbitrárias das funções de Heun confluentes. Mostramos também que o regime estacionário deste modelo possui simetria de Lie $SO(2,1)$ tipo Lorentz. Esta simetria é análoga à simetria do momento angular, porém com um sinal “errado”. O invariante desta álgebra define a meia-vida relativa do regime dinâmico do gene. O equivalente do momento angular azimutal é uma medida indireta do nível de atividade do gene. Os operadores levantamento e abaixamento conectam diferentes processos estocásticos de expressão proteínica. As flutuações destes processos estocásticos são classificadas em termos das relações entre os etiquetadores de um elemento da representação da álgebra. No arcabouço da teoria dos grupos, o modelo estocástico para um gene externamente regulado aparece como um caso particular do modelo para um gene auto-regulado. Mostramos, por fim, uma comparação entre estas duas estratégias de regulação. Demonstramos que um gene auto-regulado pode expressar proteínas em regimes sub Poisson, Poisson ou super Poisson. Por seu turno, o gene externamente regulado somente expressa proteínas em regimes Poisson ou super Poisson. Portanto, num processo estocástico, a auto-regulação mostra-se como uma forma de controle mais precisa. Também mostramos que a dinâmica de genes auto-regulados possui meia-vida mais curta que a de genes externamente regulados. Ou seja, a auto-regulação permite respostas mais rápidas à perturbações externas.

Palavras-chave: Simetrias de Lie. Regulação gênica. Métodos estocásticos. Funções especiais.

Abstract

RAMOS, A. F. **Lie symmetries and stochastic modeling of gene expression regulation.** 2008. 82 f. Thesis (Doctoral) - Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2008.

In this thesis we show that the stochastic binary model to protein synthesis by an auto-regulated gene is completely solvable. The time-dependent solution is written in terms of the confluent Heun functions. We present an example of probability dynamics to this gene. To get that, we developed a recurrence relation between arbitrary derivatives of the confluent Heun functions. We also show the existence of a Lorentz-like Lie symmetry $SO(2,1)$. This is analogous to the angular momentum symmetry but presenting one wrong sign in its preserved form. This invariant defines the relative half-life of the dynamical regime of the gene. The equivalent of the azimuth angular momentum measures indirectly the activity level of the gene. The ladder operators connect distinct stochastic processes of protein synthesis. The fluctuations of these processes are classified in terms of the relation between labeling numbers of a representation of the algebra. In the group theory formalism, the stochastic model to an externally regulated gene is a particular case of the model to an auto-regulated gene. We compare these two gene regulation strategies, and show that the auto-regulated gene can synthesize proteins into the super Poisson, Poisson, and sub Poisson fluctuating regimes. The externally regulated gene only presents the super Poisson and Poisson regimes. Therefore, the auto-regulation is responsible for a more precise control of gene expression. We also show that the dynamics of the auto-regulated genes has a shorter half-life. Thus, the auto-regulation permits faster responses of the system to external perturbation.

Keywords: Lie symmetries. Gene regulation. Stochastic methods. Special functions.

Sumário

1	Introdução	13
2	Um modelo estocástico para expressão gênica	19
2.1	A expressão gênica	19
2.1.1	Modelo simplificado	20
2.2	Modelo estocástico	21
2.2.1	Equações de Markov	21
2.2.2	O método das funções geratrizes	23
2.3	Solução na base de autofunções	25
3	Regime estacionário e simetrias	28
3.1	A teoria matemática das simetrias	28
3.2	O regime estacionário	32
3.2.1	Gene auto-regulado	32
3.2.2	Gene externamente regulado	34
3.3	Simetria de Lie	36
3.4	O significado biológico dos operadores algébricos	38
3.5	Domínio biológico dos parâmetros algébricos	38
3.5.1	Nível de atividade do gene	40
3.5.2	Classificação da estrutura das flutuações	42
4	Dinâmica estocástica	46
4.1	Solução geral em termos das funções de Heun	46
4.1.1	Probabilidades	48
4.1.2	Derivadas das autofunções	50
4.2	Dinâmica estocástica	51
5	Os genes auto e externamente regulados	54
5.1	Regime estacionário	55
5.1.1	Flutuações	55
5.1.2	Distribuições de probabilidades	56
5.2	Dinâmica	59
6	Conclusões	62
A	Derivação das autofunções	70
B	Álgebra de Lie de dimensão infinita	75

Capítulo 1

Introdução

Compreender como a informação genética armazenada no genoma é convertida em função celular é um dos problemas centrais da Biologia. Muito se avançou neste campo nos últimos 50 anos, e os mecanismos básicos da regulação do controle da transcrição do RNA, e conseqüentemente da síntese de proteínas em procariontes, foram caracterizados em experimentos com o *fago* λ [1] e o *operon-lac* [2]. No entanto, a despeito dos muitos dados gerados durante a era genômica, diversos fenômenos biofísicos ocorridos no interior celular permanecem incompreendidos. Um exemplo é a precisão espacial e temporal com que os genes são expressos no processo de segmentação de um embrião de *Drosophila* [3]. O estabelecimento e decodificação do código genético responde que proteína será sintetizada a partir de um gene, mas não quando e onde esta síntese manifestar-se-á. Em contraste com a beleza do código genético, universal e simétrico, os mecanismos de expressão gênica revelam uma enorme complexidade e forte correlação com a espécie investigada. A questão teórica central resume-se não somente à intrínseca complexidade da dinâmica da matéria viva, mas também à ausência de modelos quantitativos simples que possam ser utilizados como pontos-de-partida, e também blocos básicos, na construção de modelos mais detalhados.

Daí surge a motivação para o estudo de um gene isolado, [4, 5, 6]. Esta estratégia de decomposição de um sistema em suas componentes é bastante útil pois, estes componentes são constituídos por um número menor de elementos e executam uma função específica no

circuito gênico. Esta abordagem pode ser empregada experimentalmente, computacionalmente e teoricamente e deve contemplar a estocasticidade inerente ao interior celular. Esta aleatoriedade deve-se ao pequeno número de cópias de moléculas reagentes na célula e era prevista há mais de meio século [7].

Experimentalmente, uma rede gênica sintética ou natural é investigada pela detecção das intensidades luminosas de células em que proteínas fluorescentes são expressadas [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26]. As redes gênicas sintéticas podem ser construídas para o estudo da auto-regulação [8, 9, 10], o *genetic toggle* [11], o sistema *repressilator* [12], propagação de flutuações [13, 14], e assim por diante. Estas redes sintéticas também são úteis à classificação das fontes das flutuações em intrínsecas e extrínsecas [15, 16]. As primeiras devem-se à flutuações inerentes à ocorrência ou não de uma determinada reação química. As segundas são um cômputo das flutuações transmitidas pela variação no número de fatores de transcrição, polimerases, ribossomos, etc. Entre as redes gênicas naturais, dois sistemas são amplamente utilizados, especificamente, *lac operon* [25] e o *fago λ* [26], tidos como sistemas modelo no estudo da regulação gênica.

O avanço na investigação experimental quantitativa da expressão gênica motivou um aumento significativo na modelagem destas redes [27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45]. Simulações computacionais foram realizadas com auxílio do algoritmo de Gillespie [46]. Estas foram aplicadas com sucesso na investigação do circuito decisório do *fago λ* [27, 28]. A vantagem da aplicação desta técnica é o fornecimento de uma descrição da rede gênica, tão detalhada quanto requerida, em termos das reações químicas ocorridas. Por outro lado, essas simulações requerem longas horas de processamento e, para cada conjunto de parâmetros, uma nova simulação [47].

Aqui ressaltamos a importância de modelos que possuam solução analítica. De posse destas, rapidamente obtemos inferências acerca do comportamento quantitativo e qualitativo do sistema modelado. Estas soluções são obtidas a partir de equações macroscópicas, mesoscópicas ou microscópicas [48]. As primeiras lidam com equações para a concentração dos reagentes, as segundas abordam as probabilidades de ocorrência de um determinado

número de reagentes e as terceiras consideram equações de movimento para cada uma das moléculas reagentes. Modelos determinísticos foram aplicados para a descrição de experimentos com o *fago* λ [29] e com o *operon lac* [30], sem a inclusão do papel das flutuações. Estas, no entanto, foram introduzidas via técnica de Langevin [31]. Alternativamente, há sofisticções recentes desta metodologia em que os agentes causadores das flutuações são decompostos termo à termo [32, 49, 50].

Investigações em termos de equações mestras lidam principalmente com componentes abstratas de uma rede. Em geral estes oferecem conceitos gerais para o entendimento do papel das variações na síntese proteínica [33, 34, 35, 36, 37, 38, 51, 39, 40, 41, 42, 43, 44, 45]. Alguns destes trabalhos consideram apenas o desvio padrão calculado a partir de equações mestras [38, 51, 44, 45]. Tal abordagem mostra-se útil quando distribuições de probabilidade com um pico são estudadas. No entanto, mostrar-se-á imprecisa quando distribuições com múltiplos picos forem objeto de estudo. Uma descrição mais acurada requer distribuições de probabilidade explícitas, as quais podem ser obtidas analiticamente [37, 41, 43]. Nesta abordagem, as flutuações devidas à criação e degradação aleatórias dos mRNA e proteínas pode ser modelada com boa precisão. No entanto, fontes externas de flutuação como variações no número de RNA polimerases, indutores ou temperatura não são levados em conta. Neste contexto, um modelo que lida com a estocasticidade advinda dos processos de criação e degradação proteínica é tratado aqui [37].

Simetrias em biologia

Embora um sistema complexo *per se* [52], a teoria das simetrias revela-se bastante útil na abordagem de problemas físicos ou biológicos. Em Biologia, as simetrias baseadas na geometria de um organismo possuem uma longa história. Desde a classificação de Linnaeus, as simetrias são um aspecto essencial na taxonomia das flores. Neste esquema, a linaria (*Linaria vulgaris*) foi considerada em termos da simetria irregular de suas flores, que possuem um plano de invariância por reflexão. A descoberta de uma linaria regular, flores apresentando múltiplos planos de simetria, sugeria o aparecimento de uma nova, e completamente diferente, espécie. Linnaeus, porém, classificou-a como uma “transformação” da linaria comum e a chamou de *Peloria*, ‘monstro’ em grego. Esta hipótese levou

duzentos anos para ser comprovada, sob abordagem de Biologia Molecular. Um gene, *ci-clóidea*, apresenta um gradiente de ativação no botão da flor e, por isso, a flor desenvolve-se assimetricamente. Nos mutantes regulares, este gene é completamente reprimido ao longo do botão e isto possibilita um desenvolvimento uniforme da flor. Nas flores, de uma maneira geral, considera-se que aquelas que apresentam simetria regular são ancestrais às irregulares, o que mostra consequências evolutivas tanto das simetrias quanto das quebras das simetrias florais [53, 54, 55]. No caso das cápsulas virais, as simetrias originam-se do fato que estas estruturas estão em um mínimo de energia [56]. Este fato foi demonstrado com a constatação de que partículas virais isoladas agregam-se espontaneamente em forma de bastão, o que indica ser esta a estrutura estável da capa proteínica viral. Já a simetria icosaédrica deve-se à composição da capa viral por unidades idênticas, as quais apenas podem arranjar-se em alguma simetria cúbica [57].

Apesar da importância das simetrias em objetos, grupos discretos e cristalografia, dirigimo-nos às simetrias de um processo, ou simetrias dinâmicas, presentes nas equações que governam um fenômeno. Estas simetrias protagonizaram o desenvolvimento da física moderna. Em Física Atômica, a descrição completa do átomo de hidrogênio foi desenvolvida em termos das simetrias contínuas de Lie [58]. O trabalho seminal de Pauli [59] mostrou que a simetria rotacional de um sistema sob ação de uma força central não era suficiente para capturar as degenerescências no espectro de um problema Kepleriano. A adição do vetor de Runge-Lenz como uma quantidade conservada explica a degenerescência causada pelo vetor momento angular e estabelece a simetria ortogonal quadridimensional do átomo de hidrogênio. Outro aspecto importante da Teoria dos Grupos na Teoria Atômica advém da descrição de átomos multi-eletrônicos, sistemas reconhecidamente complexos se considerarmos a equação de Schrödinger do sistema completo. Ao invés disso, uma regra de composição entre funções de onda de átomos de hidrogênio foi encontrada para uma descrição aproximada do nível fundamental do sistema completo, chamado determinante de Slater [60]. A classificação das partículas elementares deve-se à conceitos de Teoria dos Grupos, propostos no “*Eightfold way*” independentemente, por Gell-mann e Ne’eman [61]. Aqui, um espaço vetorial abstrato têm seus elementos consid-

erados como partículas elementares. O grupo de simetrias que emerge desta classificação é o $SU(3)$ (*grupo especial unitário tridimensional*). As simetrias também são úteis em Física Molecular para a geração de espectros e amplitudes de transição [62, 63].

Um dos resultados importantes desta tese é considerar a aplicação das simetrias dinâmicas à modelagem estocástica da expressão gênica. A simetria de Lie $SO(2,1)$ foi desvelada nas equações mesoscópicas de um gene bimodal isolado, resultado publicado recentemente por nós [42]. As implicações biológicas deste fato são precisas: o invariante da álgebra caracteriza a meia-vida da dinâmica do gene; o nível de atividade do gene é caracterizado pelo “momento angular azimutal”; os operadores escada conectam diferentes processos estocásticos de expressão gênica; o regime de flutuações destes processos é caracterizado pela relação entre etiquetadores de uma representação; o modelo para um gene externamente regulado é um caso particular do modelo para um gene auto-regulado. A simetria descoberta possui interpretação biológica direta e, por isso, este ferramental possui aplicação potencial na área. Especificamente, na elaboração de uma regra de composição entre dois ou mais genes.

Funções de Heun

A equação de Heun é uma generalização das equações diferenciais de segunda ordem [64, 65, 66, 67, 68, 69]. Ela possui quatro singularidades e é bastante importante em física matemática. Isto pois, casos especiais dessa equação, incluem as equações hipergeométricas, as hipergeométricas confluentes, Lamé, Bessel, Legendre e Laguerre [65]. Estas funções são aplicadas em relatividade geral, na descrição de um buraco negro sob perturbações lineares [66]. As propriedades matemáticas desta função são objeto de intensa investigação [65, 67, 68, 69]. Nesta tese, as funções de Heun aparecem em sua forma confluyente, formando a base de autofunções em que expandimos a solução dependente do tempo do modelo estocástico para o gene auto-regulado. De posse destas soluções, obtivemos um exemplo de dinâmica estocástica desse gene. Para o cálculo das probabilidades, desenvolvemos uma relação de recorrência entre derivadas arbitrárias dessas funções.

Controle da regulação gênica

Um tema candente na busca por mecanismos moleculares subjacentes à evolução

das espécies é compreender o papel da regulação da expressão gênica neste processo [70, 71, 72]. O sequenciamento dos genomas revelou que o aumento no número de genes não implica em maior complexidade, *e. g.*, o verme (*C. elegans*) possui aproximadamente 20 mil genes enquanto a mosca-das-frutas (*Drosophila*) possui menos de 14 mil. Por seu turno, a mosca apresenta maior diversidade de células e tecidos. Aqui entra a regulação da expressão gênica, que pode estar no cerne de grande parte da diferenciação entre espécies e dentro destas. Nesse contexto, é mister analisar diferentes estratégias de regulação e seus efeitos na expressão gênica. Sendo assim, analisamos comparativamente os genes auto e externamente regulados. Mostramos, partindo de um modelo estocástico, que a auto-regulação permite síntese proteínica em regimes de baixa flutuação. Ou seja, resulta num controle mais estrito do número de proteínas produzidas pelo gene. Este resultado, embora já previsto na literatura e medido experimentalmente, não havia sido obtido a partir de equações estocásticas. Também verificamos que um gene auto-regulado pode ser mais rápido na resposta à perturbações que o tirem de um regime estacionário. Estes resultados também já estão presentes na literatura, porém baseados em modelos determinísticos.

Esta tese é dividida em quatro Capítulos. A divisão visa apresentar o modelo qualitativo e quantitativo que trataremos nesta tese, juntamente com sua solução completa, no início, Capítulo 2. Os Capítulos restantes são devotados à exploração destes resultados. O Capítulo 3 trata da investigação do papel da simetria de Lie do regime estacionário. Um exemplo de dinâmica probabilística é apresentado no Capítulo 4. No Capítulo 5 fazemos uma análise comparativa dos modelos estocásticos para os genes auto e externamente regulados. Por fim, no Capítulo 6, apresentamos nossas conclusões.

Capítulo 2

Um modelo estocástico para expressão gênica

Neste capítulo discorreremos acerca da expressão gênica. Apresentamos o modelo que consideramos para a síntese proteínica. Em seguida abordamos este modelo em termos de equações de Markov. Estas possuem soluções analíticas que são calculadas via método das funções geratrizes, tanto em regime estacionário quanto dinâmico. As soluções do modelo são obtidas ao escrevermos as equações mestras em formalismo de autovalores. Obtemos as autofunções, as quais são as funções de Heun confluentes. Por fim, escrevemos as soluções completas das equações de Markov como uma expansão sobre a base das autofunções.

2.1 A expressão gênica

Um modelo que vise abarcar a expressão gênica por completo, tal qual esta é compreendida atualmente, requer a consideração de diversos fenômenos [73]. Nesta tese utilizamos de uma versão muito simplificada deste modelo. Todavia, este esquema mostra-se suficiente para compreendermos o papel da regulação da transcrição gênica no controle da população de proteínas numa célula.

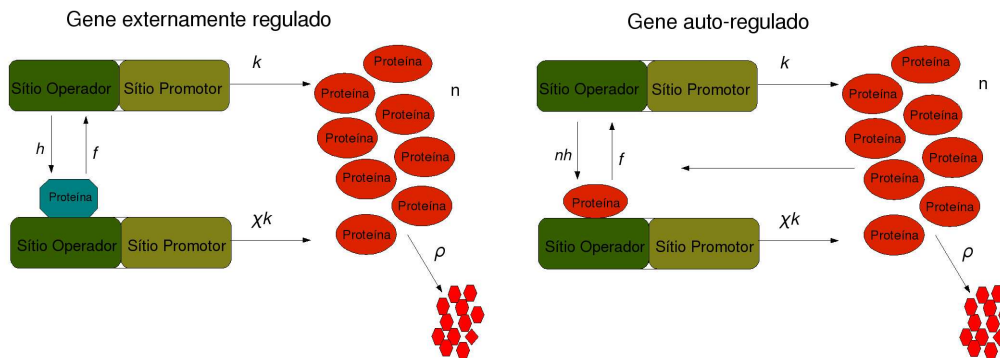


Figura 2.1: Um modelo simplificado para expressão gênica

2.1.1 Modelo simplificado

Nesta tese, quando utilizarmos o termo expressão gênica, teremos em mente o esquema apresentado na Figura 2.1. Nesta, o gene é composto por um sítio promotor e um sítio operador. Ao sítio promotor liga-se a RNAPolimerase (RNAP) para iniciar o processo de síntese proteínica. Ao sítio operador liga-se uma proteína reguladora do processo de síntese proteínica. Também consideraremos que os processos de transcrição e tradução são acoplados e chamá-lo-emos simplesmente de síntese proteínica. Isto implica considerar que as proteínas são criadas diretamente a partir do gene, sem que exista o mRNA como agente intermediário da expressão do gene. Convencionaremos que o gene estará em estado reprimido se houver uma proteína reguladora ligada ao sítio operador. Caso contrário, o gene estará em estado ativo.

A taxa de síntese proteínica depende de o gene estar em estado ativo ou reprimido e é, respectivamente, k ou χk , onde k é um número real positivo e χ é um número real entre zero e um. A degradação das proteínas independe do estado do gene e ocorre com taxa ρ real positiva. O número de proteínas sintetizadas pelo gene é dado por n e é um inteiro positivo.

A proteína reguladora da expressão do gene pode fazê-lo negativa ou positivamente. A regulação será negativa se a presença de proteínas reguladoras induzir o gene a expressar em estado reprimido. Caso contrário a regulação será positiva. Nesta tese trataremos a regulação como negativa. O leitor interessado em analisar resultados referentes à regu-

lação positiva pode obtê-lo diretamente dos resultados que apresentaremos se inverter a interpretação dos estados do gene: ativo torna-se reprimido e vice-versa.

A proteína reguladora da expressão do gene pode ser codificada por ele mesmo. Neste caso o gene é chamado auto-regulado. Como escolhemos a regulação negativa, a taxa de ligamento das proteínas ao sítio operador é proporcional ao número destas na célula, nh , sendo h um parâmetro de proporcionalidade. A taxa de desligamento da proteína reguladora do sítio operador é dada por \tilde{f} . Tanto \tilde{f} quanto h são números reais positivos. Vale advertir que a dependência funcional em n pode ser não-linear. Por exemplo, há autores que adotam dependência quadrática [44].

Quando a proteína reguladora é expressa por outro gene, este é chamado externamente regulado. A taxa de ligamento da proteína ao sítio operador não apresenta dependência funcional em n e é dada por h . A taxa de desligamento é dada por \tilde{f} . Novamente \tilde{f} e h são números reais positivos. Estes parâmetros podem apresentar dependência funcional em outras variáveis que não o número de proteínas expressas pelo gene ao qual se referem, conforme foi adotado em [43].

2.2 Modelo estocástico

Apresentamos um modelo estocástico para um gene auto-regulado e/ou sob regulação externa. O modelo conterá os dois processos aleatórios da seção anterior, um deles representando criação e destruição de proteínas. O segundo representa a mudança aleatória entre os estados ativo e reprimido do gene que, nesta tese, designaremos como chaveamento gênico.

2.2.1 Equações de Markov

O sistema mostrado na seção anterior pode ser modelado estocasticamente em termos de um processo de Markov de um passo. Neste caso, a variável estocástica varia pela unidade, com o tempo, à taxas definidas. Probabilidades dependentes do tempo são associadas aos valores permitidos à variável aleatória. A probabilidade de o sistema estar

num estado varia positivamente com o tempo se, ocorrer uma transformação que traga o sistema para este estado. Caso contrário, se a transformação tirar o sistema deste estado, a probabilidade varia negativamente.

Nas equações do modelo, o número de proteínas livres na célula, n , e os estados do gene, ativo e reprimido, são tratados como as variáveis aleatórias. A probabilidade de o gene codificador desta proteína estar em estado ativo quando n proteínas são encontradas dentro da célula é α_n . A probabilidade de o gene estar reprimido é β_n . As equações do modelo são:

$$\frac{d\alpha_n}{dt} = k[\alpha_{n-1} - \alpha_n] + \rho[(n+1)\alpha_{n+1} - n\alpha_n] - (h_1n + h_2)\alpha_n + \tilde{f}\beta_n, \quad (2.1)$$

$$\frac{d\beta_n}{dt} = \chi k[\beta_{n-1} - \beta_n] + \rho[(n+1)\beta_{n+1} - n\beta_n] + (h_1n + h_2)\alpha_n - \tilde{f}\beta_n, \quad (2.2)$$

onde as constantes k e χk são, respectivamente, as taxas de criação de proteínas quando o gene assume os estados ativo e reprimido. As proteínas são degradadas com taxa ρ . A taxa com que uma proteína reguladora desliga-se do sítio operador é \tilde{f} . As taxas h_1 e h_2 representam taxas de ligamento das proteínas reguladoras aos sítios operadores. No gene regulado externamente, esta taxa não apresenta dependência em n , portanto escolhemos h_1 nulo. No gene auto-regulado a taxa de ligamento da proteína reguladora ao sítio operador é dependente do número de proteínas expressas pelo gene. E, conforme afirmado anteriormente, escolhemos dependência linear em n e a taxa h_2 nula. Em caso de h_2 não nulo, o gene será auto e externamente regulado, como pode ocorrer em organismos mais complexos [74].

As equações (2.1) e (2.2) representam os processos químicos efetivos representados na Tabela 2.1 abaixo. Neste caso, dois processos de criação e degradação são acoplados através do chaveamento gênico. Os processos de criação e degradação são modelados em termos de distribuições de Poisson e representam os dois modos de síntese de proteínas. Ou seja, no regime estacionário, o número médio de produtos da reação é dado por, respectivamente, k/ρ e $\chi k/\rho$ para os estados ativo e reprimido do gene. O acoplamento entre estes dois eventos Poissonianos é dado em termos das taxas \tilde{f} e h_i . A mudança de esta-

dos aleatória entre estados de síntese proteínica corresponde ao chaveamento gênico. Um sistema que transita fortuitamente entre dois estados é chamado, em teoria dos métodos estocásticos, de processo do telégrafo aleatório [75].

Tabela 2.1: Processos químicos			
Criação e degradação de proteínas			
Reação	Símbolo	Taxa	Estado do gene
Degradação proteínica	$\mathcal{M} \rightarrow \emptyset$	ρ	Irrelevante
Síntese proteínica	$\emptyset \rightarrow \mathcal{M}$	k	Livre
		χk	Reprimido
Chaveamento gênico			
Reação	Símbolo	Taxa	Regulação
Desligamento	$\mathcal{O}_\beta \rightarrow \mathcal{O}_\alpha$	\tilde{f}	Irrelevante
Ligamento	$\mathcal{O}_\alpha \rightarrow \mathcal{O}_\beta$	$h_1 n$	Auto
		h_2	Externa

O acoplamento de dois processos aleatórios distintos induz a existência de duas taxas de decaimento para o equilíbrio. Uma advinda dos processo de criação e degradação proteínica, dada por ρ . A segunda taxa, devida ao chaveamento gênico, é $\tilde{f} + h_2$ para um gene sob regulação externa [75] e $\tilde{f} + h_1 k / (\rho + h_1)$ para o caso auto-regulado, conforme calculado recentemente por nós [42]. Note que a taxa de decaimento de um gene auto-regulado possui dependência com as taxas de criação e degradação. Isto é causado pela dependência em n da taxa de ligamento, o que leva a um acoplamento mais forte entre os dois processos aleatórios modelados.

2.2.2 O método das funções geratrizes

As equações de diferença-diferenciais acopladas, (2.1) e (2.2), podem ser transformadas nas seguintes equações diferenciais parciais acopladas:

$$\frac{\partial \alpha}{\partial t} = (z - 1) \left(k\alpha - \rho \frac{\partial \alpha}{\partial z} \right) - h_1 z \frac{\partial \alpha}{\partial z} - h_2 \alpha + \tilde{f} \beta, \quad (2.3)$$

$$\frac{\partial \beta}{\partial t} = (z - 1) \left(\chi k \beta - \rho \frac{\partial \beta}{\partial z} \right) + h_1 z \frac{\partial \alpha}{\partial z} + h_2 \alpha - \tilde{f} \beta, \quad (2.4)$$

que resultam da aplicação de uma generalização do método das funções geratrizes. Neste formalismo, as probabilidades são tratadas como coeficientes de uma expansão de Taylor de uma função analítica na variável z , explicitamente:

$$\alpha(z, t) = \sum_{n=0}^{\infty} \alpha_n(t) z^n, \quad \beta(z, t) = \sum_{n=0}^{\infty} \beta_n(t) z^n. \quad (2.5)$$

Na formulação original desta metodologia, a função geratriz efetuada em z igual à um, corresponde à soma de todas as probabilidades e resulta, portanto, na unidade. Na formulação aqui empregada, tal substituição resulta na probabilidade total de o gene estar ativo, $\alpha(1, t)$, ou reprimido, $\beta(1, t)$. A soma destas probabilidades resulta na unidade:

$$\alpha(1, t) + \beta(1, t) = 1. \quad (2.6)$$

No problema aqui tratado, a probabilidade total de o gene estar ativo pode ser entendida como a fração de tempo que o gene assume configuração ativa, num longo intervalo de tempo e dentre vários chaveamentos entre seus estados.

As probabilidades e os momentos de n -ésima ordem do processo aleatório definido pelas equações (2.1) e (2.2) são obtidos por diferenciação e efetuação das funções geratrizes nos pontos $z = 0$ and $z = 1$, respectivamente. As probabilidades são:

$$\alpha_n = \frac{1}{n!} \left. \frac{\partial^n \alpha}{\partial z^n} \right|_{z=0}, \quad \beta_n = \frac{1}{n!} \left. \frac{\partial^n \beta}{\partial z^n} \right|_{z=0},$$

e as médias:

$$\langle n_\alpha^p \rangle = \sum_{n=0}^{\infty} \alpha_n n^p = \left[\left(z \frac{\partial}{\partial z} \right)^p \alpha \right]_{z=1}, \quad \langle n_\beta^p \rangle = \sum_{n=0}^{\infty} \beta_n n^p = \left[\left(z \frac{\partial}{\partial z} \right)^p \beta \right]_{z=1}.$$

Tais definições são verificadas por substituição direta das definições 2.5.

2.3 Solução na base de autofunções

Nesta seção limitamo-nos ao cálculo da solução dependente do tempo para o modelo do gene auto-regulado. As equações para um gene externamente regulado já foram solucionadas em trabalhos anteriores. Sendo assim, escolhemos h_2 nulo. A solução do modelo será calculada por meio de expansão na base de autofunções do operador dependente de z . Apresentamos aqui apenas os passos seguidos para este cálculo, sendo os detalhes apresentados no Apêndice A.

A solução dinâmica do modelo é escrita explicitamente para as funções geratrizes dos estados ativo e reprimido, conforme se segue:

$$\alpha(z, t) = \sum_{j=0}^{\infty} \sum_{i=1}^2 C_j^{(i)} A_j^{(i)}(z) e^{-\lambda_j^{(i)} t}, \quad (2.7)$$

$$\beta(z, t) = \sum_{j=0}^{\infty} \sum_{i=1}^2 D_j^{(i)} B_j^{(i)}(z) e^{-\lambda_j^{(i)} t}, \quad (2.8)$$

onde $C_j^{(i)}$ e $D_j^{(i)}$ são coeficientes da expansão em série. As autofunções de \mathbb{H} são dadas por $A_j^{(i)}$ e $B_j^{(i)}$. Seus correspondentes autovalores são dados por $\lambda_j^{(i)}$. Os autovalores são calculados a partir da imposição da analiticidade às autofunções e formam dois conjuntos distintos:

$$\lambda_j^{(1)} = j\rho, \quad \text{e} \quad \lambda_j^{(2)} = (j+b)\frac{\rho}{z_0}, \quad \text{em que } j = 0, 1, 2, \dots, \quad (2.9)$$

e $b = f + M$. Os autovalores do conjunto 1 garantem analiticidade em torno do ponto z igual à unidade. Os autovalores do conjunto 2 garantem analiticidade em torno de z_0 .

As autofunções são expressas em termos das funções de Heun confluentes:

$$A_j^{(1)}(z) = \mathcal{H} \left(\Delta_1, -jz_0 + b, -1 - n, \Delta_2(\Lambda_j^{(1)}), \Delta_3(\Lambda_j^{(1)}), \frac{z - z_0}{1 - z_0} \right), \quad (2.10)$$

$$A_j^{(2)}(z) = (z - z_0)^j \mathcal{H} \left(\Delta_1, j, -1 - (j+b)/z_0, \Delta_2(\Lambda_j^{(2)}), \Delta_3(\Lambda_j^{(2)}), \frac{z - z_0}{1 - z_0} \right), \quad (2.11)$$

onde \mathcal{H} representa a função de Heun confluyente. Os parâmetros Δ_i são definidos em

termos dos parâmetros do modelo:

$$\begin{aligned}\Delta_1 &= (\chi - z_0) \frac{M}{z_0}, \\ \Delta_2(\Lambda_j^{(i)}) &= \frac{\Delta_1}{2} \left(2a + 1 - b - \frac{1 + z_0}{z_0} \Lambda_j^{(i)} \right), \\ \Delta_3(\Lambda_j^{(i)}) &= \frac{\Delta_1 \left(b - 2a - 1 + \Lambda_j^{(i)} \right) + 1 + \Lambda_j^{(i)} \left(\Lambda_j^{(i)} + 1 - b \right) / z_0}{2},\end{aligned}$$

onde $a = f(1 - \chi)/(z_0 - \chi)$ e $\Lambda_j^{(i)} = \lambda_j^{(i)} z_0 / \rho$. As soluções apresentadas acima são obtidas escrevendo-se as equações (2.3) e (2.4) como segue-se:

$$\frac{z_0}{\rho} \frac{\partial \psi}{\partial t} = \mathbb{H} \psi, \quad (2.12)$$

onde

$$\psi = \begin{bmatrix} \alpha \\ \beta \end{bmatrix}, \text{ e } \mathbb{H} = \begin{bmatrix} M \frac{z-1}{1-z_0} - (z-z_0) \frac{\partial}{\partial z} & f \\ (1-z_0)z \frac{\partial}{\partial z} & \chi M \frac{z-1}{1-z_0} - f - z_0(z-1) \frac{\partial}{\partial z} \end{bmatrix}, \quad (2.13)$$

e os novos parâmetros do modelo são definidos como:

$$z_0 = \frac{\rho}{\rho + h_1}, \quad M = k \frac{z_0}{\rho} (1 - z_0), \quad f = \tilde{f} \frac{z_0}{\rho}.$$

Note que a solução da equação (2.12) é proporcional à exponencial do operador \mathbb{H} , $\exp(\mathbb{H}t)$, e pode ser expandida na base de autofunções deste operador:

$$\sum_j e^{-\lambda_j t} \tilde{\psi}_j(z), \quad (2.14)$$

onde λ_j e $\tilde{\psi}_j$ são, respectivamente, autovalores e auto-vetores de \mathbb{H} . O cálculo de ambos advém da solução da equação de autovalores abaixo:

$$\mathbb{H} \tilde{\psi} = -\lambda \tilde{\psi}. \quad (2.15)$$

As expressões das soluções do modelo, mostradas nas equações (2.7) e (2.8), são obtidas

pelos cálculo explícito de $\tilde{\psi}$ e λ .

Baseando-nos na equação (2.9), verificamos que o modelo apresenta um espectro de decaimentos para o regime estacionário, quando as probabilidades tornam-se constantes com o tempo. O modo fundamental, em que n é nulo, implica em $\lambda_0^{(1)}$ nulo e $\lambda_0^{(2)} = b\rho/z_0$. Sendo assim, a componente da solução proporcional à $A_0^{(1)}$ possui derivada temporal nula e representa a solução do estado estacionário. Podemos mostrar que o regime estacionário satisfaz a simetria de Lie $SO(2,1)$ e que o invariante da álgebra é proporcional à $\lambda_0^{(2)}$. A exploração destes dois resultados é apresentada no próximo capítulo. A análise do espectro apresentado na equação (2.9) é tema do capítulo posterior.

Em suma, neste capítulo apresentamos um modelo para a expressão de um gene que inclui os fenômenos de criação e degradação proteínica, bem como o chaveamento gênico entre dois níveis de expressão. Este modelo baseia-se em equações de Markov e apresenta solução analítica para os regimes estacionário e dinâmico. As soluções são escritas por expansão, numa base de autofunções, cujos elementos são as funções de Heun confluentes.

Capítulo 3

Regime estacionário e simetrias

De posse das soluções gerais do modelo estocástico aqui tratado, analisamos o regime estacionário. As soluções deste regime são dadas em termos das funções de Kummer. As equações a partir das quais as soluções estacionárias são calculadas obedecem às relações de comutação da álgebra de Lie tipo-Lorentz, $SO(2, 1)$. O invariante desta simetria apresenta um sinal invertido em relação ao invariante da álgebra das rotações. No modelo aqui tratado, o significado do invariante é preciso, ele é a taxa de decaimento para o equilíbrio do chaveamento gênico. As autofunções da álgebra são etiquetadas por dois números. As flutuações no número de proteínas expressas pelo gene são caracterizadas pela relação entre esses números.

3.1 A teoria matemática das simetrias

Apresentamos informal e brevemente os conceitos de Teoria dos Grupos necessários ao entendimento do restante deste capítulo. Leitores interessados em um tratamento mais formal são dirigidos às referências [76, 77, 78].

O grupo das rotações tridimensionais é o protótipo de um grupo contínuo, ou grupo de Lie. Ele é composto pelas matrizes ortogonais \mathbf{g} , as quais satisfazem: $\mathbf{g}^T = \mathbf{g}^{-1}$. Tal propriedade garante a preservação de ângulos e magnitudes dos vetores rotacionados. Um conjunto de objetos é chamado grupo se soubermos como multiplicar e inverter seus

elementos. A inversa e a identidade também são elementos do conjunto. No caso de uma matriz rotação, sua forma geral é algo complicado envolvendo ângulos de Euler ou outra família de parâmetros.

O conjunto das matrizes anti-simétricas \mathbf{X} é bastante próximo às matrizes ortogonais: a exponencial de uma matriz anti-simétrica é uma matriz de rotação, explicitamente:

$$\mathbf{e}^{\mathbf{X}} = \mathbf{1} + \mathbf{X} + \frac{\mathbf{X}^2}{2!} + \dots \quad (3.1)$$

O conjunto das matrizes anti-simétricas possui uma estrutura bastante simples: elas formam um espaço vetorial. A soma de duas matrizes anti-simétricas resulta numa matriz anti-simétrica. Contudo não podemos multiplicar duas matrizes anti-simétricas, dado que isto não resultará num elemento do conjunto inicial, uma propriedade muito interessante destas matrizes emerge ao calcularmos seu comutador, definido como:

$$[X, Y] = XY - YX. \quad (3.2)$$

Podemos checar que o comutador de duas matrizes anti-simétricas é uma matriz anti-simétrica. Devido às propriedades mencionadas, o conjunto das matrizes anti-simétricas é chamado uma álgebra de Lie.

Resumindo, temos um espaço vetorial n -dimensional (o espaço carregador) e dois conjuntos de matrizes $n \times n$ agindo sobre ele. O primeiro, o grupo, fecha sob multiplicação e o último, que fecha sob adição, é a álgebra. Estes conjuntos relacionam-se pelo mapa exponencial. A álgebra de Lie é manipulada com mais facilidade e possui a multiplicação de Lie: o comutador.

Em muitas das aplicações em ciência desejamos lidar com funções ao invés de vetores n -dimensionais. O espaço de funções também é um espaço vetorial com a propriedade de adição usual, porém sua dimensão é infinita. Ao invés de um problema, isto é uma vantagem: temos uma estrutura mais rica, vários subespaços e a possibilidade de tomar derivadas, integrais, gráficos e assim sucessivamente. A álgebra de Lie correspondente

deve ser formada por operadores diferenciais que “rotacionam” as funções.

Podemos representar as rotações tridimensionais através das matrizes:

$$\mathbf{L}_x = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ 0 & i & 0 \end{pmatrix}, \mathbf{L}_y = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{pmatrix}, \mathbf{L}_z = \begin{pmatrix} 0 & -i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (3.3)$$

que obedecem

$$[\mathbf{L}_x, \mathbf{L}_y] = i\mathbf{L}_z, \quad [\mathbf{L}_y, \mathbf{L}_z] = i\mathbf{L}_x, \quad [\mathbf{L}_z, \mathbf{L}_x] = i\mathbf{L}_y, \quad (3.4)$$

sendo i a unidade imaginária.

As matrizes na equação (3.3) agem sobre o espaço tridimensional. Também podemos considerar os operadores diferenciais

$$\begin{aligned} \tilde{L}_x &= i \left(z \frac{\partial}{\partial y} - y \frac{\partial}{\partial z} \right), \\ \tilde{L}_y &= i \left(x \frac{\partial}{\partial z} - z \frac{\partial}{\partial x} \right), \\ \tilde{L}_z &= i \left(y \frac{\partial}{\partial x} - x \frac{\partial}{\partial y} \right), \end{aligned} \quad (3.5)$$

obedecendo as mesmas relações de comutação da equação (3.4) mas que agem sobre o espaço das funções diferenciáveis nas variáveis x, y, z .

Nesta tese trabalharemos com uma álgebra tridimensional similar à álgebra das rotações. Esta álgebra de Lie é definida pelos operadores diferenciais:

$$\begin{aligned} L_x &= \eta \frac{d^2}{d\eta^2} + (2l + 2 - \theta\eta) \frac{d}{d\eta} + \frac{\theta^2 + 1}{4} \eta - \theta(1 + l), \\ L_y &= -i\eta \frac{d}{d\eta} + i \frac{\theta\eta - 2 - 2l}{2}, \\ L_z &= \eta \frac{d^2}{d\eta^2} + (2l + 2 - \theta\eta) \frac{d}{d\eta} + \frac{\theta^2 - 1}{4} \eta - \theta(1 + l), \end{aligned} \quad (3.6)$$

onde η é uma variável complexa e l e θ são constantes reais. Estes operadores fecham sob as seguintes relações de comutação:

$$[L_x, L_y] = -iL_z, \quad [L_y, L_z] = iL_x, \quad [L_z, L_x] = iL_y. \quad (3.7)$$

A diferença entre esta álgebra, chamada $so(2, 1)$, e a das rotações, $so(3)$, é somente um sinal nas relações de comutação, porém com consequências profundas. Elas possuem a mesma relação que o grupo de rotação quadridimensional e o grupo de Lorentz.

Os operadores L_x , L_y e L_z podem ser escritos na forma de Cartan-Weyl. Nós os combinamos como segue-se, $L_{\pm} = L_x \pm iL_y$, tal que:

$$L_+ = \eta \frac{d^2}{d\eta^2} + [2l + 2 + (1 - \theta)\eta] \frac{d}{d\eta} + \frac{(1 - \theta)^2}{4} \eta + (1 - \theta)(1 + l), \quad (3.8)$$

$$L_- = \eta \frac{d^2}{d\eta^2} + [2l + 2 - (1 + \theta)\eta] \frac{d}{d\eta} + \frac{(1 + \theta)^2}{4} \eta - (1 + \theta)(1 + l),$$

e os nomeamos operadores escada, onde L_+ (L_-) é o operador levantamento (abaixamento).

Neste caso, as relações de comutação tornam-se:

$$[L_z, L_{\pm}] = \pm L_{\pm}, \quad [L_+, L_-] = -2L_z. \quad (3.9)$$

A sub-álgebra de Cartan é formada pelo operador L_z e sua ação sobre qualquer outro elemento da álgebra é proporcional ao próprio elemento, conforme podemos notar na equação (3.9).

Também definimos um operador que comuta com todos os operadores da álgebra, o operador de Casimir. Um operador com tal característica é importante pois será proporcional à identidade, segundo o lema de Schur [76]. No caso da álgebra $so(2, 1)$, ele é definido como:

$$\mathbb{C} = -L_z^2 + L_x^2 + L_y^2. \quad (3.10)$$

3.2 O regime estacionário

A componente estacionária do processo estocástico definido pelas equações (2.1) e (2.2) pode ser obtida tomando-se a componente $n = 0$ da solução geral do modelo, equações (2.7) e (2.8). Alternativamente, podemos focar-nos apenas ao caso em que as derivadas temporais das equações (2.3) e (2.4) são nulas. Isto corresponde a tomarmos $\Lambda = 0$ na equação de autovalores de \mathcal{H} (A.7). Como visamos verificar que a equação do regime estacionário satisfaz a simetria de Lie apresentada anteriormente, $so(2, 1)$, optamos por escrever as equações do modelo:

$$\left(M \frac{z-1}{1-z_0} - \frac{h_2}{\rho+h_1} \right) \alpha + (z_0 - z) \frac{d\alpha}{dz} + f\beta = 0, \quad (3.11)$$

$$\left(\chi M \frac{z-1}{1-z_0} - f \right) \beta - z_0(z-1) \frac{d\beta}{dz} + (1-z_0)z \frac{d\alpha}{dz} - \frac{h_2}{\rho+h_1} \alpha = 0, \quad (3.12)$$

onde

$$z_0 = \frac{\rho}{\rho+h_1}, \quad M = k \frac{z_0}{\rho} (1-z_0), \quad f = \tilde{f} \frac{z_0}{\rho}.$$

Para cada uma das estratégias de regulação gênica reservamos uma subsecção, dadas as particularidades de interpretação dos modelos.

3.2.1 Gene auto-regulado

Nesta subsecção tomamos h_2 nulo. A solução das equações (3.11) e (3.12) é obtida resolvendo-se a expressão (3.11) para β , o que resulta em:

$$\beta = \frac{M}{f} \frac{1-z}{1-z_0} \alpha + \frac{z-z_0}{f} \frac{d\alpha}{dz}. \quad (3.13)$$

Substituímos a expressão acima na equação (3.12), a expandimos e, dividindo-a pelo fator:

$$\frac{M}{f} \frac{z_0}{1-\chi/z_0} \frac{1-z}{1-z_0},$$

obtemos uma equação diferencial de segunda ordem, conforme escrito abaixo:

$$\eta \frac{d^2 \alpha}{d\eta^2} + \left(1 + b - \frac{z_0 + \chi}{z_0 - \chi} \eta\right) \frac{d\alpha}{d\eta} + \left[\frac{z_0 \chi}{(z_0 - \chi)^2} \eta + \frac{z_0 + \chi}{z_0 - \chi} \frac{1 + b}{2} \right] \alpha - \left(a + \frac{1 - b}{2}\right) \alpha = 0. \quad (3.14)$$

A nova variável η é escrita como função de z ,

$$\eta = \frac{M}{z_0} \frac{z_0 - \chi}{1 - z_0} (z - z_0) \quad (3.15)$$

e os parâmetros a e b são escritos como função dos parâmetros do modelo:

$$a = \frac{1 - \chi}{z_0 - \chi} f, \quad e \quad b = f + M. \quad (3.16)$$

A equação (3.14) apresenta duas singularidades regulares, uma na origem e a segunda no infinito. Isto sugere uma solução em termos das funções hipergeométricas confluentes. De fato, a solução é escrita como:

$$\alpha(\eta) = c^{-1} \frac{a}{b} \frac{z_0 - \chi}{1 - \chi} \exp\left(\chi \frac{\eta - \eta_1}{z_0 - \chi}\right) \mathcal{M}(1 + a, 1 + b, \eta), \quad (3.17)$$

onde,

$$c = \mathcal{M}(a, b, \eta_1), \quad e \quad \eta_1 = M(z_0 - \chi)/z_0. \quad (3.18)$$

O símbolo \mathcal{M} representa a função de Kummer M. O termo η_1 é obtido ao substituirmos z igual a unidade na definição da variável η , equação (3.15). A função de Kummer M é definida em termos da série:

$$\mathcal{M}(a, b, \theta) = \sum_{n=0}^{\infty} \frac{(a)_n \theta^n}{(b)_n n!}, \quad (3.19)$$

onde $(x)_n$ representa a função de Pochhammer:

$$(x)_n = x(x + 1)(x + 2) \dots (x + n - 1), \quad (x)_0 = 1. \quad (3.20)$$

A componente irregular da solução da equação (3.14), escrita em termos da função de Kummer U , é eliminada. Esta é uma vantagem das soluções analíticas frente às numéricas. Podemos eliminar divergências diretamente nas primeiras enquanto nas segundas buscamos soluções que estejam no intervalo de validade.

A função geratriz das probabilidades de o gene estar em estado reprimido é obtida pela ação da equação (3.13) sobre α e considerando a relação 13.4.14, contida na referência [79]. Escrevemo-na como:

$$\beta(\eta) = c^{-1} \exp\left(\chi \frac{\eta - \eta_1}{z_0 - \chi}\right) \mathcal{M}(a, b, \eta) - \alpha(z). \quad (3.21)$$

Também definimos a probabilidade de n proteínas serem encontradas dentro da célula, independente do estado do gene, como $\alpha_n + \beta_n = \phi_n$. Neste caso a função geratriz é obtida diretamente da soma de $\alpha + \beta$, resultando em:

$$\phi(\eta) = c^{-1} \exp\left(\chi \frac{\eta - \eta_1}{z_0 - \chi}\right) \mathcal{M}(a, b, \eta), \quad (3.22)$$

note que em η igual a η_1 , a função geratriz acima é igual a unidade. Ou seja, esta função está normalizada em termos de c .

3.2.2 Gene externamente regulado

A solução estacionária do modelo para um gene externamente regulado é obtida seguindo-se algoritmo similar ao empregado ao modelo para o gene auto-regulado. Inicialmente devemos impor h_1 nulo. A equação de segunda ordem para α será indêntica à equação (3.14), se: impusermos z_0 igual a unidade e a variável η for definida em termos de z , como segue-se,

$$\eta = \frac{k}{\rho}(1 - \chi)(z - 1). \quad (3.23)$$

Note que, neste caso, η_1 é nulo. Os parâmetros a e b são redefinidos como:

$$b = \frac{\tilde{f} + h_2}{\rho}, \quad \text{e} \quad a = \frac{\tilde{f}}{\rho}. \quad (3.24)$$

Os parâmetros do modelo podem ser redefinidos em termos das características dos processos estocásticos modelados:

$$N = \frac{k}{\rho}, \quad p_\alpha = \frac{\tilde{f}}{\tilde{f} + h_2} = \frac{a}{b}, \quad p_\beta = 1 - p_\alpha, \quad \text{e } \epsilon = b, \quad (3.25)$$

em que N é o número médio de proteínas produzidas pelo gene em estado ativo no regime estacionário, conforme discutido na Seção (2.2.1), à página 22.

Os parâmetros p_α ou p_β são, respectivamente, as probabilidades de o gene estar ativo ou reprimido no regime estacionário. Isto pode ser verificado se consideramos as equações:

$$\begin{aligned} -h_2\alpha(1) + \tilde{f}\beta(1) &= 0, \\ \alpha(1) + \beta(1) &= 1, \end{aligned} \quad (3.26)$$

as quais foram obtidas considerando-se z igual a unidade na equação (3.11) e a conservação da probabilidade total, conforme a equação (2.6). Note que h_1 é nulo. Resolvemos o sistema de equações acima considerando a notação $p_x \equiv x(1)$ e recuperamos a expressão apresentada nas equações (3.25).

O parâmetro ϵ é uma quantidade adimensional, que relaciona as duas taxas de decaimento para o regime estacionário dos processos estocásticos aqui modelados. A taxa de decaimento dos processos de criação e degradação proteínica é dada por ρ . Já o processo de chaveamento gênico possui taxa de decaimento dada por $\epsilon\rho$.

As funções geratrizes das probabilidades do gene externamente regulado são:

$$\begin{aligned} \alpha(\eta) &= \frac{a}{b} \exp\left(\frac{\chi\eta}{1-\chi}\right) \mathcal{M}(1+a, 1+b, \eta), \\ \beta(\eta) &= \frac{b-a}{b} \exp\left(\frac{\chi\eta}{1-\chi}\right) \mathcal{M}(a, 1+b, \eta), \\ \phi(\eta) &= \exp\left(\frac{\chi\eta}{1-\chi}\right) \mathcal{M}(a, b, \eta), \end{aligned} \quad (3.27)$$

e foram obtidas igualando z_0 à unidade nas funções geratrizes das probabilidades do

gene auto-regulado. A constante c foi à unidade, pois a função Kummer M , quando seu argumento é nulo, é igual a um. A função β foi escrita considerando a transformação número 13.4.3 da referência [79].

3.3 Simetria de Lie

A existência de soluções analíticas para a equação (3.14) e a similaridade entre o operador diferencial que age sobre α e o operador L_z , definido na equação (3.6), nos direciona à busca de uma simetria de Lie no modelo estocástico aqui tratado [42, 80, 81, 76]. De fato, a ação do operador L_z sobre sua autofunção pode ser escrita como:

$$L_z \alpha = m \alpha, \quad (3.28)$$

em que α é a função geratriz das probabilidades de o gene estar em estado ativo e as equações (3.14) e (3.28) são idênticas se impusermos:

$$l = \frac{b-1}{2}, \quad \text{e} \quad m = a + (1-b)/2. \quad (3.29)$$

A simetria considerada será referente ao modelo de um gene auto ou externamente regulado de acordo com a forma da constante θ . Num gene auto-regulado ela é:

$$\theta = \frac{z_0 + \chi}{z_0 - \chi}, \quad (3.30)$$

enquanto no gene externamente regulado ela fica:

$$\theta = \frac{1 + \chi}{1 - \chi}. \quad (3.31)$$

Dessarte, concluímos que a função geratriz das probabilidades do estado ativo do gene é autofunção da sub-álgebra de Cartan da simetria $SO(2,1)$. A ação do operador levantamento sobre a autofunção leva-a noutra autofunção cujo parâmetro a vai em $a+1$.

Por sua vez, sob a ação do operador abaixamento, o parâmetro a da autofunção vai em $a - 1$. Explicitamente, num caso geral:

$$L_+ \exp\left(\frac{\theta - 1}{2}\eta\right) \mathcal{M}(a, 1 + b, \eta) = a \exp\left(\frac{\theta - 1}{2}\eta\right) \mathcal{M}(1 + a, 1 + b, \eta), \quad (3.32)$$

$$L_- \exp\left(\frac{\theta - 1}{2}\eta\right) \mathcal{M}(1 + a, 1 + b, \eta) = (a - b) \exp\left(\frac{\theta - 1}{2}\eta\right) \mathcal{M}(a, 1 + b, \eta). \quad (3.33)$$

Por conseguinte, podemos gerar um conjunto completo de funções de geratrizes, cujos parâmetros a diferem pela unidade, através da ação dos operadores escada. Cada uma das regulações é recuperada pela substituição de seu respectivo θ .

O operador de Casimir pode ser escrito como segue:

$$\mathbb{C} = -L_z^2 + L_z + L_+ L_-. \quad (3.34)$$

Sua ação sobre a função geratriz do estado ativo, α , possui como autovalor a expressão canônica [76]:

$$\mathbb{C}\alpha = -l(l + 1)\alpha. \quad (3.35)$$

Em termos do parâmetro b , obtemos a ação do operador de Casimir com o auxílio das equações (3.32) e (3.33), e seu autovalor fica:

$$\mathbb{C}\alpha = \frac{1 - b^2}{4}\alpha. \quad (3.36)$$

O parâmetro b é um número real positivo, como podemos verificar de suas definições nas equações (3.16) e (3.24). Portanto, o autovalor do operador de Casimir será um número real menor que $1/4$ e o número l satisfaz $l < -1/2$.

A representação que obtivemos é irredutível e de dimensão infinita (ilimitada acima e abaixo) [76]. Ela é etiquetada pelo número real contínuo b e cada um dos elementos da representação é etiquetado pelo número real a . Dado a_0 , um número infinito de autofunções de L_z são obtidas pela ação dos operadores escada, cujos a 's são $a_0, a_0 \pm 1, a_0 \pm 2$, etc.

3.4 O significado biológico dos operadores algébricos

O invariante da álgebra, o operador de Casimir, têm um significado biológico claro, ele indica o regime de decaimento ao equilíbrio do chaveamento gênico. Tal fato é demonstrado ao analisarmos a componente dois da solução dependente do tempo e o nível fundamental ($n = 0$) de sua meia vida:

$$\lambda_0^{(2)} = b \frac{\rho}{z_0}. \quad (3.37)$$

Note que esta taxa de decaimento para o equilíbrio é proporcional ao autovalor do operador de Casimir, b .

O autovalor do operador L_z é proporcional à afinidade entre o sítio operador e a proteína reguladora. Esta conclusão é sustentada pela análise das equações (3.28) e (3.29). O parâmetro b é definido pela ação do operador de Casimir. O parâmetro restante, a , é definido pela ação de L_z . Como será mostrado adiante, a probabilidade estacionária de o gene estar ativo apresenta dependência que cresce monotonicamente com a . Portanto, a afinidade entre o sítio operador e a proteína reguladora é inferida pelo cálculo de a .

Os operadores escada agem conectando processos estocásticos distintos, cujos valores de a diferem pela unidade. Tal fato decorre da definição da ação dos operadores levantamento e abaixamento sobre as funções geratrizes, conforme as equações (3.32) e (3.33). Mostraremos na próxima seção que a satisfaz um domínio de validade. Assim, a ação dos operadores escada dá origem à funções geratrizes apenas se a permanecer em seu domínio de validade. Caso contrário a função obtida não terá significado biológico embora permaneça válida do ponto-de-vista matemático.

3.5 Domínio biológico dos parâmetros algébricos

Do ponto-de-vista de Teoria das Representações, os etiquetadores de um elemento do espaço carregador b e a são reais. A etiqueta b é, conforme afirmado anteriormente, um real positivo, enquanto a pode assumir qualquer valor real. No entanto, para que este

elemento do espaço carregador seja função geratriz de um processo estocástico, a deverá satisfazer as restrições que mostraremos a seguir.

No gene auto-regulado, a pode ser um número real positivo ou negativo. Será positivo se z_0 for maior que χ e negativo em caso contrário. Nomearemos setor positivo o conjunto de valores assumidos por a 's maiores que zero:

$$0 \leq a \leq \frac{1 - \chi}{z_0 - \chi} b, \quad \text{se } z_0 > \chi; \quad (3.38)$$

e de setor negativo o conjunto de valores assumidos por a 's menores que zero:

$$\frac{1 - \chi}{z_0 - \chi} b \leq a \leq 0, \quad \text{se } z_0 < \chi. \quad (3.39)$$

Note que o intervalo de a é dado em função de b , o invariante da álgebra.

Ambas as equações acima são verificadas a partir da análise das definições de a e b no gene auto-regulado:

$$a = f \frac{1 - \chi}{z_0 - \chi}, \quad \text{e } b = f + M. \quad (3.40)$$

em que resolvemos a primeira das equações (3.40) para f , substituímos na segunda equação e, após isolarmos M , impomos que este seja positivo. Daí resultam as relações (3.38) e (3.39).

Note, na equação (3.40), que a apresenta um singularidade para z_0 e χ idênticos. No entanto, isto não implica em divergência da função geratriz das probabilidades do estado ativo. Isto deve-se ao argumento da função de KummerM também anular-se neste limite, conforme a definição (3.15). Assim, a função de KummerM tende à unidade e a função geratriz à zero.

No gene externamente regulado, analisamos os valores possíveis de a considerando $z_0 = 1$ no conjunto de valores possíveis no gene auto-regulado. Destarte, z_0 será necessariamente maior que χ e os valores de a no gene auto-regulado estão somente no setor

positivo. Além disso, note que a desigualdade (3.38), torna-se:

$$0 \leq a \leq b, \quad (3.41)$$

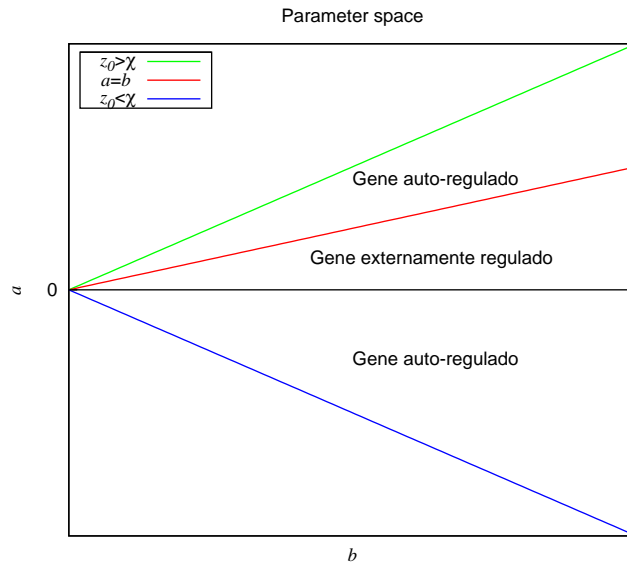


Figura 3.1: Domínios biológicos de a .

Visualizamos os domínios de validade de a na Figura (3.1). No eixo horizontal ficam os valores possíveis de b e no eixo vertical os valores de a . A linha horizontal preta no centro da figura é a origem dos valores de a . Para um dado valor de b , os valores possíveis de a estão numa linha vertical entre a origem e uma das linhas coloridas. O domínio de a no gene auto-regulado fica entre a linha verde e a origem, no setor positivo. No setor negativo, esse domínio está entre a origem e a linha azul. No gene externamente regulado, o domínio de a está entre a origem e a linha vermelha. Note que o domínio de validade biológica do gene externamente regulado é um sub-domínio de a no gene auto-regulado.

3.5.1 Nível de atividade do gene

O módulo do parâmetro a é uma medida indireta da probabilidade de o gene estar ativo, em outras palavras, de seu nível de atividade em regime estacionário. No gene externamente regulado isto é explícito, pois a probabilidade de o gene estar ativo apresenta dependência linear com a , $p_\alpha = a/b$. Portanto, se a assume seu valor mínimo, o gene possui

probabilidade um de estar reprimido. Por outro lado, se a assume seu valor máximo, o gene estará totalmente ativo.

No gene auto-regulado, esta relação não é tão clara do ponto-de-vista algébrico, visto que a função geratriz das probabilidades de o gene estar ativo, dada na equação (3.17), quando efetuada em η_1 fica:

$$p_\alpha = \frac{a z_0 - \chi}{b} \frac{1 - \chi}{1 - \chi} \frac{\mathcal{M}(1 + a, 1 + b, \eta_1)}{\mathcal{M}(a, b, \eta_1)}, \quad (3.42)$$

onde p_α é a probabilidade de o gene estar ativo em regime estacionário. Note que quando a é nulo, p_α vai à zero, enquanto que o valor máximo do módulo de a corresponde à total atividade do gene. Dada a dificuldade de invertermos a equação (3.42), na Figuras (3.2) construímos um gráfico da probabilidade de o gene estar ativo como função de a . Ambos os setores positivo e negativo do domínio de validade de a são contemplados. Em cada gráfico escolhemos um valor distinto de b , mantendo χ fixo em ambos. Há seis curvas em cada gráfico, cada uma representa p_α como função de a para um dado z_0 . O conjunto de valores de z_0 é o mesmo em ambos os gráficos. A probabilidade de o gene estar ativo cresce de zero à um monotonicamente com o módulo de a . No entanto, esta dependência não é linear.

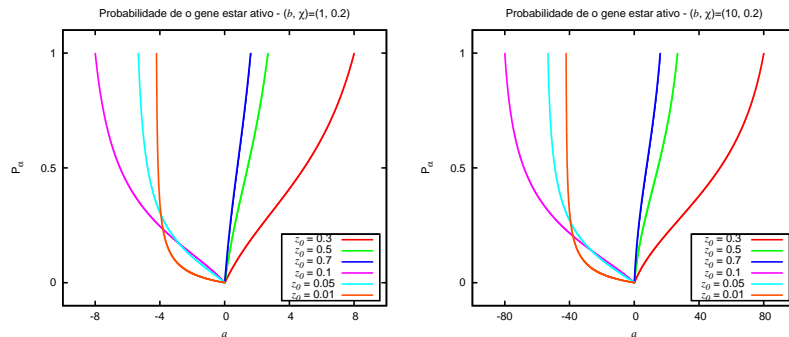


Figura 3.2: Nível de atividade do gene auto-regulado para dois valores distintos de b .

3.5.2 Classificação da estrutura das flutuações

Nesta subsecção, mostramos como a relação entre os etiquetadores da representação e seus elementos, b e a , é útil na classificação da estrutura das flutuações no número de proteínas expressas pelo gene. Definiremos as flutuações em termos do fator de Fano [82], o qual é dado por:

$$\mathcal{F} = \frac{\langle n^2 \rangle - \langle n \rangle^2}{\langle n \rangle}. \quad (3.43)$$

A distribuição de Poisson possui fator de Fano igual a unidade e é denominada Poisson ou Fano. Uma distribuição cujo fator de Fano é maior que um é denominada super Poisson e será mais larga. Uma distribuição sub Poisson possui fator de Fano menor que um e caracteriza-se por ser estreita.

Os valores médios de $\langle n \rangle$ e de $\langle n^2 \rangle$ são obtidos pela aplicação das derivadas primeira e segunda sobre as funções geratrizes das probabilidades totais, ϕ , em $\eta = \eta_1$:

$$\begin{aligned} \langle n \rangle &= \left. \frac{d\eta}{dz} \frac{d\phi}{d\eta} \right|_{\eta=\eta_1} = \frac{M [p_\alpha(1-\chi) + \chi]}{z_0(1-z_0)}, \\ \langle n^2 \rangle &= \langle n \rangle + \left(\frac{d\eta}{dz} \right)^2 \left. \frac{d^2\phi}{d\eta^2} \right|_{\eta=\eta_1}, \end{aligned} \quad (3.44)$$

$$\langle n^2 \rangle = \langle n \rangle^2 + \langle n \rangle + \left(\frac{M z_0 - \chi}{z_0(1-z_0)} \right)^2 \frac{a}{b} \left[\frac{a+1}{b+1} \frac{\mathcal{M}(a+2, b+2, \eta_1)}{\mathcal{M}(a, b, \eta_1)} - \frac{b}{a} \left(\frac{1-\chi}{z_0-\chi} \right)^2 p_\alpha^2 \right].$$

Nas figuras seguintes apresentaremos gráficos do fator de Fano de uma distribuição de probabilidades como função do número médio de proteínas expressas pelo gene auto-regulado. Na Figura 3.3 apresentamos um gráfico representando o setor positivo. O setor negativo é contemplado na Figura 3.4. Para os gráficos de cada setor, consideramos dois valores distintos de b . O valor de χ é idêntico em todas as figuras. Cada curva é obtida fixando um valor distinto de a . Em seguida, variamos z_0 no intervalo $\chi < z_0 < 1$ (ou $0 < z_0 < \chi$) para definirmos as curvas do setor positivo (ou setor negativo), tomamos os respectivos valores de \mathcal{F} e $\langle n \rangle$, e obtivemos o conjunto de pontos com que desenhemos as

curvas apresentadas.

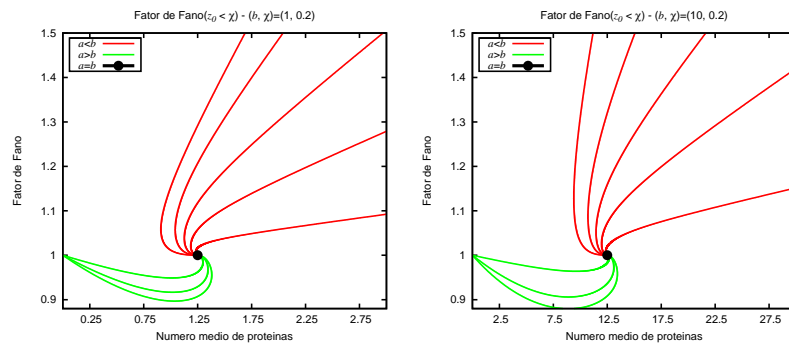


Figura 3.3: Fator de Fano no setor positivo

Na Figura 3.3, as linhas vermelhas correspondem à distribuições em que $a < b$. As linhas verdes representam distribuições em que $a > b$. O disco preto representa distribuições em que $a = b$, sendo $\langle n \rangle = b/(1 - \chi)$. O parâmetro b etiqueta a representação do grupo $SO(2, 1)$ e a denota um elemento desta representação. Quando $a < b$ temos um processo estocástico super Poisson e teremos um processo de Poisson em caso de a e b serem idênticos. Processos sub Poisson ocorrem quando $a > b$.

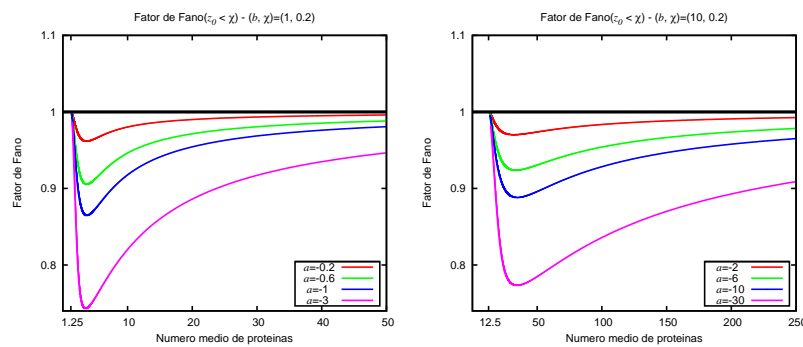


Figura 3.4: Fator de Fano no setor negativo

Na Figura 3.4 cada curva colorida corresponde a um valor distinto de a . Há uma linha de referência horizontal preta. É conclusão direta que qualquer processo estocástico definido no setor negativo será sub Poisson.

Um gene externamente regulado apresenta processos estocásticos de síntese proteínica que são super Poisson ou Poisson. Esta afirmação é demonstrada partindo-se do fato que os processos estocásticos do gene externamente regulado são definidos exclusivamente no setor positivo. Adicionalmente, esses processos são etiquetados por números que satis-

fazem a condição $a \leq b$. De acordo com os resultados apresentados na Figura 3.3, isto corresponde à processos estocásticos de Fano e super Fano. Isto é verificado algebricamente pela equação para o fator de Fano do gene externamente regulado, explicitamente:

$$\mathcal{F} = 1 + \frac{N(1 - \chi)^2 a}{1 + b} \frac{b - a}{b a(1 - \chi) + \chi b}, \quad (3.45)$$

obtida a partir das equações (3.44), substituindo-se, nesta ordem $M = N(1 - z_0)$, $z_0 = 1$ e $p_\alpha = a/b$.

O gene auto-regulado realiza a síntese proteínica estocástica nos três regimes distintos de flutuação: Poisson e sub ou super Poisson. Se o processo estocástico de síntese de proteínas for definido no setor positivo, os etiquetadores dos elementos da representação, a , poderão ser maiores, menores ou iguais à b . Conseqüentemente, o gene poderá expressar proteínas estocasticamente nas três classes de flutuações definidas em termos do fator de Fano. Em caso de o processo estocástico ser definido no setor negativo, o gene auto-regulado necessariamente expressará em regime sub-Poisson.

Em suma, mostramos que as soluções do regime estacionário dos modelos estocásticos apresentados possuem simetria de Lie tipo Lorentz $SO(2,1)$. O invariante da álgebra caracteriza a taxa de decaimento para o equilíbrio do chaveamento gênico. O autovalor do operador de Cartan de um elemento da representação da álgebra mede, indiretamente, o nível de atividade do gene. A ação dos operadores escada conecta processos estocásticos cujos a 's diferem pela unidade. O regime de flutuações no número de proteínas expressas pelo gene é caracterizado pela relação entre os etiquetadores de um elemento do espaço carregador. O modelo estocástico para um gene externamente regulado é obtido a partir de uma restrição no espaço de validade do modelo para um gene auto-regulado.

Em perspectiva, podemos utilizar o ferramental matemático das simetrias de Lie na busca por uma regra de composição entre modelos para genes isolados, *e. g.* calculando os coeficientes de Clebsch-Gordon [76]. Podemos investigar o significado biológico da álgebra de dimensão infinita que forma a base em que as equações do modelo, Apêndice B. O parâmetro z_0 é, provavelmente, autovalor de um operador de Casimir de uma álgebra que

contenha a álgebra $so(2, 1)$ já descoberta.

Capítulo 4

Dinâmica estocástica

Neste Capítulo exploramos um exemplo de dinâmica obtida a partir das soluções exatas do modelo estocástico mostrado no Capítulo 1. A série em termos da qual a solução do modelo para o gene auto-regulado é escrita não foi somada. Sendo assim, a solução dinâmica do modelo para o gene auto-regulado é decomposta em suas taxas de decaimento para o equilíbrio. Esta decomposição é realizada numericamente, pois ainda inexistem produtos internos entre funções de Heun confluentes. Mostramos um exemplo de evolução das probabilidades na seção final.

4.1 Solução geral em termos das funções de Heun

As funções geratrizes das probabilidades dependentes do tempo, conforme apresentado no primeiro capítulo desta tese, são dadas por:

$$\alpha(z, t) = \sum_{j=0}^{\infty} \sum_{i=1}^2 (-1)^{(i-1)j} C_j^{(i)} A_j^{(i)}(z) e^{-\lambda_j^{(i)} t}, \quad (4.1)$$

$$\beta(z, t) = \sum_{j=0}^{\infty} \sum_{i=1}^2 (-1)^{(i-1)j} D_j^{(i)} B_j^{(i)}(z) e^{-\lambda_j^{(i)} t}, \quad (4.2)$$

onde $C_j^{(i)}$ e $D_j^{(i)}$ são coeficientes da expansão na base de autofunções, $A_j^{(i)}(z)$ e $B_j^{(i)}(z)$ são as autofunções e $\lambda_j^{(i)}$ são seus autovalores correspondentes, dados por:

$$\lambda_j^{(1)} = j\rho, \quad \text{e} \quad \lambda_j^{(2)} = \frac{b+j}{z_0}\rho, \quad \text{em que } j = 0, 1, 2, \dots \quad (4.3)$$

Neste capítulo, os autovalores do modelo são as taxas de decaimento para o regime estacionário. O espectro correspondente é apresentado na Figura 4.1. No eixo vertical escrevemos a taxas de decaimento em unidades de ρ . Na direção horizontal, da esquerda para a direita, escrevemos os conjuntos de autovalores em termos de $\lambda_j^{(1)}$ e três situações distintas para $\lambda_j^{(2)}$, quando a fração $\lambda_0^{(2)}/\lambda_1^{(1)}$ é, respectivamente, menor, igual ou maior que um. O espaçamento entre dois autovalores consecutivos são 1 ou $1/z_0$ para, respectivamente, $\lambda_j^{(1)}$ ou $\lambda_j^{(2)}$. O nível fundamental do conjunto 1 é nulo e corresponde ao regime estacionário. O nível fundamental de decaimento do conjunto 2 é dado por b/z_0 e é determinado pela ação do operador de Casimir, conforme apresentado no Capítulo 2.

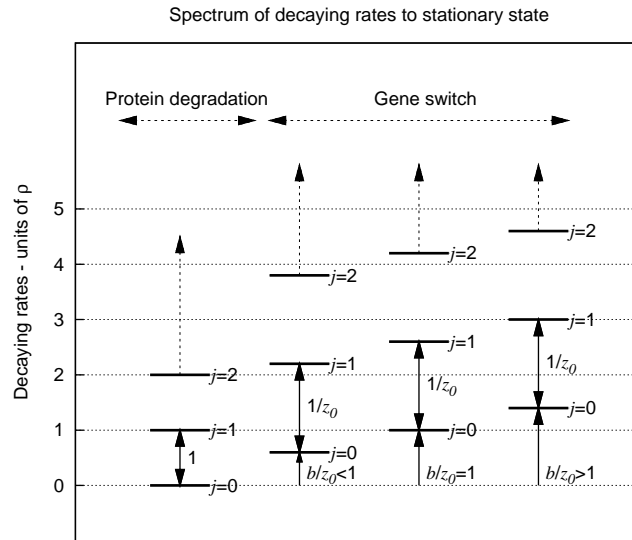


Figura 4.1: Espectro de taxas de decaimento para o regime estacionário.

Os coeficientes $C_0^{(1)}$ e $D_0^{(1)}$ são as constantes de normalização do regime estacionário. Este regime é completamente determinado a partir dos valores dos parâmetros do modelo, conforme aprendemos no Capítulo 2. Sendo assim, ao calcularmos as probabilidades estacionárias, impomos que as constantes multiplicativas sejam idênticas, $C_0^{(1)} = D_0^{(1)}$, e

que a probabilidade total valha um. Daí, resolvemos a equação resultante para a constante de normalização. Isto resulta na constante de normalização do regime estacionário, obtida independentemente das componentes dinâmicas das soluções. Mostraremos sua fórmula explícita mais adiante.

A dinâmica probabilística será obtida ao determinarmos uma condição inicial. O estado inicial poderá ser dado por uma função não normalizada do número de proteínas. Esta função será escrita como combinação linear das autofunções apresentadas em (4.1) e (4.2), onde as constantes $C_j^{(i)}$ e $D_j^{(i)}$ são os seus coeficientes.

A dinâmica de um processo estocástico, obtida a partir deste formalismo, é fortemente correlacionada com seu estado inicial. Este estado será uma combinação linear das autofunções, a partir da qual fixamos os i, j -ésimos coeficientes da expansão e, conseqüentemente, as correspondentes i, j -ésimas taxas de decaimento para o equilíbrio. Eventualmente, a série obtida pode ser somada em funções dependentes das variáveis z e t concomitantemente. Todavia, mesmo que tal soma não seja viável, haverá vários decaimentos exponenciais. A menor dessas taxas determina a meia-vida do regime dinâmico do gene. Em caso de os parâmetros satisfazerem $0 < b < z_0 < 1$, a constante de chaveamento gênico domina a dinâmica do sistema. Para z_0 e b idênticos, a degradação e o chaveamento gênico atingem equilíbrio simultaneamente. A taxa de degradação proteínica define a meia-vida da dinâmica quando b/z_0 é maior que um.

4.1.1 Probabilidades

A dinâmica das probabilidades é obtida derivando-se as funções geratrizes mostradas nas equações (4.1) e (4.2). Sendo assim, a probabilidade de o sítio operador estar ligado ou não ligado à proteína reguladora é, respectivamente, obtida a partir de:

$$\alpha_n = \frac{1}{n!} \left. \frac{\partial^n \alpha}{\partial z^n} \right|_{z=0}, \quad \text{ou} \quad \beta_n = \frac{1}{n!} \left. \frac{\partial^n \beta}{\partial z^n} \right|_{z=0}, \quad (4.4)$$

quando n proteínas são encontradas no citoplasma.

A expressão resultante pode ser escrita como:

$$\alpha_n = \frac{1}{N} \sum_{j=0}^{\infty} \sum_{i=1}^2 \frac{1}{j!} \left[\frac{(-1)^{(i-1)n}}{n!} C_j^{(i)} A_j^{(i),n} - \bar{Y}_j^{(i)} p_{\alpha}^e \right] \exp(-\lambda_j^{(i)} t), \quad (4.5)$$

e

$$\beta_n = \frac{1}{N} \sum_{j=0}^{\infty} \sum_{i=1}^2 \frac{1}{j!} \left[\frac{(-1)^{(i-1)n}}{n!} D_j^{(i)} B_j^{(i),n} - \bar{Y}_j^{(i)} p_{\beta}^e \right] \exp(-\lambda_j^{(i)} t), \quad (4.6)$$

em que $X_j^{(i),n}$ representa a n -ésima derivada de $X_j^{(i)}$ em relação à z , efetuada em z nulo,

$$X_j^{(i),n} \equiv \left. \frac{d^n X_j^{(i)}}{dz^n} \right|_{z=0}.$$

A constante de normalização N é calculada a partir de:

$$N = \sum_{n=0}^{\infty} \sum_{i=1}^2 \sum_{j=j_0}^{\infty} (-1)^{(i-1)n} \frac{C_j^{(i)} A_j^{(i),n} + D_j^{(i)} B_j^{(i),n}}{n! j!}, \quad (4.7)$$

onde $j_0 = 1$ se $i = 1$ e $j_0 = 0$ se $i = 2$. Note que o regime estacionário não é computado no cálculo da constante de normalização. As constantes da expansão, $C_j^{(i)}$ e $D_j^{(i)}$, são definidas a partir das condições iniciais. Em particular, a constante de normalização do regime estacionário é obtida a partir de:

$$\tilde{C} = \left(\sum_{n=0}^{\infty} \frac{A_0^{(1),n} + B_0^{(1),n}}{n!} \right)^{-1}, \quad \text{portanto } C_0^{(1)} = D_0^{(1)} = N\tilde{C}. \quad (4.8)$$

As constantes com barra, $\bar{Y}_j^{(i)}$ são definidas tal que em t nulo, a componente estacionária se anule e durante a evolução temporal a probabilidade total seja um . Elas são a fração que as componentes i, j -ésimas da expansão em série representam em relação à soma N :

$$\bar{Y}_j^{(i)} = \sum_{n=0}^{\infty} (-1)^{(i-1)n} \frac{C_j^{(i)} A_j^{(i),n} + D_j^{(i)} B_j^{(i),n}}{n!}, \quad (4.9)$$

sendo $j = 1, 2, 3 \dots$ se $i = 1$ e $j = 0, 1, 2 \dots$ se $i = 2$. As probabilidades totais de o gene estar, respectivamente, ativo ou reprimido em regime estacionário são p_{α}^e ou p_{β}^e .

Demonstramos a constância da probabilidade total ao efetuarmos a soma em n nas equações (4.5) e (4.6), o que resulta em:

$$p_\alpha(t) = p_\alpha^e + \frac{1}{N} \sum_{i=1}^2 \sum_{j=j_0}^{\infty} \frac{1}{j!} \left[C_j^{(i)} p_\alpha^{(i)}(j) - \bar{Y}_j^{(i)} p_\alpha^e \right] e^{-\lambda_j^{(i)} t}, \quad (4.10)$$

e

$$p_\beta(t) = p_\beta^e + \frac{1}{N} \sum_{i=1}^2 \sum_{j=j_0}^{\infty} \frac{1}{j!} \left[D_j^{(i)} p_\beta^{(i)}(j) - \bar{Y}_j^{(i)} p_\beta^e \right] e^{-\lambda_j^{(i)} t}, \quad (4.11)$$

onde $p_x^{(i)}(j) = \sum_{n=0}^{\infty} (-1)^{(i-1)n} X^{(i),n}/n!$. Somamos ambas as expressões acima e, considerando $p_\alpha^e + p_\beta^e = 1$, obtemos:

$$p_\alpha(t) + p_\beta(t) = 1 + \frac{1}{N} \sum_{i=1}^2 \sum_{j=j_0}^{\infty} \frac{1}{j!} \left[C_j^{(i)} p_\alpha^{(i)}(j) + D_j^{(i)} p_\beta^{(i)}(j) - \bar{Y}_j^{(i)} \right] e^{-\lambda_j^{(i)} t}. \quad (4.12)$$

O termo entre colchetes anula-se ao considerarmos a equação (4.9) e portanto:

$$p_\alpha(t) + p_\beta(t) = 1. \quad (4.13)$$

4.1.2 Derivadas das autofunções

As derivadas n -ésimas das autofunções $A_j^{(i)}$ e $B_j^{(i)}$ são calculadas numericamente. Isto porque estas autofunções são escritas em termos das funções de Heun confluentes. Diferente das funções de Kummer, as funções de Heun não possuem uma fórmula fechada para o cálculo de sua derivada n -ésima. Como alternativa, poderíamos utilizar a rotina contida no programa Maple 11. No entanto, esta é lenta no cálculo de derivadas de ordem superior à dez e, leva cerca de dez minutos para o cálculo das mesmas.

Visando tornar estes cálculos mais rápidos, podemos escrever uma relação de recorrência entre as derivadas das funções de Heun. Esta relação é obtida pela ação do operador derivada n -ésima em z sobre a equação diferencial de segunda ordem para $A_j^{(i)}$. Dessarte, escrevemos a derivada segunda de $A_j^{(i)}$ como função da derivada primeira e da própria função. E assim sucessivamente nas derivadas de ordem superior. Também aplicamos

este operador sobre a equação para $B_j^{(i)}$, que é função de $A_j^{(i)}$ e de $A_j^{(i),1}$. Obtivemos uma fórmula para a derivada n -ésima de $B_j^{(i)}$ em relação à z como função das derivadas de $A_j^{(i)}$.

O algoritmo que escrevemos resultou bastante rápido para o cálculo de derivadas de ordem superior, *e. g.* de ordem 150, levando menos de um segundo para efetuar esta e todas as derivadas anteriores. Os cálculos e a relação de recorrência a que nos referimos são apresentados no Apêndice C.

4.2 Dinâmica estocástica

Nesta seção apresentamos um exemplo de evolução temporal. Mostramos quatro instantes distintos da dinâmica das distribuições de probabilidade. Nosso objetivo é mostrar que a metodologia empregada para a obtenção das soluções dinâmicas do modelo funciona. Para tal, apenas utilizamos dos primeiros termos da expansão na base de autofunções.

Neste exemplo de evolução temporal fixamos os momentos de ordem inferior. A partir daí, calculamos valores para os coeficientes da expansão que resultem em distribuições de probabilidade correspondentes aos momentos que escolhemos. No gráfico correspondente à cada instante de tempo apresentamos as probabilidades de encontrarmos n proteínas na célula, ϕ_n , e as respectivas probabilidades de o gene estar ativo, α_n , ou reprimido, β_n . As probabilidades totais são indicadas em preto nas figuras. As probabilidades de o gene estar ativo ou reprimido são, respectivamente, apresentadas em verde ou azul.

Na Figura 4.2 apresentamos a dinâmica de uma distribuição cuja probabilidade total de estar ativo, número médio de proteínas produzidas e fator de Fano da distribuição inicial são, respectivamente, dados por:

$$p_\alpha^o = 0.9, \quad \langle n \rangle^o = 72., \quad \mathcal{F}^o = \frac{\langle n^2 \rangle^o - (\langle n \rangle^o)^2}{\langle n \rangle^o} = 0.68.$$

Na expansão na base de autofunções consideramos termos de até segunda ordem, ou seja, $j = 2$. Os coeficientes foram escolhidos para satisfazer as condições iniciais acima. Em

regime estacionário a probabilidade total de o gene estar ativo, número médio de proteínas produzidas e fator de Fano da distribuição são, respectivamente, dados por:

$$p_{\alpha}^e = 0.12, \quad \langle n \rangle^e = 13.5, \quad \mathcal{F}^e = \frac{\langle n^2 \rangle^e - (\langle n \rangle^e)^2}{\langle n \rangle^e} = 0.84.$$

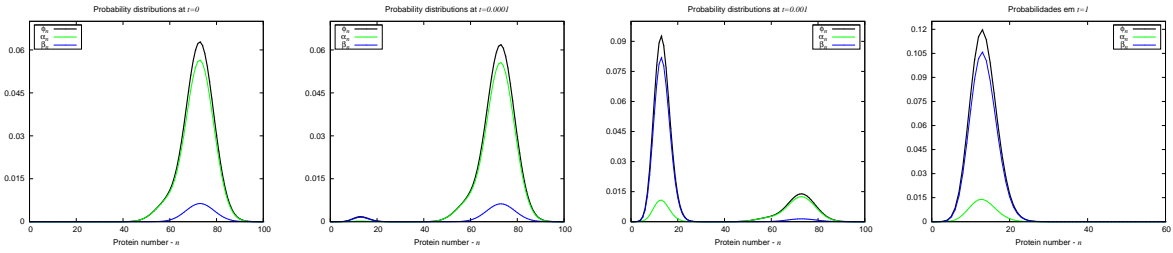


Figura 4.2: Dinâmica das distribuições de probabilidade. Os parâmetros do modelo assumem os respectivos valores: $(b, a, z_0, \chi) = (10, 20, 0.08, 0)$.

Na Figura 4.2 há quatro gráficos. Cada um deles representa as probabilidades em instantes distintos. A distribuição inicial é centrada em torno de $\langle n \rangle^o$, enquanto a distribuição estacionária é centrada em $\langle n \rangle^e$. Quando o tempo atinge valor igual à vida média de uma proteína, ou seja, $t = 1$, o sistema alcança o regime estacionário. A razão para isso é que os coeficientes do tipo 2 na expansão do estado inicial são muito maiores que as componentes do tipo 1. Portanto, a influência efetiva da degradação proteínica é pequena. Outrossim, a distribuição inicial não desloca-se suavemente para a esquerda. Ao contrário, a altura do pico inicial diminui gradualmente, enquanto o pacote estacionário cresce. Isto corresponde à probabilidade alta de degradação de muitas proteínas em frações de tempo de $1/\rho$.

Nosso objetivo com a dinâmica aqui apresentada é apenas ilustrativo de uma aplicação das soluções que obtivemos. A razão para isto é que consideramos uma expansão em série com termos de até segunda ordem. Para que dinâmicas mais precisas e realistas sejam obtidas, é necessário que as expansões tomem em conta termos de até, pelo menos, $j = 100$. Para tal, é necessário resolver numericamente um sistema de equações para os

coeficientes $C_j^{(i)}$, dado por:

$$\alpha_n(0) = \frac{1}{N} \sum_{j=0}^{\infty} \sum_{i=1}^2 \frac{1}{j!} \frac{(-1)^{(i-1)n}}{n!} C_j^{(i)} A_j^{(i),n}, \quad (4.14)$$

em que $\alpha_n(0)$ são probabilidades iniciais. A alternativa analítica para o procedimento acima é encontrar o produto interno entre funções de Heun confluentes. Neste caso, as probabilidades iniciais seriam escritas em termos de funções geratrizes convertidas em funções de Heun confluentes.

Em suma, neste Capítulo apresentamos um exemplo de dinâmica estocástica de um gene auto-regulado. Para obtê-la, mostramos uma forma geral para escrevermos as funções geratrizes como combinação linear na base de autofunções, tal que a probabilidade total seja preservada. Desenvolvemos uma relação de recorrência entre as derivadas das funções de Heun confluentes que permitem o cálculo de derivadas superiores com rapidez. Esta metodologia é aplicável à qualquer equação diferencial.

Em perspectiva, as funções de Heun confluentes ainda carecem de uma fórmula fechada para o produto interno. Este é um problema de interesse imediato tanto para o estudo da dinâmica do gene auto-regulado quanto para a comunidade que investiga as propriedades das funções de Heun. Alternativamente, resta resolver numericamente a expansão de um estado inicial em termos das autofunções até componentes de ordens superiores.

Capítulo 5

Os genes auto e externamente regulados

Nesta tese apresentamos dois modelos estocásticos para a síntese proteínica por um gene auto ou externamente regulado. Ambos são de grande interesse à investigação do funcionamento de uma rede gênica. Entre outros motivos, porque o chamado módulo de entrada simples, composto por um gene auto regulado e genes externamente regulados, é bastante recorrente nessas redes [5, 83, 6]. Neste caso, a proteína expressa por um gene auto-regulado também controla o nível de expressão dos genes externamente regulados. Investigar as razões que guiam a “escolha” da auto-regulação ou regulação externa da expressão gênica é tema de pesquisa em aberto. Um paradigma nesta busca é a teoria do controle, na qual atribui-se à auto-regulação a performance ótima de um sistema. Em biologia, argumenta-se que esta estratégia foi selecionada pela evolução e algumas das razões para isso são apresentadas na literatura [8, 10, 84, 85, 86]. Sistemas auto-regulados são mais eficazes no controle da expressão gênica, sendo mais precisos na manutenção de um determinado número médio de proteínas no interior celular [8, 84, 74]. Eles também atingem tal quantidade com maior rapidez [85]. O nível de atividade do gene auto-regulado responde mais rápida e proporcionalmente à perturbações causadas por agentes externos [86]. As flutuações no número de proteínas expressas por este gene tendem a ser menores [8, 10].

Os modelos teóricos comparando estas duas estratégias de regulação lidam com equações determinísticas para a concentração proteínica [85, 86]. Análises do papel das flu-

tuações na população de proteínas são introduzidas via técnica de Langevin. Ou seja, partem-se de equações determinísticas adidas de um termo aleatório, mais especificamente, ruído branco [10]. Neste capítulo, utilizando-nos dos modelos estocásticos apresentados nesta tese e de suas propriedades, comparamos a auto-regulação e a regulação externa de um gene e recuperamos resultados presentes na literatura. Mostramos que o gene auto-regulado é mais preciso no controle do número de proteínas no interior celular, além de ser mais rápido na resposta à perturbações externas.

5.1 Regime estacionário

Nesta seção analisamos as flutuações exibidos por um gene auto ou externamente regulado, em regime estacionário. Também mostramos distribuições de probabilidades exibidas por ambos em regime estacionário e as comparamos.

5.1.1 Flutuações

Na subseção 3.5.2, à página 42, demonstramos que o regime de flutuações do gene é classificado em termos da relação entre os parâmetros algébricos a e b . Mostramos que num gene auto-regulado o domínio biológico é composto por dois setores, o positivo ou o negativo. Estes, respectivamente, ocorrem quando a é um número real maior ou menor que zero. O setor positivo é devido à z_0 ser maior que χ , sendo que a está no intervalo:

$$0 \leq a \leq \frac{1 - \chi}{z_0 - \chi} b. \quad (5.1)$$

O setor negativo deve-se à z_0 ser menor que χ , sendo que a satisfaz:

$$0 \geq a \geq \frac{1 - \chi}{z_0 - \chi} b. \quad (5.2)$$

Um gene auto-regulado operando no setor positivo terá o número de proteínas satisfazendo uma distribuição de probabilidades que será: super-Poisson quando a for menor

que b ; Poisson quando a e b forem idênticos; em caso de a ser maior que b , sub-Poisson. Por sua vez, se este gene operar no setor negativo, assumirá somente o regime sub-Poisson de flutuações, independente da relação entre a e b .

Para analisarmos o gene externamente regulado, basta considerarmos z_0 igual à unidade nas equações para o gene auto-regulado. Há duas consequências diretas, sendo a eliminação do setor negativo a primeira delas. A segunda consequência decorre da equação (5.1), em que verificamos que o intervalo de validade de a reduziu-se à $0 \leq a \leq b$. Portanto, os regimes de flutuação possíveis ao gene externamente regulado serão: Poisson quando a e b forem idênticos; super-Poisson quando a for menor que b .

Destarte, concluímos que um gene auto-regulado pode apresentar um controle da população proteínica mais estrito que o gene externamente regulado. Isto ocorrerá sempre que a for um real negativo, ou um real positivo maior que b . Este resultado recupera dados já presentes na literatura [8, 10]. Outrossim, mostra limites em que tanto um gene sob regulação externa quanto um gene sob auto-regulação podem ser equivalentes do ponto-de-vista do regime de flutuações no número de proteínas sintetizadas.

5.1.2 Distribuições de probabilidades

Apresentamos as distribuições de probabilidades do número de proteínas em regime estacionário para ambos os sistemas de regulação da síntese proteínica. No caso de um gene auto-regulado, as probabilidades são obtidas com o auxílio das expressões seguintes:

$$\phi_n = \left(\frac{M}{z_0}\right)^n \frac{\exp\left(\frac{M}{z_0} \frac{\chi}{z_0 - 1}\right)}{c n! (1 - z_0)^n} \sum_{k=0}^n \binom{n}{k} \chi^{n-k} (z_0 - \chi)^k \frac{(a)_k}{(b)_k} \mathcal{M}\left(a + k, b + k, M \frac{\chi - z_0}{1 - z_0}\right), \quad (5.3)$$

para a probabilidade total,

$$\alpha_n = \frac{a}{b} \frac{z_0 - \chi}{1 - \chi} \left(\frac{M}{z_0}\right)^n \frac{\exp\left(\frac{M}{z_0} \frac{\chi}{z_0 - 1}\right)}{c n! (1 - z_0)^n} \sum_{k=0}^n \binom{n}{k} \chi^{n-k} (z_0 - \chi)^k \frac{(a + 1)_k}{(b + 1)_k} \mathcal{M}\left(a + 1 + k, b + 1 + k, M \frac{\chi - z_0}{1 - z_0}\right), \quad (5.4)$$

para a probabilidade de o gene estar ativo e, por fim:

$$\beta_n = \phi_n - \alpha_n, \quad (5.5)$$

para a probabilidade de o gene estar reprimido. A constante c é definida como:

$$c = \mathcal{M}(a, b, M(1 - \chi/z_0)). \quad (5.6)$$

As expressões para cálculo das probabilidades foram obtidas com a ação do operador derivada k -ésima em relação à z sobre as funções geratrizes das probabilidades do regime estacionário, apresentadas na Seção 3.2, à página 32. Nas equações de (5.3) à (5.6), consideramos:

$$M = b - \frac{z_0 - \chi}{1 - z_0} a.$$

As fórmulas para as probabilidades do gene sob regulação externa podem ser obtidas diretamente a partir das equações de (5.3) à (5.6), se impusermos, na ordem em que são apresentadas, as seguintes condições:

$$M = N(1 - z_0) \text{ e } z_0 = 1.$$

Sendo assim, não escreveremos aqui as equações para obtenção das probabilidades do gene externamente regulado. O leitor interessado em escrever as fórmulas para obtenção dessas probabilidades pode utilizar-se das relações acima apresentadas.

As distribuições de probabilidade foram construídas visando ambos os genes expressando o mesmo número médio de proteínas. Também impusemos que b fosse o mesmo, ou seja, do ponto-de-vista das simetrias de Lie, os genes estão na mesma representação. Nos gráficos, representamos as probabilidades no eixo vertical e o número de proteínas no eixo horizontal. Em cada gráfico apresentamos as probabilidades correspondentes à um determinado conjunto de parâmetros. A probabilidade total de encontrarmos n proteínas na célula é representada em preto. As probabilidades correspondentes de o gene estar

ativo ou reprimido são, respectivamente, indicadas em verde ou azul.

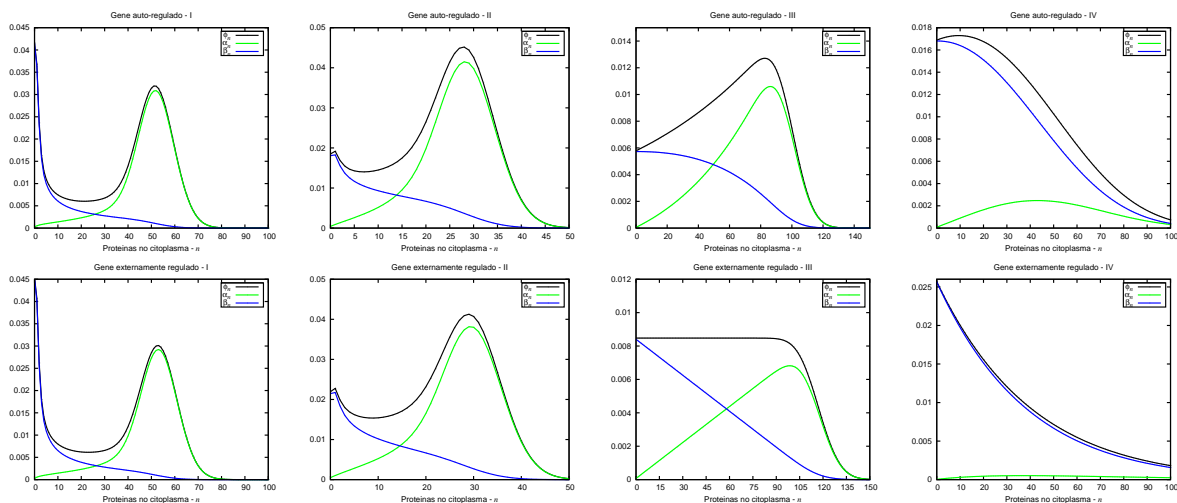


Figura 5.1: Distribuições de probabilidade no regime estacionário para os dois modelos. Os gráficos superiores referem-se ao gene auto-regulado e, ao externamente regulado, referem-se os gráficos inferiores. Da esquerda para a direita escolhemos b e a com os respectivos valores: I - 0.5 e 0.34; II - 1 e 0.7; III - 2 e 1; IV - 20 e 1. Os números médios de proteínas correspondentes à essas distribuições são: I - 37.5; II - 22.5; III - 20.7; IV - 36.1. Os valores de χ são: I - 0.01; II - 0.01; III - 0; IV - 0. Nas distribuições do gene auto-interagentes consideramos os seguintes valores para z_0 : I - 0.997; II - 0.99; III - 0.99; IV - 0.9. No caso externamente regulado, os valores de N são: I - 55; II - 32; III - 118; IV - 722.

No conjunto de gráficos exposto na Figura 5.1, escolhemos os parâmetros algébricos a idênticos para ambos os modelos de gene, auto ou externamente regulado. Escolhemos os parâmetros z_0 , no gene auto-regulado, e N , no gene externamente regulado, para que ambos os genes expressem a mesma quantidade média de proteínas. Os quatro gráficos superiores mostram os resultados obtidos para um gene auto-regulado, enquanto os quatro gráficos inferiores mostram resultados para um gene externamente regulado. Um gráfico da linha superior terá como seu correspondente, em termos de b , a e $\langle n \rangle$, o gráfico imediatamente abaixo. Há diferenças quantitativas entre as distribuições de probabilidades em relação à estratégia de regulação da síntese proteínica. Note que esta diferença se torna maior à medida em que z_0 se distancia da unidade, da esquerda para a direita. No entanto, esta diferenciação não é qualitativamente significativa e o comportamento estocástico dos genes é similar. Outrossim, as probabilidades dos genes estarem ativos ou reprimidos são parecidas em ambos os modelos e aproximadamente 1/2 em todos os casos.

No conjunto de Figuras 5.2 apresentamos distribuições de probabilidade em que os

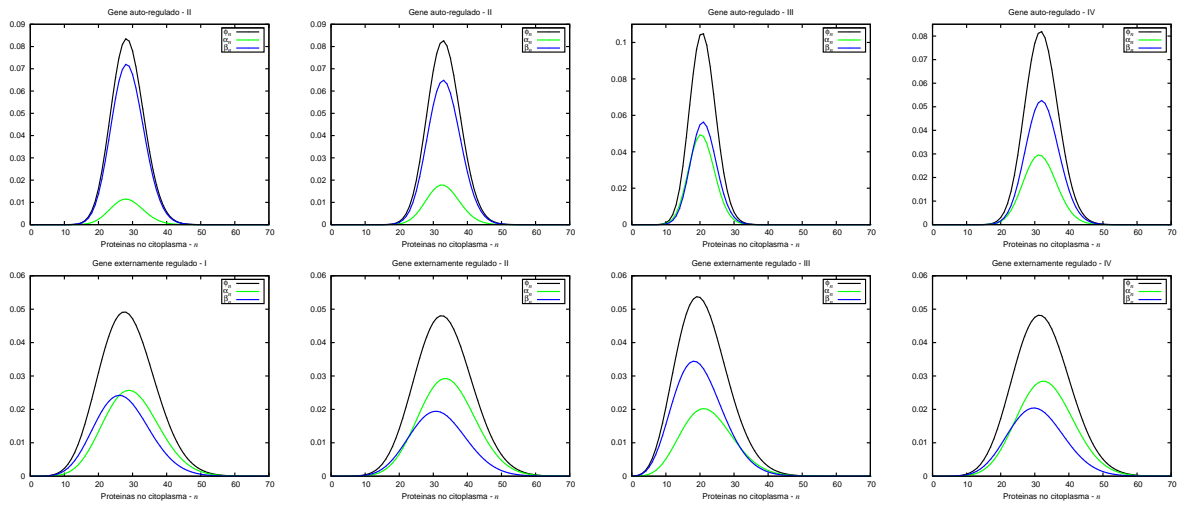


Figura 5.2: Distribuições de probabilidade no regime estacionário para os dois modelos. Os gráficos superiores referem-se ao gene auto-regulado e, ao externamente regulado, referem-se os gráficos inferiores. Em todos os gráficos b é 20. Da esquerda para a direita, no gene auto-regulado a assume os respectivos valores: I - 50; II - 100; III - 200; IV - -90; no gene externamente regulado: I - 10.4; II - 12.06; III - 7.6; IV - 11.7. Os números médios de proteínas correspondentes à essas distribuições são: I - 28.5; II - 33.3; III - 20.9; IV - 32.2. Os valores de χ são todos nulos com exceção das distribuição superior número IV, em que este é dado por: 0.2. Em todas as distribuições inferiores inferiores escolhemos $N = 55$. O valor de z_0 é 0.083 em todas as distribuições superiores com exceção da número IV, em que $z_0 = 0.05$.

valores de a são distintos. As distribuições na linha superior referem-se ao gene auto-regulado e aquelas da inferior ao gene externamente regulado. Uma distribuição inferior possui mesmo número médio que aquela imediatamente acima. As distribuições do gene auto-regulado são todas sub-Fano enquanto as do gene externamente regulado são super-Fano. Há também uma clara distinção na estrutura dos chaveamentos gênicos, dado que os níveis de atividade dos genes diferem. Isto é visualizado nas probabilidades de o gene estar reprimido, que é sempre maior que a probabilidade de o gene estar ativo no gene auto-regulado. No gene externamente regulado, a relação entre essas probabilidades varia dentre os gráficos.

5.2 Dinâmica

Aqui apresentamos a dinâmica da probabilidade de um gene auto ou externamente regulado estar em estado ativo. Também mostramos a evolução temporal do número

médio de proteínas correspondente. Os estados iniciais e regimes estacionários das probabilidades, e números médios, são os mesmos em ambos os genes. A unidade temporal é definida em termos de $1/\rho$ e será representada no eixo horizontal. Adotaremos o mesmo valor de ρ nos dois modelos. Isto implica assumirmos que a proteína expressada por ambos os genes é a mesma. O valor de b também é idêntico em ambos os modelos. Os parâmetros do gene auto-regulado são os mesmos que os utilizados no Capítulo 3, ou seja $b = 10, a = 20, \chi = 0, z_0 = 0,08$. No gene externamente regulado, escolhemos $a = 1.2, N = 114, \chi = 0$. Os valores iniciais e estacionários de p_α e $\langle n \rangle$ são idênticos aos escolhidos no Capítulo 3:

$$p_\alpha^o = 0.9, \quad \langle n \rangle^o = 72.,$$

$$p_\alpha^e = 0.12, \quad \langle n \rangle^e = 13.5.$$

Nos gráficos, a curva em rosa representa o gene auto-regulado (*gar*). O gene externamente regulado (*ger*) é representado pela curva verde.

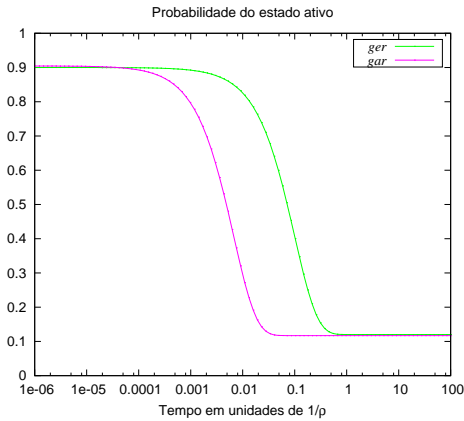


Figura 5.3: Evolução temporal das probabilidades de o genes auto ou externamente regulados estarem ativos.

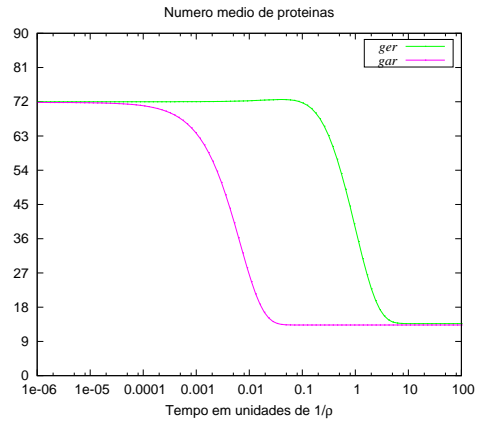


Figura 5.4: Evolução temporal do número médio de proteínas expressas pelos genes auto ou externamente regulados.

Na Figura 5.3, mostramos a dinâmica da probabilidade de o gene estar ativo. Estas probabilidades são representadas no eixo vertical. Verificamos, a partir do gráfico, que o gene auto-regulado atinge o nível de atividade estacionário antes do gene externamente

regulado.

Na Figura 5.4, mostramos a dinâmica do número médio de proteínas no interior celular, o qual é representado no eixo vertical. A partir do gráfico, verificamos que o número médio de proteínas expressas pelo gene auto-regulado atinge o regime estacionário com antecedência.

Em ambas as dinâmicas apresentadas, o gene auto-regulado atinge o regime estacionário antes do gene externamente regulado. No caso das taxas de decaimento do chaveamento gênico, isto deve-se à b ser dividido por z_0 , que no gene auto-regulado está no intervalo entre zero e um. Portanto, essa taxa de decaimento será maior que a do gene externamente regulado. Outra causa disso são os estados iniciais que foram escolhidos, os quais levam os termos de degradação proteínica a terem influência desprezível sobre a dinâmica do gene auto-regulado.

Em suma, neste Capítulo analisamos comparativamente os modelos estocásticos para expressão gênica estudados ao longo desta Tese. Mostramos que o gene auto-regulado pode expressar proteínas em regimes de baixas flutuações. Portanto esta estratégia de regulação é mais eficaz na manutenção de um número médio de proteínas no interior celular. Este resultado é consequência direta da existência de simetrias no modelo, desde que o regime de flutuações é definido em termos da relação entre b e a . Outrossim, mostramos um exemplo de dinâmica em que o gene auto-regulado atinge o regime estacionário antes do gene externamente regulado. Este resultado está em consonância com dados já presentes na literatura.

Em perspectiva, o modelo estocástico para o gene externamente regulado, sob a ótica da teoria dos grupos, resulta de um caso particular do modelo para o gene auto-regulado. Isto leva à hipótese de que a estratégia de regulação adotada por um gene é modificável. Se tal ocorre, pode ser via mutação na região reguladora e/ou na região codificadora do gene.

Capítulo 6

Conclusões

Nesta tese mostramos que o modelo estocástico binário para um gene auto-regulado possui solução analítica completa. A dinâmica probabilística é obtida em termos das funções de Heun confluentes. Como não há uma expressão fechada para as derivadas gerais destas funções, introduzimos uma relação de recorrência entre derivadas das mesmas.

Introduzimos a noção de simetria e invariância no campo da expressão gênica estocástica. Isto é obtido reescrevendo as equações de diferença-diferenciais em formalismo de operadores diferenciais no regime estacionário. Obtivemos uma álgebra $SO(2,1)$, construída explicitamente com operadores diferenciais. O invariante da álgebra caracteriza a meia-vida da dinâmica do gene. O regime de flutuações do processo estocástico é classificado pela relação entre os etiquetadores de um elemento da representação do grupo. O gene auto-regulado apresenta os regimes sub Fano, Fano e super Fano. O gene externamente regulado apresenta apenas os regimes Fano ou super Fano. Outra consequência das simetrias é que o modelo para o gene externamente regulado é um caso particular do modelo para o gene auto-regulado.

Fizemos uma análise comparativa entre os modelos para os genes auto e externamente regulados. Mostramos, partindo de equações estocásticas, que a auto-regulação representa uma estratégia de controle mais precisa. Também mostramos que o sistema auto-regulado responde com maior rapidez à perturbações por agentes externos.

Em perspectiva, há diversas ramificações deste trabalho. Faz-se mister encontrar uma

forma de ortogonalização das auto-funções do operador \mathbb{H} . Isto ampliará as possibilidades de sistemas dinâmicos descritos por este modelo, além de ser uma abordagem mais elegante. A existência de simetrias possibilita que se investigue um princípio de composição entre vários genes. Uma possibilidade bastante clara é que as proteínas do gene auto-regulado passem a regular o gene externamente regulado. Outra hipótese é impor representações de dimensões superiores à 2 da álgebra $u(2)$ que forma a base do processo de chaveamento. Neste caso, o gene possuiria um maior número de níveis de regulação. Os elementos de matriz são os elementos da álgebra de Heisenberg adida do termo zd/dz , veja Apêndice B. Ainda neste tópico, podemos investigar se a álgebra de Lie de dimensão infinita, advinda do modelo aqui tratado, possui significado biológico.

Do ponto-de-vista biológico há duas questões que podem ser levantadas diretamente. A primeira trata da verificação experimental da conservação da meia-vida da dinâmica de um gene. Resultados experimentais mostram que o sistema *operon lac* leva um tempo fixo para atingir regime estacionário, independente de seu estado inicial ser totalmente induzido ou não-induzido [25]. A segunda questão diz respeito às estratégias de regulação. Verificamos, a partir do modelo aqui apresentado, que um sistema externamente regulado surge como um caso particular do sistema auto-regulado. Daí, pode-se especular que a regulação externa de um gene, e portanto a interação entre genes distintos e a complexidade em redes gênicas, surgiu como uma transformação do sistema auto-regulado.

O modelo que apresentamos nesta tese é bastante simples para a modelagem de expressão gênica. Ele possui apenas dois níveis de regulação e omite o processo de transcrição. Porém, desta simplicidade advém seu forte, que é o de discriminar com bastante clareza os efeitos de duas estratégias de regulação amplamente adotadas na natureza.

Referências

- [1] PTASHNE, M. *A genetic switch: phage lambda and higher organisms*. 2ed. Cambridge: Cell press and blackwell science; 1992.
- [2] JACOB, F.; MONOD, J. Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, London, v.3, n.3, p. 318–56, 1961.
- [3] JAEGER, J. et al. Dynamical analysis of regulatory interactions in the gap gene system of *Drosophila melanogaster*. *Genetics*, Baltimore, v.167,n.4,p. 1721–37, 2004.
- [4] HARTWELL, L.H., et al From molecular to modular cell biology. *Nature*, London, v.402,n.6761,Suppl,p.C47–52, 1999.
- [5] SHEN-ORR, S.S., et al Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics*, New York,v.31,n.1,p.64–8, 2002.
- [6] ALON, U. Network motifs: theory and experimental approaches. *Nature Reviews Genetics*, London, v.8, n.6, p.450–461, 2007.
- [7] DELBRÜCK, M.. Statistical fluctuations in autocatalytic reactions. *Journal of Chemical Physics*, Melville, v.8, n.1, p.120–124, 1940.
- [8] BECSKEI, A., SERRANO, L. Engineering stability in gene networks by autoregulation. *Nature*, London, v.405, n.6786, p.590–3, 2000.
- [9] BECSKEI, A.; SERAPHIN, B.; SERRANO, L. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO Journal*, Oxford, v.20, n.10, p.2528–35, 2001.
- [10] SIMPSON, M.L., COX, C.D, SAYLER, G.S. Frequency domain analysis of noise in autoregulated gene circuits. *Proceedings of the National Academy of Sciences of the United States of America*, Washington,v.100,n.8,p.4551–6, 2003.
- [11] GARDNER, T.S., CANTOR, C.R., COLLINS, J.J. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, London,v.403,n.6767,p.339–42, 2000.
- [12] ELOWITZ, M.B., LEIBLER, S. A synthetic oscillatory network of transcriptional regulators. *Nature*, London, v.403,n.6767,p.335–8, 2000.

- [13] HOOSHANGI, S., THIBERGE, S., WEISS, R. Ultrasensitivity and noise propagation in a synthetic transcriptional cascade. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, v.102, n.10, p.3581–6, 2005.
- [14] PEDRAZA, J.M., VAN OUDENAARDEN, A. Noise propagation in gene networks. *Science*, Washington, v.307, n.5717, p.1965–9, 2005.
- [15] ELOWITZ, M.B. et al Stochastic gene expression in a single cell. *Science*, Washington, v.297, n.5584, p.1183–6, 2002.
- [16] RASER, J.M., O'SHEA, E.K. Control of stochasticity in eukaryotic gene expression. *Science*, Washington, v.304, n.5678, p.1811–4, 2004.
- [17] RASER, J.M., O'SHEA, E.K.. Noise in gene expression: origins, consequences, and control. *Science*, Washington, v.309, n.5743, p.2010–2013, 2005.
- [18] OZBUDAK, E.M. et al Regulation of noise in the expression of a single gene. *Nature Genetics*, New York, v.31, n.1, p.69–73, 2002.
- [19] KAERN, M., BLAKE, W.J., COLLINS, J.J. The engineering of gene regulatory networks. *Annual Review of Biomedical Engineering*, Palo Alto, v.5, p.179–206, 2003.
- [20] BLAKE, W.J. et al Noise in eukaryotic gene expression. *Nature*, London, v.422, n.6932, p.633–7, 2003.
- [21] ISAACS, F.J. et al Prediction and measurement of an autoregulatory genetic module. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, v.100, n.13, p.7714–9, 2003.
- [22] ROSENFELD, N. et al A fluctuation method to quantify in vivo fluorescence data. *Biophysical Journal*, Bethesda, v.91, n.2, p.759–66, 2006.
- [23] CAI, L., FRIEDMAN, N., XIE, X.S. Stochastic protein expression in individual cells at the single molecule level. *Nature*, London, v.440, n.7082, p.358–62, 2006.
- [24] BHALLA, U.S., IYENGAR, R. Emergent properties of networks of biological signaling pathways. *Science*, Washington, v.283, n.5400, p.381–7, 1999.
- [25] METTETAL, J.T. et al Predicting stochastic gene expression dynamics in single cells. *Proceedings of the National Academy of Sciences of the United States of America*, Washington, v.103, n.19, p.7304–9, 2006.
- [26] LITTLE, J.W., SHEPLEY, D.P., WERT, D.W. Robustness of a gene regulatory circuit. *EMBO Journal*, Oxford, v.18, n.15, p.4299–307, 1999.
- [27] 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.814–9, 1997.
- [28] ARKIN, A., ROSS, J., MCADAMS, H.H. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. *Genetics*, Baltimore, v.149, n.4, p.1633–48, 1998.

- [29] REINITZ, J., VAISNYS, J.R. Theoretical and experimental analysis of the phage lambda genetic switch implies missing levels of co-operativity. *Journal of Theoretical Biology*, London,v.145,n.3,p.295–318, 1990.
- [30] SANTILLAN, M., MACKEY, M.C. Influence of catabolite repression and inducer exclusion on the bistable behavior of the lac operon. *Biophysical Journal*, Bethesda,v.86,n.3,p.1282–92, 2004.
- [31] AURELL, E., SNEPPEN, K. Epigenetics as a first exit problem. *Physical Review Letters*, College PK,v.88,n.4,048101, 2002.
- [32] ZHU, X.M., et al Robustness, stability and efficiency of phage lambda genetic switch: dynamical structure analysis. *Journal of Bioinformatics Computational Biology*, London,v.2,n.4,p.785–817, 2004.
- [33] PECCOUD, J., YCART,B.. Markovian modelling of gene product synthesis. *Theoretical Population Biology*, San Diego,v.48,n.2,p.222–234,.
- [34] KEPLER, T.B., ELSTON, T.C. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophysical Journal*, Bethesda,v.81,n.6,p.3116–36, 2001.
- [35] 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.8614–9, 2001.
- [36] 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.2374–9, 2003.
- [37] HORNOS, J.E., et al Self-regulating gene: an exact solution. *Physical Review E*, College PK,v.72,n.5,Pt1,051907, 2005.
- [38] PAULSSON, J.. Models of stochastic gene expression. *Physics of Life Reviews*, Amsterdam,v.2,n.2,p.157–175, 2005.
- [39] TIAN, T., BURRAGE, K. Stochastic models for regulatory networks of the genetic toggle switch. *Proceedings of the National Academy of Sciences of the United States of America*, Washington,v.103,n.22,p.8372–7, 2006.
- [40] KAERN, M., et al Stochasticity in gene expression: *Nature Reviews Genetics*, London,v.6,n.6,p.451–64, 2005.
- [41] FRIEDMAN, N., CAI, L., XIE, X.S. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Physical Review Letters*, College PK,v.97,n.16,168302, 2006.
- [42] RAMOS, A.F., HORNOS, J.E.M. Symmetry and stochastic gene regulation. *Physical Review Letters*, College PK,v.99,n.10,108103, 2007.
- [43] INNOCENTINI GC, HORNOS JE. Modeling stochastic gene expression under repression. *Journal of Mathematical Biology*, New York,v.55,n.3,p.413–31, 2007.

- [44] KIM, K.Y., LEPZELTER, D., WANG, J.. Dynamics and intrinsic statistical fluctuations of a gene switch. *Journal of Physical Chemistry B*, Washington,v.111,n.34,p.10239-47, 2007.
- [45] KIM, K.Y., LEPZELTER, D., WANG, J. Single molecule dynamics and statistical fluctuations of gene regulatory networks: a repressilator. *Journal of Chemical Physics*, Melville,v.126,n.3,034702, 2007.
- [46] GILLESPIE, D.T.. Exact stochastic simulation of coupled chemical reactions. *Journal of Physical Chemistry*, Washington,v.81,n.25,p.2340–2351, 1977.
- [47] HASTY, J., et al. Computational studies of gene regulatory networks: in numero molecular biology. *Nature Reviews Genetics*, London,v.2,n.4,p.268–79, 2001.
- [48] ANDRIEUX, D., GASPARD, P.. Temporal disorder and fluctuation theorem in chemical reactions. *Physical Review E*, College PK,v.77,n.3,Pt1,031137, 2008.
- [49] YIN, L., AO, P.. Existence and construction of dynamical potential in nonequilibrium processes without detailed balance. *Journal of Physics A: Mathematical and General*, Bristol,v.39,n.27,p.8593–8601, 2006.
- [50] KWON, C., AO, P., THOULESS, D.J. Structure of stochastic dynamics near fixed points. *Proceedings of the National Academy of Sciences of the United States of America*, Washington,v.102,n.37,p.13029–33, 2005.
- [51] LU, T., et al. Phenotypic variability of growing cellular populations. *Proceedings of the National Academy of Sciences of the United States of America*, Washington,v.104,n.48,p.18982–7, 2007.
- [52] FOOTE, R.. Mathematics and complex systems. *Science*, Washington,v.318,n.5849,p.410–412, 2007.
- [53] COEN ES, et al. Evolution of floral symmetry. *Philosophical Transactions of the Royal Society of London Series B: Biological Sciences*, London,v.350,n.1331,p.35–38, 1995.
- [54] ENDRESS, P.K.. Evolution of floral symmetry. *Current Opinion in Plant Biology*, London,v.4,n.1,p.86–91, 2001.
- [55] NEAL, P.R., DAFNI, A., GIURFA, M.. Floral symmetry and its role in plant-pollinator systems: terminology, distribution, and hypotheses. *Annual Review of Ecology and Systematics*, Palo Alto,v.29,p.345–373, 1998.
- [56] FRAENKEL-CONRAT, H., SINGER, B., WILLIAMS, R.C.. Infectivity of viral nucleic acid. *Biochimica Biophysica Acta*, Amsterdam,v.25,n.1,p.87–96, 1957.
- [57] CRICK, F.H.C., WATSON, J.D.. Structure of small viruses. *Nature*, London,v.177,n.4506,p.473–475, 1956.
- [58] BARUT, A.O., KLEINERT, H. Transition probabilities of the hydrogen atom from noncompact dynamical groups. *Physical Review*, College PK,v.156,n.5,p.1541–1545, 1967.

- [59] PAULI, W. On the hydrogen spectrum from the standpoint of the new quantum mechanics. *Zeitschrift fur Physik*,v.36,n.5,p.336–363, 1926.
- [60] SLATER, J.C. The theory of complex spectra. *Physical Review*, College PK,v.34,n.10,p.1293–1322, 1929.
- [61] GELL-MANN, M., NE'EMAN, Y. The eightfold way. New York: Benjamin; 1964.
- [62] IACHELLO, F., LEVINE, R.D. Algebraic theory of molecules. Oxford: Oxford university press; 1995.
- [63] HORNOS, J.E.M., IACHELLO, F. The overtone spectrum of acetylene in the vibron model. *Journal of Chemical Physics*, Melville,v.90,n.10,p.5284–5291, 1989.
- [64] ERDÉLYI, A., BATEMAN, H. Higher transcendental functions. vol. 1-3 of Bateman Manuscript Project. New York: McGraw-Hill; 1953.
- [65] BECKER, P.A. Normalization integrals of orthogonal Heun functions. *Journal of Mathematical Physics*, Woodbury,v.38,n.7,p.3692–3699, 1997.
- [66] FIZIEV, P.P. Exact solutions of Regge-Wheeler equation and quasi-normal modes of compact objects. *Classical and Quantum Gravity*, Bristol,v.23,n.7,p.2447-68, 2006.
- [67] STAHLHOFEN, A.A. Susy, Gauss, Heun and physics: a magic square? *Journal of Physics A: Mathematical and General*, Bristol,v.37,n.43,p.10129–10138, 2004.
- [68] CHEB-TERRAB, E.S. Solutions for the general, confluent and biconfluent Heun equations and their connection with Abel equations. *Journal of Physics A: Mathematical and General*, Bristol,v.37,n.42,p.9923–9949, 2004.
- [69] KAZAKOV, A.Y. Symmetries of the confluent Heun equation. *Journal of Mathematical Sciences*, London,v.117,n.2,p.3918–3927, 2003.
- [70] CHEN, K., RAJEWSKY, N. The evolution of gene regulation by transcription factors and microRNAs. *Nature Reviews Genetics*, London,v.8,n.2,p.93–103, 2007.
- [71] DOEBLEY, J., LUKENS, L. Transcriptional regulators and the evolution of plant form. *Plant Cell*, Rockville,v.10,n.7,p.1075–82, 1998.
- [72] LEVINE, M., TJIAN, R. Transcription regulation and animal diversity. *Nature*, London,v.424,n.6945,p.147–51, 2003.
- [73] ORPHANIDES, G., REINBERG, D. A unified theory of gene expression. *Cell*, Cambridge,v.108,n.4,p.439–51, 2002.
- [74] LOPES, F.J.P., et al Spatial bistability generates hunchback expression sharpness in Drosophila embryo. *PLoS Computational Biology*, San Francisco, 2008. In Press.
- [75] GARDINER, C.W. Handbook of stochastic methods. 3ed. Berlin: Springer-Verlag, 2004.
- [76] WYBOURNE, B.G. Classical groups for physicists. New York: John Wiley; 1974.

- [77] BARUT, A.O.; RACZKA, R. Theory of group representations and applications. 2ed. Singapore: World scientific, 1986.
- [78] GILMORE, R. Lie groups, Lie algebras, and some of their representations. 3ed. New York: Dover, 2006.
- [79] ABRAMOWITZ, M.; STEGUN, I.A., ed. Handbook of mathematical functions. Washington: U.S. GPO, 1972.
- [80] LORENTE, M. Raising and lowering operators and their factorization for generalized orthogonal polynomials of hypergeometric type on homogeneous and non-homogeneous lattices. *Journal of Physics A: Mathematical and General*, Bristol,v.36,n.1,p.183–198, 2003.
- [81] MILLER W Jr. Lie theory and difference equations. II. *Journal of Mathematical Analysis and Applications*, San Diego,v.39,n.2,p.406–422, 1972.
- [82] FANO, U. Ionization yield of radiations II. The fluctuations of the number of ions. *Physical Review*, College PK,v.72,n.1,p.26–29, 1947.
- [83] MILO, R., et al Network motifs: simple building blocks of complex networks. *Science*, Washington,v.298,n.5594,p.824–827, 2002.
- [84] SAVAGEAU, M.A. Comparison of classical and autogenous systems of regulations in inducible operons. *Nature*, London,v.252,n.5484,p.546–549, 1974.
- [85] ROSENFELD, N., ELOWITZ, M.B., ALON, U. Negative autoregulation speeds the response times of transcription networks. *Journal of Molecular Biology*, London,v.323,n.5,p.785–93, 2002.
- [86] CAMAS, F.M., BLAZQUEZ, J., POYATOS, J.F. Autogenous and nonautogenous control of response in a genetic network. *Proceedings of the National Academy of Sciences of the United States of America*, Washington,v.103,n.34,p.12718–23, 2006.

Apêndice A

Derivação das autofunções

Neste apêndice apresentamos o cálculo da solução analítica do modelo estocástico de um gene auto-interagente. As equações diferenciais parciais e acopladas do modelo estocástico tratado nesta tese podem ser escritas como:

$$\frac{z_0}{\rho} \frac{\partial \psi}{\partial t} = \mathbb{H}\psi. \quad (\text{A.1})$$

As soluções da equação acima são da forma:

$$\tilde{\psi} \exp\left(-\frac{\rho\Lambda}{z_0}t\right),$$

onde, $\tilde{\psi}$ e Λ são, respectivamente, autofunção e autovalor de \mathbb{H} . As autofunções de \mathbb{H} são obtidas a partir de sua respectiva equação de autovalores:

$$\mathbb{H}\tilde{\psi} = -\Lambda\tilde{\psi}$$

onde

$$\psi = \begin{bmatrix} A \\ B \end{bmatrix}, \quad \text{e} \quad \mathbb{H} = \begin{bmatrix} M \frac{z-1}{1-z_0} - (z-z_0) \frac{d}{dz} & f \\ (1-z_0)z \frac{d}{dz} & \chi M \frac{z-1}{1-z_0} - f - z_0(z-1) \frac{d}{dz} \end{bmatrix}.$$

A ação de \mathbb{H} sobre $\tilde{\psi}$ resulta no seguinte par de equações diferenciais acopladas:

$$\left(M \frac{z-1}{1-z_0} + \Lambda\right) A - (z-z_0) \frac{dA}{dz} + fB = 0, \quad (\text{A.2})$$

$$\left(\chi M \frac{z-1}{1-z_0} - f + \Lambda\right) B - z_0(z-1) \frac{dB}{dz} + (1-z_0)z \frac{dA}{dz} = 0. \quad (\text{A.3})$$

A solução das equações acima é obtida isolando-se B na equação (A.2):

$$B = - \left(\frac{M}{f} \frac{z-1}{1-z_0} + \frac{\Lambda}{f}\right) A + \frac{z-z_0}{f} \frac{dA}{dz}, \quad (\text{A.4})$$

e, substituindo a expressão acima na equação (A.3), obtemos uma equação diferencial de segunda ordem:

$$\frac{d^2 A}{dz^2} + P \frac{dA}{dz} + QA = 0, \quad (\text{A.5})$$

sendo:

$$P = -\frac{M}{z_0} \frac{z_0 + \chi}{1-z_0} - \frac{\Lambda/z_0}{z-1} + \frac{1+f+M-\Lambda}{z-z_0}$$

e

$$Q = \frac{1}{(z-1)(z-z_0)} \left[\frac{M}{z_0} \frac{z-1}{1-z_0} \left(\chi M \frac{z-1}{1-z_0} + \Lambda + \chi\Lambda - f - z_0 \right) + \Lambda \frac{\Lambda-f}{z_0} \right]$$

A equação (A.5) possui três singularidades regulares nos pontos z_0 , 1 e infinito. Isto sugere uma solução em termos das funções de Heun confluentes. De fato, impondo:

$$A(z) = \exp\left(\frac{\chi M}{z_0} \frac{z-1}{1-z_0}\right) \tilde{A}(z),$$

na equação (A.5), obtemos:

$$\frac{d^2 \tilde{A}}{dz^2} + \tilde{P} \frac{d\tilde{A}}{dz} + \tilde{Q} \tilde{A} = 0, \quad (\text{A.6})$$

sendo

$$\tilde{P} = \frac{\Delta_1}{1-z_0} - \frac{\Lambda/z_0}{z-1} + \frac{1+b-\Lambda}{z-z_0}$$

e

$$\tilde{Q} = \frac{2\Delta_2 + \Delta_1(1 + b - \Lambda - \Lambda/z_0)}{2(1 - z_0)(z - 1)} + \frac{2\Delta_3 - 1 - \Lambda/z_0 + \Delta_1(\Lambda - b + a)}{(z - z_0)(z - 1)}$$

onde:

$$a = f \frac{1 - \chi}{z_0 - \chi} \quad \text{e} \quad b = f + M,$$

e

$$\begin{aligned} \Delta_1 &= M \frac{\chi - z_0}{z_0}, \\ \Delta_2(\Lambda) &= \frac{\Delta_1}{2} \left(1 + 2a - b - \Lambda \frac{1 - z_0}{z_0} \right), \\ \Delta_3(\Lambda) &= \frac{1}{2} - \Delta_2 - \frac{\Lambda}{z_0} \frac{1 - \Delta_1 + \Lambda - b}{2}. \end{aligned}$$

A solução da equação (A.6) é obtida em termos das funções Heun confluentes e podem ser escritas como se segue:

$$\tilde{A}(z) = C^{(1)}A^{(1)}(z) + C^{(2)}(z - z_0)^{\Lambda - b}A^{(2)}(z),$$

sendo

$$\begin{aligned} \tilde{A}^{(1)}(z) &= \mathcal{H}(\Delta_1, b - \Lambda, -1 - \Lambda/z_0, \Delta_2, \Delta_3, \frac{z - z_0}{1 - z_0}), \\ \tilde{A}^{(2)}(z) &= \mathcal{H}(\Delta_1, \Lambda - b, -1 - \Lambda/z_0, \Delta_2, \Delta_3, \frac{z - z_0}{1 - z_0}), \end{aligned}$$

em que \mathcal{H} representa a função de Heun confluyente.

A solução final das equações dependentes do tempo é obtida por calculando-se os conjuntos de autovalores. Há dois conjuntos e o primeiro deles, o número 1, é dado por:

$$\Lambda_j^{(1)} = jz_0, \quad \text{sendo } j = 0, 1, 2, \dots, \quad (\text{A.7})$$

e é obtido a partir de $\tilde{A}^{(1)}(z)$. Consideramos a transformação:

$$\mathcal{H}\left(\Delta_1, b - \Lambda, -1 - \Lambda/z_0, \Delta_2, \Delta_3, \frac{z - z_0}{1 - z_0}\right) = \left(\frac{1 - z}{1 - z_0}\right)^{1 + \Lambda/z_0} \mathcal{H}\left(\Delta_1, b - \Lambda, 1 + \Lambda/z_0, \Delta_2, \Delta_3, \frac{z - z_0}{1 - z_0}\right),$$

em $\tilde{A}^{(1)}$ e impomos analiticidade em torno de z igual à um. Isto implica no expoente de $(1 - z)$ ser um número inteiro positivo, ou seja:

$$1 + \Lambda/z_0 = j,$$

onde j é um inteiro positivo. A expressão (A.7) é obtida isolando-se Λ na relação acima e reescalando $j \rightarrow j + 1$, para que o autovalor seja sempre positivo.

O segundo conjunto de autovalores, o número 2, é dado por:

$$\Lambda_j^{(2)} = b + j, \quad \text{sendo } j = 0, 1, 2, \dots, \quad (\text{A.8})$$

e é obtido a partir de $\tilde{A}^{(2)}(z)$. Neste caso, impomos analiticidade em torno do ponto z_0 . Sendo assim, o expoente de $(z - z_0)$ satisfaz:

$$\Lambda - b = j,$$

onde j é um inteiro positivo. A expressão (A.8) é obtida isolando-se Λ na expressão acima.

Os conjuntos de autofunções correspondentes aos autovalores acima calculados assumem as respectivas formas:

$$\begin{aligned} \tilde{A}_j^{(1)}(z) &= \mathcal{H}\left(\Delta_1, b - jz_0, -1 - j, \Delta_2^{(j)}(1), \Delta_3^{(j)}(1), \frac{z - z_0}{1 - z_0}\right), \\ \tilde{A}_j^{(2)}(z) &= (z - z_0)^j \mathcal{H}\left(\Delta_1, j, -1 - (b + j)/z_0, \Delta_2^{(j)}(2), \Delta_3^{(j)}(2), \frac{z - z_0}{1 - z_0}\right), \end{aligned}$$

onde os parâmetros $\Delta_k^{(j)}(i)$ são obtidos explicitamente por substituição direta do autovalor

correspondente à cada uma das autofunções:

$$\Delta_k^{(j)}(i) = \Delta_k(\Lambda_j^{(i)}),$$

em que $i = \{1, 2\}$ e $k = \{2, 3\}$. A solução da equação (A.6) é dada por:

$$A_j^{(i)} = \exp\left(\frac{\chi M}{z_0} \frac{z-1}{1-z_0}\right) \tilde{A}_j^{(i)}.$$

As autofunções $B_j^{(i)}(z)$ são obtidas a partir da ação do operador (A.4) sobre uma autofunção $A_j^{(i)}$, substituindo-se o respectivo autovalor em lugar de Λ :

$$B_j^{(i)}(z) = \left(\frac{M}{f} \frac{1-z}{1-z_0} - \frac{\Lambda_j^{(i)}}{f} + \frac{z-z_0}{f} \frac{d}{dz} \right) A_j^{(i)}(z).$$

A solução final da equação (A.1) é dada por:

$$\begin{aligned} \alpha(z, t) &= \sum_{j=0}^{\infty} \sum_{i=1}^2 C_j^{(i)} A_j^{(i)} \exp\left(-\Lambda_j^{(i)} t\right), \\ \beta(z, t) &= \sum_{j=0}^{\infty} \sum_{i=1}^2 D_j^{(i)} B_j^{(i)} \exp\left(-\Lambda_j^{(i)} t\right). \end{aligned}$$

Apêndice B

Álgebra de Lie de dimensão infinita

Neste apêndice apresentamos a álgebra de Lie de dimensão infinita que é obtida a partir do regime estacionário do modelo estocástico apresentado. O operador \mathbb{H} definido anteriormente pode ser escrito na forma seguinte:

$$\mathbb{H} = \hat{e}_{11} \left(\frac{M}{z_0 - 1} a_0 + \frac{M}{1 - z_0} z + z_0 \frac{d}{dz} - z \frac{d}{dz} \right) + \hat{e}_{12} f a_0 + \hat{e}_{21} (1 - z_0) z \frac{d}{dz} + \hat{e}_{22} \left[\left(\frac{\chi M}{z_0 - 1} - f \right) a_0 + \frac{M}{1 - z_0} z + z_0 \frac{d}{dz} - z \frac{d}{dz} \right],$$

onde $\{M, z_0, \chi, f\}$ é o conjunto de parâmetros. Os operadores $\hat{e}_{\mu\nu}$ são matrizes 2×2 cujos elementos são $(\hat{e}_{\mu\nu})_{\iota\kappa} = \delta_{\mu\iota} \delta_{\nu\kappa}$, onde $\delta_{\mu\nu}$ é o símbolo de Kronecker e satisfaz:

$$\delta_{\mu\nu} = \begin{cases} 1, & \text{se } \mu = \nu, \\ 0, & \text{se } \mu \neq \nu. \end{cases}$$

Estas matrizes formam a base da álgebra $u(2)$, cujas relações de comutação são:

$$[\hat{e}_{\mu\nu}, \hat{e}_{\iota\kappa}] = \delta_{\nu\iota} \hat{e}_{\mu\kappa} - \delta_{\kappa\mu} \hat{e}_{\iota\nu}.$$

O operador a_0 representa a identidade e juntamente com z e $\frac{d}{dz}$ forma a álgebra de Heisenberg: $H \equiv \{a_0, z, \frac{d}{dz}\}$, com relações de comutação:

$$[a_0, z] = \left[a_0, \frac{d}{dz} \right] = 0, \quad \left[\frac{d}{dz}, z \right] = a_0.$$

O elemento $z\frac{d}{dz}$ é adicionado por meio de uma soma semi-direta à H . Isto porque H é um ideal da álgebra formada por $\bar{H} \equiv \{a_0, z, \frac{d}{dz}, z\frac{d}{dz}\}$. Isto é verificado a partir das relações de comutação de \bar{H} :

$$\left[a_0, z\frac{d}{dz} \right] = 0, \quad \left[\frac{d}{dz}, z\frac{d}{dz} \right] = -\frac{d}{dz}, \quad \left[z, z\frac{d}{dz} \right] = z,$$

em que o comutador de qualquer elemento de H com o operador zd/dz resulta num elemento de H .

Portanto, o operador \mathbb{H} é escrito na base:

$$a_0\hat{e}_{\mu\nu}, \quad z\hat{e}_{\mu\nu}, \quad \frac{d}{dz}\hat{e}_{\mu\nu}, \quad z\frac{d}{dz}\hat{e}_{\mu\nu}.$$

As relações de comutação entre os elementos acima fecham apenas entre eles. Assim, para que a álgebra do estado estacionário feche, a extendemos tal que um elemento arbitrário da mesma é dado por:

$$\hat{e}_{\mu\nu}z^i\frac{d^p}{dz^p},$$

em que $i \geq 0$ e $p \geq 0$ são números inteiros. Sendo assim, esta álgebra é formada por infinitos elementos e terá dimensão infinita.

As relações de comutação da álgebra acima são dadas por:

$$\begin{aligned} \left[\hat{e}_{\mu\nu}z^i\frac{d^p}{dz^p}, \hat{e}_{\nu\kappa}z^j\frac{d^q}{dz^q} \right] &= \delta_{\nu\iota}\hat{e}_{\mu\kappa} \sum_{l=0}^p \binom{p}{l} \frac{j!A_{j,l}}{(j-l)!} z^{i+j-l} \frac{d^{p+q-l}}{dz^{p+q-l}} \\ &\quad - \delta_{\kappa\mu}\hat{e}_{\nu\iota} \sum_{l=0}^q \binom{q}{l} \frac{i!A_{i,l}}{(i-l)!} z^{i+j-l} \frac{d^{p+q-l}}{dz^{p+q-l}}, \end{aligned}$$

em que,

$$A_{j,l} = \begin{cases} 1, & \text{se } j \geq l, \\ 0, & \text{se } j < l. \end{cases}$$

Apêndice C

Relação de recorrência para derivadas n -ésimas

Neste apêndice apresentamos o algoritmo utilizado para o cálculo da n -ésima derivada das funções geratrizes das probabilidades dependentes do tempo. Consideramos a equação diferencial de segunda ordem para $A(z)$ na seguinte forma:

$$\sum_{i=0}^2 R_i z^i \frac{d^2 A}{dz^2} + \sum_{i=0}^2 P_i z^i \frac{dA}{dz} + \sum_{i=0}^2 Q_i z^i A = 0, \quad (\text{C.1})$$

onde X_i são coeficientes reais ou complexos. A equação acima apresenta três singularidades regulares, uma no infinito e as duas restantes nos pontos:

$$z^{\pm} = -\frac{R_1}{2R_2} \pm \sqrt{\frac{R_1^2 - 4R_2R_0}{4R_2^2}}.$$

A equação (C.1) será uma equação de Heun confluyente em forma canônica se

$$R_2 = 1, \quad R_1 = -1, \quad R_0 = 0, \quad \text{e} \quad Q_2 = 0.$$

A equação (C.1) será equivalente à equação (A.5) se:

$$\begin{aligned}
R_2 &= 1, & R_1 &= -1 - z_0, & R_0 &= z_0, \\
P_2 &= -\frac{M}{z_0} \frac{z_0 + \chi}{1 - z_0}, & P_1 &= P_2 R_1 + 1 + f + M + \Lambda \frac{R_1}{z_0}, & P_0 &= -P_1 - P_2 + \Lambda \frac{z_0 - 1}{z_0}, \\
Q_2 &= \frac{\chi}{z_0} \frac{M^2}{(1 - z_0)^2}, & Q_1 &= -2Q_2 + \frac{M}{z_0} \frac{\Lambda + \chi\Lambda - f - z_0}{1 - z_0}, & Q_0 &= -Q_1 - Q_2 + \Lambda \frac{\Lambda - f}{z_0}.
\end{aligned} \tag{C.2}$$

A ação do operador $\frac{d^n}{dz^n}$ sobre a equação (C.1), com o auxílio do Teorema de Leibniz, relação 3.3.8 da referência [79], resulta na seguinte relação de recorrência:

$$A^{(n+2)} = -\frac{\Gamma_1 A^{(n+1)} + \Gamma_2 A^{(n)} + \Gamma_3 A^{(n-1)} + \Gamma_4 A^{(n-2)}}{\sum_{i=0}^2 R_i z^i}, \quad n = 0, 1, 2, \dots \tag{C.3}$$

onde o símbolo $A^{(n)}$ representa a derivada n -ésima da função A calculada no ponto z . Os coeficientes das derivadas são:

$$\begin{aligned}
\Gamma_1 &= 2R_2 n z + R_1 n + \sum_{i=0}^2 P_i z^i, \\
\Gamma_2 &= R_2 n(n-1) + 2P_2 n z + P_1 n + \sum_{i=0}^2 Q_i z^i, \\
\Gamma_3 &= P_2 n(n-1) + 2Q_2 n z + Q_1 n, \\
\Gamma_4 &= Q_2 n(n-1).
\end{aligned}$$

A efetuação dessa relação de recorrência requer que a função e sua derivada primeira no ponto z sejam conhecidas.

Note que esta relação de recorrência, se calculada nos pontos z^\pm , reduz-se à:

$$A^{(n+1)} = -\frac{\Gamma_2 A^{(n)} + \Gamma_3 A^{(n-1)} + \Gamma_4 A^{(n-2)}}{\Gamma_1}. \tag{C.4}$$

Em particular, podemos calcular as derivadas n -ésimas das funções $B(z)$, definidas na

equação (A.4). Neste caso teremos:

$$B^{(n)} = \frac{z - z_0}{f} A^{(n+1)} + \left(\frac{M}{f} \frac{1 - z}{1 - z_0} + \frac{\Lambda}{f} + \frac{n}{f} \right) A^{(n)} - \frac{M}{f} \frac{n}{1 - z_0} A^{(n-1)}, \quad n = 0, 1, 2, \dots,$$

e no cálculo das derivadas n -ésimas de $A(z)$, tomamos em conta as relações (C.2).