

---

# MÉTODO DE DIFERENÇAS TEMPORAIS APLICADO ÀS EQUAÇÕES DE RICCATI ACOPLADAS ENTRE SI

Oswaldo L. V. Costa\*  
oswaldo@lac.usp.br

Julio C.C. Aya\*  
julio@lac.usp.br

\*Departamento de Engenharia de Telecomunicações e Controle, Escola Politécnica da Universidade de São Paulo  
CEP: 05508 900 São Paulo SP Brazil

---

## ABSTRACT

In this paper we present an iterative technique based on Monte Carlo simulations for deriving the optimal control of the infinite horizon linear regulator problem of discrete-time Markovian jump linear systems for the case in which the transition probability matrix of the Markov chain is not known. It is well known that the optimal control of this problem is given in terms of the maximal solution of a set of coupled algebraic Riccati equations (CARE), which have been extensively studied over the last few years. We trace a parallel with the theory of  $TD(\lambda)$  algorithms for Markovian decision processes to develop a  $TD(\lambda)$  like algorithm for the optimal control associated to the maximal solution of the CARE. Some numerical examples are also presented.

**KEYWORDS:** Monte carlo simulations, coupled algebraic riccati equations, jump systems, optimal control.

## RESUMO

Neste trabalho apresentaremos uma técnica iterativa baseada em simulações de Monte Carlo para calcular o controle ótimo de um problema de regulador linear quadrático de horizonte infinito para um sistema linear com saltos Markovianos a tempo discreto, quando a matriz de transição de probabilidade não é conhecida. Sabemos que o controle ótimo deste problema é dado em termos da solução maximal de um conjunto de equações algébricas de Riccati acopladas entre si

(EARA) a tempo discreto, que foram extensivamente estudadas nos últimos anos. Traçaremos um paralelo com a teoria do algoritmo  $TD(\lambda)$  para Processos Markovianos de Decisão (PMD) para desenvolver o algoritmo  $TD(\lambda)$  para o controle ótimo associado à solução maximal de uma EARA.

**PALAVRAS-CHAVE:** :

Simulações de monte carlo, equações algébricas de Riccati acopladas entre si, sistemas com saltos, controle ótimo.

## 1 INTRODUÇÃO

Neste trabalho consideraremos o seguinte modelo em um espaço probabilístico  $(\Omega, P, \{\mathcal{F}_k\}, \mathcal{F})$  apropriado, conhecido na literatura internacional como sistemas lineares com saltos Markovianos a tempo discreto (vide (Mariton, 1990)):

$$\begin{aligned}x(k+1) &= A_{\theta(k)}x(k) + B_{\theta(k)}u(k) \\x(0) &= x_0, \theta(0) = \theta_0,\end{aligned}\tag{1}$$

onde  $\theta(k)$  é uma cadeia de Markov tomando valores em  $\{1, \dots, N\}$  com a matriz de transição de probabilidade  $\mathcal{P} = (p_{ij})$ . Seja  $\Sigma = \{u = (u(0), \dots); u(k) \text{ é } \mathcal{F}_k\text{-mensurável para cada } k\}$ . Para  $u \in \Sigma$ , considere o seguinte funcional quadrático para o sistema (1):

$$J_{(x_0, \theta_0)}(u) = E_{(x_0, \theta_0)} \left( \sum_{k=0}^{\infty} \|C_{\theta(k)}x(k)\|^2 + \|D_{\theta(k)}u(k)\|^2 \right)\tag{2}$$

Deseja-se minimizar (2) sob  $u \in \Sigma$ . Nos artigos ((Costa e Fragoso, 1995), (Ji e Chizeck, 1988), (Ji et al., 1991)) mostra-se que a solução deste problema esta associada à existência

---

Artigo submetido em 14/12/2000

1a. Revisão em 19/12/2002

Aceito sob recomendação do Ed. Assoc. Prof. Liu Hsu

de uma solução  $P = (P_1, \dots, P_N)$ ,  $P_i \geq 0$   $i = 1, \dots, N$ , do seguinte conjunto de equações algébricas de Riccati acopladas entre si (EARA), para  $i = 1, \dots, N$

$$X_i = A_i' \mathcal{E}_i(X) A_i + C_i' C_i - A_i' \mathcal{E}_i(X) B_i (B_i' \mathcal{E}_i(X) B_i + D_i' D_i)^{-1} B_i' \mathcal{E}_i(X) A_i \quad (3)$$

onde  $\mathcal{E}(X) = (\mathcal{E}_1(X), \dots, \mathcal{E}_N(X))$  é definido, para  $X = (X_1, \dots, X_N)$  como

$$\mathcal{E}_i(X) = \sum_{j=1}^N p_{ij} X_j. \quad (4)$$

Se tal solução  $P$  existe, pode-se mostrar (vide (Costa e Fragoso, 1995)) que a lei de controle ótima para o problema dado pelas equações (1), (2) e (3) é obtida pela lei de controle de realimentação

$$u(k) = F_{\theta(k)} x(k) \quad (5)$$

onde  $F = (F_1, \dots, F_N)$  é dado por

$$F_i = -(B_i' \mathcal{E}_i(P) B_i + D_i' D_i)^{-1} B_i' \mathcal{E}_i(P) A_i. \quad (6)$$

Associado aos problemas de horizonte infinito (1), (2), temos o problema de horizonte finito (7) no qual, para algum  $S = (S_1, \dots, S_N) \in \mathbb{H}^{n+}$ , deseja-se encontrar um  $u(0), \dots, u(\kappa)$  tal que minimize o seguinte funcional

$$J_{(x_0, \theta_0)}(u) = E_{(x_0, \theta_0)} \left( \sum_{k=0}^{\kappa} (\|C_{\theta(k)} x(k)\|^2 + \|D_{\theta(k)} u(k)\|^2) + x(\kappa + 1)' S_{\theta(\kappa+1)} x(\kappa + 1) \right). \quad (7)$$

Como é provado em (Costa e Fragoso, 1995), a solução deste problema pode ser obtida do seguinte conjunto de equações a diferença de Riccati para  $k = \kappa, \dots, 0$ ,

$$\begin{aligned} P_i(k) &= A_i' \mathcal{E}_i(P(k+1)) A_i + C_i' C_i \\ &\quad - A_i' \mathcal{E}_i(P(k+1)) B_i (B_i' \mathcal{E}_i(P(k+1)) B_i \\ &\quad + D_i' D_i)^{-1} B_i' \mathcal{E}_i(P(k+1)) A_i \\ &= (A_i + B_i F_i(k))' \mathcal{E}_i(P(k+1)) (A_i + B_i F_i(k)) \\ &\quad + C_i' C_i + F_i(k)' D_i' D_i F_i(k) \end{aligned} \quad (8)$$

onde  $P(\kappa + 1) = S$  e

$$F_i(k) = -(B_i' \mathcal{E}_i(P(k+1)) B_i + D_i' D_i)^{-1} B_i' \mathcal{E}_i(P(k+1)) A_i. \quad (9)$$

A lei de controle ótima de (7) é dada por

$$u(k) = F_{\theta(k)}(k) x(k)$$

e sob algumas condições (vide (Costa e Fragoso, 1995)),  $P(k)$  converge para uma solução semi-definida positiva  $P$  de (3).

Traçando um paralelo entre a teoria acima e a teoria de processos Markovianos de decisão (PMD), poderíamos relacionar as iterações (8),(9) com a técnica chamada de iteração de valores em PMD. Por outro lado, como será apresentado no Lema 5 abaixo (vide Observação 1), o método chamado de quase-linearização para obter a solução de (3) pode ser visto como a técnica de iteração de estratégias para PMD, envolvendo a parte de avaliação de estratégia e melhoramento de estratégia.

O método do  $TD(\lambda)$  tem sido aplicado para resolver problemas relacionados a PMD (veja por exemplo (Bertsekas e Tsitsiklis, 1996),(Sutton e Barto, 1998)), para o caso onde a matriz de transição de probabilidade da cadeia de Markov é desconhecida. Neste caso assumimos que se conhece o histórico do processo, isto é, que existe um conjunto de trajetórias simuladas do sistema, e que a função de custo de uma dada estratégia é calculada iterativamente através do histórico do processo.

A meta deste trabalho é aplicar o método  $TD(\lambda)$  para obter o controle ótimo (5), (6) associado à solução  $P$  do EARA dado pelas equações (3) e (4) para o caso em que a matriz de transição de probabilidade  $\mathcal{P}$  é desconhecida, sendo possível simular as trajetórias para a cadeia de Markov  $\theta(k)$ . Para o caso em que a matriz de probabilidade é conhecida, existe uma série de algoritmos numéricos para obter a solução maximal  $P$  do EARA (3) e (4) (veja ((Abou-Kandil et al., 1995), (Ait-Rami e Ghaoui, 1996), (Blair e Sworder, 1975), (Costa e Fragoso, 1995), (Costa e Marques, 1999), (Costa et al., 1997), (Gajic e Borno, 1995), (Ji e Chizeck, 1988), (Ji et al., 1991), (Mariton, 1990), (Val et al., 1998), (Costa e Aya, 1999)). Para o caso no qual a matriz de transição de probabilidade  $\mathcal{P}$  é não conhecida, propõe-se um algoritmo semelhante ao da avaliação de estratégias de Monte Carlo que atualiza as estimativas incrementalmente para obter o controle ótimo  $F$  dado pelas equações (5), (6). Obtém-se a prova deste convergência, seguindo os mesmos argumentos que em (Bertsekas e Tsitsiklis, 1996).

Este trabalho é apresentado do seguinte modo: na Seção 2 apresentamos algumas notações e resultados preliminares. Na seção 3 descrevemos o método  $TD(\lambda)$  para resolver a EARA (3). O resultado principal é o Teorema 7, que prova a convergência do método apresentado nesta seção. Na Seção 4 mostraremos um exemplo numérico, e na Seção 5 os comentários finais.

## 2 NOTAÇÃO E RESULTADOS PRELIMINARES

Para  $\mathbb{X}$  e  $\mathbb{Y}$  espaços de Banach, definimos  $\mathbb{B}(\mathbb{X}, \mathbb{Y})$  o espaço de Banach de todos os operadores lineares de  $\mathbb{X}$  em  $\mathbb{Y}$ , com a norma uniforme induzida representada por  $\|\cdot\|$ . Por simplicidade usaremos  $\mathbb{B}(\mathbb{X}) := \mathbb{B}(\mathbb{X}, \mathbb{X})$ . O raio espectral de um operador  $\mathcal{G} \in \mathbb{B}(\mathbb{X})$  será denotado por  $r_\sigma(\mathcal{G})$ . Se  $\mathbb{X}$  é um espaço de Hilbert o produto interno será denotado por  $\langle \cdot; \cdot \rangle$ , e para  $\mathcal{G} \in \mathbb{B}(\mathbb{X})$ ,  $\mathcal{G}^*$  denotará o operador adjunto de  $\mathcal{G}$ . Como é usual,  $\mathcal{G} \geq 0$  ( $\mathcal{G} > 0$  respectivamente) denotará que o operador  $\mathcal{G} \in \mathbb{B}(\mathbb{X})$  é positivo semi-definido (positivo definido). Em particular denotaremos o espaço complexo  $n$ -dimensional por  $\mathbb{C}^n$  e por  $\mathbb{B}(\mathbb{C}^n, \mathbb{C}^m)$  o espaço linear de todas as matrizes  $m \times n$  com norma limitada, e  $\mathbb{B}(\mathbb{C}^n) := \mathbb{B}(\mathbb{C}^n, \mathbb{C}^n)$  e o produto interno em  $\mathbb{B}(\mathbb{C}^n)$  é dado por  $\langle H; V \rangle = \text{tr}\{H * V\}$  para  $H, V \in \mathbb{B}(\mathbb{C}^n)$ . O superescrito  $'$  denotará a transposta de uma matriz. Usamos neste trabalho  $\|\cdot\|_2$  para a norma Euclidiana em  $\mathbb{C}^n$  e a norma induzida em  $\mathbb{B}(\mathbb{C}^n)$ .

Neste trabalho, serão ainda empregados os seguintes resultados:

**Lema 1** *Seja  $\mathbb{X}$  um espaço de Hilbert e  $\mathcal{G} \in \mathbb{B}(\mathbb{X})$ . As seguintes afirmações são equivalentes:*

- $r_\sigma(\mathcal{G}) < 1$
- existe um  $\mathcal{W} \in \mathbb{B}(\mathbb{H}^n)$  invertível e  $\mathcal{Q} \in \mathbb{B}(\mathbb{X})$  tal que  $\|\mathcal{Q}\| < 1$  e  $\mathcal{G} = \mathcal{W}\mathcal{Q}\mathcal{W}^{-1}$ .

**Prova:** Corolário 1.14 de (Kubrusly, 1997), páginas 31-32. ■

**Lema 2** *Seja  $\mathbb{X}$  um espaço de Hilbert e  $\mathcal{G} \in \mathbb{B}(\mathbb{X})$ . Se  $r_\sigma(\mathcal{G}) < 1$  então para cada  $Q \in \mathbb{X}$  existe uma única solução  $S(Q) \in \mathbb{X}$  para o sistema em  $X$*

$$X - \mathcal{G}(X) = Q. \quad (10)$$

Além disso

$$S(Q) = \sum_{k=0}^{\infty} \mathcal{G}^k(Q). \quad (11)$$

**Prova:** Veja Teorema 5.17, página 102 em (Weidman, 1980). ■

O conjunto  $\mathbb{H}^{n,m}$  é formado pelo espaço linear de todas as  $N$ -seqüência de matrizes complexas  $V = (V_1, \dots, V_N)$  com  $V_i \in \mathbb{B}(\mathbb{C}^n, \mathbb{C}^m)$ ,  $i = 1, \dots, N$ , e por simplicidade  $\mathbb{H}^n = \mathbb{H}^{n,n}$ . Neste trabalho consideraremos o conjunto  $\mathbb{H}^n$  equipado com o seguinte produto interno: para

$H = (H_1, \dots, H_N)$ ,  $V = (V_1, \dots, V_N) \in \mathbb{H}^n$  definimos o produto interno  $\langle \cdot; \cdot \rangle$  em  $\mathbb{H}^n$  como segue:

$$\langle H; V \rangle := \sum_{i=1}^N \text{tr}\{H_i^* V_i\} \quad (12)$$

e  $\|H\|^2 := \langle H; H \rangle$  (de forma que  $\|H\|^2 = \sum_{i=1}^N \|H_i\|^2$ ). Então com o produto interno acima definido (vide equação (12)),  $\mathbb{H}^n$  é um espaço de Hilbert. Nos também temos a seguinte norma  $\|\cdot\|_{max}$  em  $\mathbb{H}^n$ . Para  $H = (H_1, \dots, H_N) \in \mathbb{H}^n$ ,  $\|H\|_{max}$  é definido por

$$\|H\|_{max} := \max\{\|H_i\|_2; i = 1, \dots, N\}.$$

Definimos

$$\mathbb{H}^{n+} := \{V = (V_1, \dots, V_N) \in \mathbb{H}^n; V_i \geq 0, \forall i\}$$

e escrevemos que  $V \geq S$  onde  $V = (V_1, \dots, V_N) \in \mathbb{H}^n$  e  $S = (S_1, \dots, S_N) \in \mathbb{H}^n$ , se

$$V - S = (V_1 - S_1, \dots, V_N - S_N) \in \mathbb{H}^{n+},$$

e que  $V > S$  se  $V_i - S_i > 0$  para  $i = 1, \dots, N$ . Para,  $\Gamma = (\Gamma_1, \dots, \Gamma_N) \in \mathbb{H}^n$  definimos os seguintes operadores,  $\mathcal{L}(\cdot) = (\mathcal{L}_1(\cdot), \dots, \mathcal{L}_N(\cdot)) \in \mathbb{B}(\mathbb{H}^n)$  e  $\mathcal{G}(\cdot) = (\mathcal{G}_1(\cdot), \dots, \mathcal{G}_N(\cdot)) \in \mathbb{B}(\mathbb{H}^n)$  para  $V = (V_1, \dots, V_N) \in \mathbb{H}^n$  e  $i, j = 1, \dots, N$ ,

$$\mathcal{L}_i(V) := \Gamma_i' \mathcal{E}_i(\cdot) \Gamma_i \quad (13)$$

$$\mathcal{G}_j(V) := \sum_{i=1}^N p_{ij} \Gamma_i V_i \Gamma_i' \quad (14)$$

onde o operador  $\mathcal{E}(\cdot) = (\mathcal{E}_1(\cdot), \dots, \mathcal{E}_N(\cdot)) \in \mathbb{B}(\mathbb{H}^n)$  é definido como em (4). É simples verificar que os operadores  $\mathcal{E}$ ,  $\mathcal{L}$ , e  $\mathcal{G}$  mapeiam  $\mathbb{H}^{n+}$  em  $\mathbb{H}^{n+}$ , e que  $r_\sigma(\mathcal{L}) = r_\sigma(\mathcal{G})$ . No artigo (Costa e Fragoso, 1993) se prova que  $r_\sigma(\mathcal{L}) = r_\sigma(\mathcal{G})$  (de fato  $\mathcal{L} = \mathcal{G}^*$  de acordo com a equação (12)). Também definimos o operador  $\mathcal{T} \in \mathbb{B}(\mathbb{H}^n)$  como

$$\mathcal{T}_i(V) := \sum_{j=1}^N p_{ij} \Gamma_j^* V_j \Gamma_j \quad (15)$$

onde, de novo,  $V = (V_1, \dots, V_N) \in \mathbb{H}^n$  e  $\mathcal{T}(V) = (\mathcal{T}_1(V), \dots, \mathcal{T}_N(V))$ . No artigo de (Costa e Fragoso, 1993) provou-se que  $r_\sigma(\mathcal{G}) = r_\sigma(\mathcal{L}) = r_\sigma(\mathcal{T})$ .

Assumimos no modelo (1) e na função de custo (2) que  $A = (A_1, \dots, A_N) \in \mathbb{H}^n$ ,  $B = (B_1, \dots, B_N) \in \mathbb{H}^{m,n}$ ,  $C = (C_1, \dots, C_N) \in \mathbb{H}^{n,p}$  e  $D = (D_1, \dots, D_N) \in \mathbb{H}^{m,p}$ . Como for mostrado em (Costa e Fragoso, 1993), (Mariton, 1988), o modelo (1) com  $u(k) = F_{\theta(k)} x(k)$ , e  $V_i(k) = E(x(k)x(k)^* 1_{\{\theta(k)=i\}})$ ,  $V(k) = (V_1(k), \dots, V_N(k)) \in \mathbb{H}^{n+}$  leva a

$$V(k+1) = \mathcal{G}(V(k)), \quad k = 0, 1, \dots$$

onde  $\Gamma_i = A_i + B_i F_i$  em (14), e  $E\left(\|x(k)\|^2\right) = \sum_{i=1}^N \text{tr}\{V_i(k)\}$ .

O próximo passo é definir o operador para o 4<sup>essimo</sup> momento de  $x(t)$ . Para isto usamos o produto de Kronecker  $L \otimes K \in \mathbb{B}(\mathbb{C}^{n^2})$  para  $L, K \in \mathbb{B}(\mathbb{C}^n)$  e o operador  $\text{vec}\{\cdot\} : \mathbb{B}(\mathbb{C}^n) \mapsto \mathbb{C}^{n^2}$  definido da forma comum (veja (Brewer, 1978)). Seja o operador  $\mathcal{H} \in \mathbb{B}(\mathbb{H}^{n^2})$  da seguinte forma: para  $S = (S_1, \dots, S_N) \in \mathbb{H}^{n^2}$ ,  $\mathcal{H}(S) = (\mathcal{H}_1(S), \dots, \mathcal{H}_N(S))$  é definido como

$$\mathcal{H}_j(S) := \sum_{i=1}^N p_{ij} (\bar{\Gamma}_i \otimes \Gamma_i) S_i (\Gamma_i' \otimes \Gamma_i^*), \quad (16)$$

onde  $S_i(k) = E(\text{vec}\{x(k)x(k)^*\} \text{vec}\{x(k)x(k)^*\}^* 1_{\{\theta(k)=i\}})$ ,  $S(k) = (S_1(k), \dots, S_N(k)) \in \mathbb{H}^{n^2+}$ . Temos o seguinte resultado:

**Lema 3**  $S(k+1) = \mathcal{H}(S(k))$  e  $E\left(\|x(k)\|^4\right) = \sum_{i=1}^N \text{tr}\{S_i(k)\}$ .

**Prova:** Como  $x(k+1) = \Gamma_{\theta(k)} x(k)$ , é fácil de se verificar que

$$x_j(k+1)x_j(k+1)^* = 1_{\{\theta(k+1)=j\}} \sum_{i=1}^N \Gamma_i x_i(k)x_i(k)^* \Gamma_i^*$$

onde  $x_i(k) := x(k)1_{\{\theta(k)=i\}}$ . Seja  $z_i(k) = \text{vec}\{x_i(k)x_i(k)^*\}$ ,  $i = 1, \dots, N$ . Depois de algumas manipulações segue que

$$z_j(k+1)z_j(k+1)^* = 1_{\{\theta(k+1)=j\}} \cdot \sum_{i=1}^N (\bar{\Gamma}_i \otimes \Gamma_i) z_i(k)z_i(k)^* (\Gamma_i' \otimes \Gamma_i^*)$$

e escrevemos

$$S_i(k) = E(z_i(k)z_i(k)^* 1_{\{\theta(k)=i\}}).$$

Conseqüentemente temos que

$$S_j(k+1) = \sum_{i=1}^N p_{ij} (\bar{\Gamma}_i \otimes \Gamma_i) S_i(k) (\Gamma_i' \otimes \Gamma_i^*).$$

Finalmente, note que

$$\begin{aligned} \text{tr}\{z_i(k)z_i(k)^*\} &= \|x_i(k)\|^2 \text{tr}\{x_i(k)x_i(k)\} \\ &= \|x_i(k)\|^4. \end{aligned}$$

Vamos definir o conceito de estabilidade que será usado nas seções seguintes.

**Definição 1** Diremos que  $F = (F_1, \dots, F_N) \in \mathbb{H}^{n,m}$  estabiliza  $(A, B)$  no sentido da média quadrática se, quando fazemos  $u(k) = F_{\theta(k)} x(k)$  no sistema (1), temos que  $E\left(\|x(k)\|^2\right) \rightarrow 0$  quando  $k \rightarrow \infty$  para qualquer condição inicial de  $x(0)$  e  $\theta(0)$ . Diremos que  $(A, B)$  é estabilizável na média quadrática se para algum  $F = (F_1, \dots, F_N) \in \mathbb{H}^{n,m}$ , temos que  $F$  estabiliza  $(A, B)$  no sentido da média quadrática.

O seguinte resultado, provado em (Costa e Fragoso, 1993), mostra que  $F = (F_1, \dots, F_N)$  estabiliza o sistema (1) no sentido da média quadrática se e somente se o raio espectral do operador (14) (ou (13)) em malha fechada é menor que 1.

**Lema 4**  $F = (F_1, \dots, F_N) \in \mathbb{H}^{n,m}$  estabiliza  $(A, B)$  no sentido da média quadrática se e somente se  $r_\sigma(\mathcal{G}) < 1$ , onde  $\mathcal{G}$  é como em (14) com  $\Gamma_i = A_i + B_i F_i$ .

**Definição 2** Definimos  $\mathcal{F}(\cdot) = (\mathcal{F}_1(\cdot), \dots, \mathcal{F}_N(\cdot)) : \mathbb{H}^{n+} \rightarrow \mathbb{H}^{n,m}$ ,  $\mathcal{V}(\cdot) = (\mathcal{V}_1(\cdot), \dots, \mathcal{V}_N(\cdot)) : \mathbb{H}^{n,m} \rightarrow \mathbb{H}^n$  e  $\mathcal{R} = (\mathcal{R}_1(\cdot), \dots, \mathcal{R}_N(\cdot)) : \mathbb{H}^{n+} \rightarrow \mathbb{H}^{n+}$  como

$$\begin{aligned} \mathcal{F}_i(X) &:= -(B_i' \mathcal{E}_i(X) B_i + D_i' D_i)^{-1} B_i' \mathcal{E}_i(X) A_i \\ \mathcal{V}_i(F) &:= C_i' C_i + F_i' D_i' D_i F_i \\ \mathcal{R}_i(X) &:= A_i' \mathcal{E}_i(X) A_i + C_i' C_i \\ &\quad - A_i' \mathcal{E}_i(X) B_i (B_i' \mathcal{E}_i(X) B_i \\ &\quad + D_i' D_i)^{-1} B_i' \mathcal{E}_i(X) A_i \end{aligned} \quad (17)$$

onde  $X = (X_1, \dots, X_N) \in \mathbb{H}^{n+}$ ,  $F = (F_1, \dots, F_N) \in \mathbb{H}^{n,m}$ .

A seguinte identidade será útil nas próximas seções: para qualquer  $F = (F_1, \dots, F_N) \in \mathbb{H}^{n,m}$ , temos que

$$\begin{aligned} (A_i + B_i F_i)' \mathcal{E}_i(S) (A_i + B_i F_i) + F_i' D_i' D_i F_i \\ = (A_i + B_i \mathcal{F}_i(S))' \mathcal{E}_i(S) (A_i + B_i \mathcal{F}_i(S)) \\ + \mathcal{F}_i(S)' D_i' D_i \mathcal{F}_i(S) \\ + (F_i - \mathcal{F}_i(S))' (B_i' \mathcal{E}_i(S) B_i + D_i' D_i) (F_i - \mathcal{F}_i(S)). \end{aligned} \quad (18)$$

O seguinte Lema provado em (Costa e Marques, 1999), provê a existência de uma solução maximal para (3) quando  $(A, B)$  é estável na média quadrática. É baseado na técnica de quasi-linearização para a EARA (3), e paralelamente à técnica de iteração de estratégias para PMD (vide Observação 1 abaixo).

**Lema 5** Suponha que  $(A, B)$  é estabilizável na média quadrática e considere  $F^0 = (F_1^0, \dots, F_N^0) \in \mathbb{H}^{n,m}$  tal que estabiliza  $(A, B)$  no sentido da média quadrática. Então para  $l = 0, 1, 2, \dots$ , existe  $X^l = (X_1^l, \dots, X_N^l)$  que satisfaz às seguintes propriedades:

a)  $X^0 \geq X^1 \geq \dots \geq X^l \geq X$ , para um  $X \in \mathbb{H}^{n+}$  arbitrário tal que  $X \geq \mathcal{R}(X)$ .

b)  $r_\sigma(\mathcal{L}^l) < 1$ , onde  $\mathcal{L}^l(\cdot) = (\mathcal{L}_1^l(\cdot), \dots, \mathcal{L}_N^l(\cdot))$  e para  $i = 1, \dots, N$ ,

$$\begin{aligned}\mathcal{L}_i^l(\cdot) &:= A_i^l \mathcal{E}_i(\cdot) A_i^l, \\ A_i^l &:= A_i + B_i F_i^l, \\ F_i^l &:= \mathcal{F}_i(X^{l-1}) \text{ for } l = 1, 2, \dots\end{aligned}$$

c)  $X^l$  satisfaz  $X^l = \mathcal{L}^l(X^l) + \mathcal{V}(F^l)$  e é dado por  $X^l = \sum_{k=0}^{\infty} (\mathcal{L}^l)^k(\mathcal{V}(F^l))$ .

Além disso existe um  $X^+ = (X_1^+, \dots, X_N^+) \in \mathbb{H}^{n+}$  tal que  $X^+ = \mathcal{R}(X^+)$ ,  $X^+ \geq X$  para qualquer  $X \in \mathbb{H}^{n+}$  tal que  $X \geq \mathcal{R}(X)$ , e  $X^l \rightarrow X^+$  quando  $l \rightarrow \infty$ . Mais ainda  $r_\sigma(\mathcal{L}^+) \leq 1$ , onde  $\mathcal{L}^+(\cdot) = (\mathcal{L}_1^+(\cdot), \dots, \mathcal{L}_N^+(\cdot))$  é definido como  $\mathcal{L}_i^+(\cdot) = A_i^+ \mathcal{E}_i(\cdot) A_i^+$ , para  $i = 1, \dots, N$ , e

$$\begin{aligned}F_i^+ &= \mathcal{F}_i(X^+) \\ A_i^+ &= A_i + B_i F_i^+\end{aligned}$$

**Observação 1** O passo c) do Lema 5, que corresponde ao cálculo da solução do sistema linear

$$X^l = \mathcal{L}^l(X^l) + \mathcal{V}(F^l),$$

pode ser visto como o passo de avaliação de estratégias na técnica de iteração de estratégias para PMD, enquanto que da identidade (18),

$$\begin{aligned}(A_i + B_i F_i)' \mathcal{E}_i(X^l) (A_i + B_i F_i) + C_i' C_i + F_i' D_i' D_i F_i = \\ X_i^l - (F_i^l - \mathcal{F}_i(X^l))' (B_i' \mathcal{E}_i(X^l) B_i + D_i' D_i) (F_i^l - \mathcal{F}_i(X^l)) \\ + (F_i - \mathcal{F}_i(X^l))' (B_i' \mathcal{E}_i(X^l) B_i + D_i' D_i) (F_i - \mathcal{F}_i(X^l))\end{aligned}$$

e o lado direito é minimizado em  $F_i$  escolhendo  $F_i = F_i^{l+1} = \mathcal{F}_i(X^l)$ , que pode ser visto como o passo de melhoramento de estratégias.

Encerramos esta seção com o seguinte resultado que será bastante útil ao longo desta seção. Num espaço probabilístico apropriado  $(\Phi, \mathbb{P}, \{\Sigma_t\}, \Sigma)$ , considere duas seqüências de processos estocásticos  $\{W(t); t = 0, 1, \dots\}$ ,  $\{\gamma(t); t = 0, 1, \dots\}$  tal que para cada  $t = 0, 1, \dots$ ,  $W(t)$  é uma matriz  $n \times n$  e  $\gamma(t)$  é uma variável escalar positiva  $\Sigma_t$ -adaptada. Assuma que  $\mathbb{P}$ -quase sempre temos que

$$\sum_{t=0}^{\infty} \gamma(t) = \infty \quad (19)$$

e

$$\sum_{t=0}^{\infty} \gamma(t)^2 < \infty. \quad (20)$$

Assuma também, que o termo do ruído satisfaz às seguintes condições

$$\mathbb{E}(W(t) | \Sigma_t) = 0 \quad (21)$$

e

$$\mathbb{E}(\|W(t)\|^2 | \Sigma_t) \leq A(t) \quad (22)$$

onde  $\{A(t); t = 0, 1, \dots\}$  é um processo estocástico tal que para cada  $t = 0, 1, \dots$ ,  $A(t)$  é uma variável escalar positiva  $\Sigma_t$ -adaptada. Considere um processo estocástico  $\{R(t); t = 0, 1, \dots\}$ ,  $R(t)$  uma matriz  $n \times n$ , dada pela seqüência

$$R(t+1) = (1 - \gamma(t))R(t) + \gamma(t)W(t) \quad (23)$$

**Lema 6** Suponha que as equações (19)-(22) são satisfeitas e que as seqüências  $\{R(t); t = 0, 1, \dots\}$  de matrizes  $n \times n$  são dadas por (23). Se a seqüência  $A(t)$  é limitada  $\mathbb{P}$ -quase sempre então  $R(t)$  converge a zero  $\mathbb{P}$ -quase sempre.

**Prova:** Veja Corolário 4.1, página 161 em (Bertsekas e Tsitsiklis, 1996). ■

### 3 MÉTODO DE DIFERENÇA TEMPORAL QUANDO NÃO SE CONHECE A MATRIZ $\mathcal{P}$

Se a matriz de transição de probabilidade  $\mathcal{P}$  é conhecida, podemos utilizar o algoritmo descrito no Lema 5, para encontrar a solução maximal  $P$  do EARA dada pelas equações (3),(4) e então obter a lei de controle ótima  $F$  dada pela equação (6). Se a matriz de transição de probabilidade  $\mathcal{P}$  não é conhecida poderíamos tentar desenvolver um algoritmo de Monte Carlo para obter a solução  $X^l$  como no Lema 5 parte c). Entretanto, isto não é suficiente para obter  $F^{l+1} = \mathcal{F}(X^l)$ , como se pode observar da equação (17), pois há uma dependência em  $\mathcal{P}$  através do operador  $\mathcal{E}(\cdot)$ . Uma alternativa é calcular diretamente o termo  $S^l := \mathcal{E}(X^l)$ , que nos leva à seguinte equação:

$$\begin{aligned}S_i^l = \sum_{j=1}^N p_{ij} (A_j + B_j F_j^l)^* S_j^l (A_j + B_j F_j^l) \\ + \sum_{j=1}^N p_{ij} \mathcal{V}_j(F^l)\end{aligned}$$

ou

$$S^l = \mathcal{T}^l(S^l) + \mathcal{E}(\mathcal{V}(F^l)).$$

Escrevemos  $Q^l = \mathcal{E}(\mathcal{V}(F^l))$  para simplificar a notação. Como foi visto na seção 2,  $r_\sigma(\mathcal{T}^l) = r_\sigma(\mathcal{L}^l) < 1$ , e o Lema 2 pode ser usado de forma que a equação em  $Y$

$$Y = \mathcal{T}^l(Y) + Q^l \quad (24)$$

tem uma única solução  $S^l$  dada por

$$S^l = \sum_{k=0}^{\infty} (T^l)^k (Q^l). \quad (25)$$

Um vez que se tenha calculado  $S^l$ ,  $F^{l+1}$  pode ser obtida da equação (17) como

$$F_i^{l+1} = -(B_i^* S_i^l B_i + D_i^* D_i)^{-1} B_i^* S_i^l A_i, \quad i = 1, \dots, N. \quad (26)$$

O restante desta seção será dedicado a calcular  $S^l$ , via simulações de Monte Carlo, e traçando um paralelo com o método do  $TD(\lambda)$  (veja (Bertsekas e Tsitsiklis, 1996)). Para simplificar a notação, eliminaremos o superescrito  $l$ . Para qualquer  $\lambda \in [0, 1)$  definimos o operador  $\mathcal{J} \in \mathbb{B}(\mathbb{H}^n)$  e  $Z \in \mathbb{H}^n$  da seguinte maneira

$$\mathcal{J}(\cdot) := (1 - \lambda) \sum_{k=0}^{\infty} \lambda^k T^{k+1}(\cdot)$$

e

$$Z := \sum_{k=0}^{\infty} (\lambda T)^k (Q).$$

Obtemos então o seguinte resultado:

**Proposição 1**  $r_\sigma(\mathcal{J}) < 1$ .

**Prova:** Como foi mencionado acima,  $r_\sigma(T) = r_\sigma(\mathcal{L}) < 1$ . Suponha que tenhamos  $r_\sigma(\mathcal{J}) \geq 1$ . Então para algum  $\beta \in \mathbb{C}$ ,  $|\beta| \geq 1$ , e  $V \in \mathbb{H}^n$

$$\mathcal{J}(V) = \beta V.$$

Neste caso,

$$(1 - \lambda) |\beta| \geq (1 - \lambda) \Rightarrow \frac{|\beta|}{1 + \lambda(|\beta| - 1)} \geq 1$$

e

$$\begin{aligned} \lambda \beta T(V) &= \lambda T(\mathcal{J}(V)) = (1 - \lambda) \sum_{k=0}^{\infty} \lambda^{k+1} T^{k+2}(V) \\ &= (1 - \lambda) \sum_{k=1}^{\infty} \lambda^k T^{k+1}(V) \\ &= \mathcal{J}(V) - (1 - \lambda) T(V) \\ &= \beta V - (1 - \lambda) T(V) \end{aligned}$$

de forma que

$$T(V) = \frac{\beta}{1 + \lambda(\beta - 1)} V$$

e  $\frac{\beta}{1 + \lambda(\beta - 1)}$  é um autovalor de  $T$  com o autovetor associado  $V$ . Porém

$$\begin{aligned} \left| \frac{\beta}{1 + \lambda(\beta - 1)} \right| &= \frac{|\beta|}{\sqrt{(1 - \lambda)^2 + 2\lambda(1 - \lambda)\Re\{\beta\} + \lambda^2 |\beta|^2}} \\ &\geq \frac{|\beta|}{\sqrt{(1 - \lambda)^2 + 2\lambda(1 - \lambda)|\beta| + \lambda^2 |\beta|^2}} \\ &= \frac{|\beta|}{1 + \lambda(|\beta| - 1)} \geq 1 \end{aligned}$$

que é uma contradição com o fato que  $r_\sigma(T) < 1$ . Deste modo  $r_\sigma(\mathcal{J}) < 1$ . ■

Fazendo iterações na equação (24) com  $Y = S$ , temos que

$$S = T^{k+1}(S) + \sum_{t=0}^k T^{k-t}(Q), \quad k = 0, 1, \dots \quad (27)$$

Temos da equação (27) que

$$\begin{aligned} S &= \sum_{k=0}^{\infty} (1 - \lambda) \lambda^k S \\ &= (1 - \lambda) \left[ \sum_{k=0}^{\infty} \lambda^k \left( T^{k+1}(S) + \sum_{t=0}^k T^{k-t}(Q) \right) \right] \quad (28) \\ &= \sum_{t=0}^{\infty} (\lambda T)^t (Q) + (1 - \lambda) \sum_{k=0}^{\infty} \lambda^k T^{k+1}(S) = Z + \mathcal{J}(S) \\ &= \sum_{k=0}^{\infty} \lambda^k \left[ T^k(Q) + T^{k+1}(S) - T^k(S) \right] + S. \end{aligned}$$

A equação (28) evidencia o seguinte método de diferenças temporais. Seja  $\Xi := \{1, \dots, N\}^\infty$  e para cada  $i = 1, \dots, N$  e  $t = 1, 2, \dots$  considere as seguintes variáveis aleatórias  $\Theta_i(t) = (\theta_i(t, 0), \theta_i(t, 1), \dots) \in \Xi$  tal que  $\theta_i(t, 0) = i$  e  $\{\theta_i(t, k)\}$  tem a mesma distribuição que  $\{\theta(k)\}$ . Para  $k = 0, 1, \dots$  o operador afim limitado  $\mathcal{D}(t, k, \cdot)$  é definido, para  $V = (V_1, \dots, V_N) \in \mathbb{H}^n$ , em termos de  $\mathcal{B}_i(t, k) \in \mathbb{H}^n$  e  $\mathcal{C}_i(t, k, \cdot) \in \mathbb{B}(\mathbb{H}^n)$  como

$$\Upsilon_i(t, k) := (A_{\theta_i(t, k)} + B_{\theta_i(t, k)} F_{\theta_i(t, k)}) \Upsilon_i(t, k - 1)$$

$$\Upsilon_i(t, 0) := I$$

$$\mathcal{D}_i(t, k, V) := \mathcal{B}_i(t, k) + \mathcal{C}_i(t, k, V)$$

$$\begin{aligned} \mathcal{B}_i(t, k) &:= \Upsilon_i(t, k)^* \left[ C_{\theta_i(t, k+1)}^* C_{\theta_i(t, k+1)} \right. \\ &\quad \left. + F_{\theta_i(t, k+1)}^* D_{\theta_i(t, k+1)}^* D_{\theta_i(t, k+1)} F_{\theta_i(t, k+1)} \right] \Upsilon_i(t, k) \end{aligned}$$

$$\mathcal{C}_i(t, k, V) := \Upsilon_i(t, k)^* \left[ (A_{\theta_i(t, k+1)} + B_{\theta_i(t, k+1)} F_{\theta_i(t, k+1)})^* \right.$$

$$\left. V_{\theta_i(t, k+1)} (A_{\theta_i(t, k+1)} + B_{\theta_i(t, k+1)} F_{\theta_i(t, k+1)}) - V_{\theta_i(t, k)} \right] \Upsilon_i(t, k),$$

$$\mathcal{D}(t, k, V) := (\mathcal{D}_1(t, k, V), \dots, \mathcal{D}_N(t, k, V)).$$

Note que

$$E(\mathcal{D}_i(t, k, V) | \theta(0) = i) = T_i^k(Q) + T_i^{k+1}(V) - T_i^k(V)$$

e definimos para  $i = 1, \dots, N$

$$W_i(t) := \sum_{k=0}^{\infty} \lambda^k \left( \mathcal{D}_i(t, k, Y(t)) - (\mathcal{T}_i^k(Q) + \mathcal{T}_i^{k+1}(Y(t)) - \mathcal{T}_i^k(Y(t))) \right),$$

$$W(t) = (W_1(t), \dots, W_N(t)).$$

Considere  $\{\gamma(t); t = 1, 2, \dots\}$  satisfazendo as equações (19), (20) com probabilidade 1 e para um  $Y(0) = (Y_1(0), \dots, Y_N(0)) \in \mathbb{H}^n$  arbitrário defina para  $t = 1, 2, \dots$ , a seqüência  $Y(t) = (Y_1(t), \dots, Y_N(t)) \in \mathbb{H}^n$  da seguinte maneira:

$$Y_i(t+1) = Y_i(t) + \gamma(t) \sum_{k=0}^{\infty} \lambda^k \mathcal{D}_i(t, k) \quad i = 1, \dots, N,$$

ou de uma forma mais compacta,

$$Y(t+1) = Y(t) + \gamma(t) \sum_{k=0}^{\infty} \lambda^k \mathcal{D}(t, k, Y(t)) \quad (29)$$

Esta equação será fundamental no desenvolvimento do método pois, como será demonstrado no Teorema 7,  $Y(t)$  converge para  $S$ . Conforme fica evidenciado pelo somatório na equação (29), para a implementação do algoritmo necessitamos de um grande número de realizações da cadeia de Markov  $\theta(t)$ , geradas computacionalmente ou através de observações passadas, e representadas na equação (29) por  $\theta_i(t, k), k = 0, 1, \dots$ . Note que da equação (28) podemos escrever a equação (29) da seguinte maneira:

$$Y(t+1) = (1 - \gamma(t))Y(t) + \gamma(t)(Z + \mathcal{J}(Y(t)) + W(t)). \quad (30)$$

Denotamos por  $\Sigma_t$  a história do processo até o tempo  $t$ , a qual é definida como

$$\Sigma_t = \sigma \left\{ Y(0), \dots, Y(t), \Theta_i(s), s = 1, \dots, t-1, i = 1, \dots, N, \gamma(s), s = 1, \dots, t \right\}.$$

Então para cada  $i = 1, \dots, N$

$$\mathbb{E}(W_i(t) | \Sigma_t) = \sum_{k=0}^{\infty} \lambda^k \left( E \left( \mathcal{D}_i(t, k, Y(t)) | \theta(0) = i \right) - \left( \mathcal{T}_i^k(Q) + \mathcal{T}_i^{k+1}(Y(t)) - \mathcal{T}_i^k(Y(t)) \right) \right) = 0.$$

Na seguinte proposição considere  $\mathcal{H}$  como na equação (16) com  $\Gamma_i = A_i + B_i F_i$ .

**Proposição 2** Se  $\lambda^2 r_\sigma(\mathcal{H}) < 1$  então existem constantes  $a > 0, b > 0$  tais que

$$\mathbb{E} \left( \|W_i(t)\|^2 | \Sigma_t \right) \leq a \|Y(t)\|^2 + b.$$

**Prova:** Visto que  $W_i(t)$  é hermitiana, podemos encontrar um  $x \in \mathbb{C}^n, \|x\|_2 = 1$ , tal que  $\|W_i(t)\|_2 = |x^* W_i(t) x|$ . Observe que

$$x^* W_i(t) x = \sum_{k=0}^{\infty} \lambda^k \left( \left\{ \|C_{\theta_i(t, k+1)} x(k+1)\|_2^2 + \|D_{\theta_i(t, k+1)} F_{\theta_i(t, k+1)} x(k+1)\|_2^2 + (1 - \lambda) x(k+2)^* Y_{\theta_i(t, k+1)}(t) x(k+2) \right\} - \left\{ E \left( \|C_{\theta_i(t, k+1)} x(k+1)\|_2^2 + \|D_{\theta_i(t, k+1)} F_{\theta_i(t, k+1)} x(k+1)\|_2^2 + (1 - \lambda) x(k+2)^* Y_{\theta_i(t, k+1)}(t) x(k+2) \right) \right\} \right)$$

onde

$$x(k+1) = (A_{\theta_i(t, k)} + B_{\theta_i(t, k)} F_{\theta_i(t, k)}) x(k)$$

$$x(0) = x, \theta_i(t, 0) = i.$$

Como visto na seção 2,  $E(\|x(k)\|^2) = \sum_{i=1}^N \text{tr}\{V_i(k)\}$ , onde  $V_i(k) = E(x(k)x(k)^* 1_{\theta_i(t, k)=i})$ ,  $V(k) = (V_1(k), \dots, V_N(k)) \in \mathbb{H}^{n^2}$

$$V(k+1) = \mathcal{G}(V(k)), \quad k = 0, 1, \dots$$

com  $\Gamma_i = A_i + B_i F_i$  em (14). Como  $r_\sigma(\mathcal{G}) < 1$ , podemos encontrar  $c_0 > 0, c_1 \in (0, 1)$  tais que

$$\|\mathcal{G}^k\| \leq c_0 c_1^k.$$

Dessa forma, definindo  $c_3 = \|C\|_{max} + \|D\|_{max} \|F\|_{max}$  e observando que

$$\sum_{i=1}^N \text{tr}\{\mathcal{G}_i^k(V(0))\} \leq nN \|\mathcal{G}_i^k(V(0))\|$$

segue que

$$\begin{aligned} & E \left( \|C_{\theta_i(t, k+1)} x(k+1)\|_2^2 + \|D_{\theta_i(t, k+1)} F_{\theta_i(t, k+1)} x(k+1)\|_2^2 \right) \\ & \leq c_3 E \left( \|x(k+1)\|_2^2 \right) \\ & = c_3 \sum_{i=1}^N \text{tr}\{\mathcal{G}_i^{k+1}(V(0))\} \leq c_3 c_0 N n c_1^{k+1} \|V(0)\| \\ & = c_3 c_0 N c_1^{k+1} \|x\|^2 = c_3 c_0 N c_1^{k+1}. \end{aligned}$$

e similarmente

$$E(x(k+2)^* Y_{\theta_i(t, k+1)} x(k+2)) \leq c_0 N n c_1^{k+2} \|Y(t)\|.$$

Portanto para alguma constante  $c_4 > 0$ ,

$$\begin{aligned} \mathbb{E} \left( \|W_i(t)\|^2 \mid \Sigma_t \right) &\leq c_4(1 + \|Y(t)\|^2) \\ &+ c_4 \left( E \left( \left( \sum_{k=0}^{\infty} \lambda^k \|x(k+1)\|_2^2 \right)^2 \right) \right) \\ &+ E \left( \left( \sum_{k=0}^{\infty} \lambda^k \|x(k+2)\|_2^2 \right)^2 \right) \|Y(t)\|^2. \end{aligned}$$

Finalmente observe que do Lema 3 temos que

$$\begin{aligned} &E \left( \left( \sum_{k=0}^{\infty} \lambda^k \|x(k+1)\|_2^2 \right)^2 \right) \\ &\leq \left( \sum_{k=0}^{\infty} \lambda^k \sqrt{E(\|x(k+1)\|_2^4)} \right)^2 \\ &= \left( \sum_{k=0}^{\infty} \lambda^k \sqrt{\sum_{i=1}^N \text{tr}\{\mathcal{H}_i^{k+1}(S(0))\}} \right)^2 \\ &\leq \left( \frac{1}{\lambda} N n^2 \sum_{k=0}^{\infty} \sqrt{\lambda^{2(k+1)} \|\mathcal{H}^{k+1}\|} \right)^2. \end{aligned}$$

Se  $\lambda^2 r_\sigma(\mathcal{H}) < 1$  então podemos encontrar  $c_5 > 0$ ,  $c_6 \in (0, 1)$  tais que

$$\lambda^{2k} \|\mathcal{H}^k\| \leq c_5 c_6^k$$

e

$$\sum_{k=0}^{\infty} \sqrt{\lambda^{2(k+1)} \|\mathcal{H}^{k+1}\|} \leq c_5^{\frac{1}{2}} \sum_{k=0}^{\infty} \left( c_6^{\frac{1}{2}} \right)^{k+1} = c_7.$$

Deste modo para algum  $c_8 > 0$ ,

$$\begin{aligned} &c_4 \left( E \left( \left( \sum_{k=0}^{\infty} \lambda^k \|x(k+1)\|_2^2 \right)^2 \right) \right) \\ &+ E \left( \left( \sum_{k=0}^{\infty} \lambda^k \|x(k+2)\|_2^2 \right)^2 \right) \|Y(t)\|^2 \\ &\leq c_8 \left( 1 + \|Y(t)\|^2 \right) \end{aligned}$$

e o resultado segue.  $\blacksquare$

Assim as equações (21) e (22) são satisfeitas para cada  $W_i(t)$ ,  $i = 1, \dots, N$ . O seguinte resultado mostra que, através de uma transformação no algoritmo dado por (30) podemos assumir que  $\|\mathcal{J}\| < 1$ .

**Observação 2** Sem perder a generalidade, assumimos que  $\|\mathcal{J}\| < 1$ .

Suponha que o resultado não seja válido. Visto que  $r_\sigma(\mathcal{J}) < 1$  (Proposição 1), sabemos que do Lema 1 existe um  $\mathcal{W} \in \mathbb{B}(\mathbb{X})$  invertível e real e  $\tilde{\mathcal{J}} \in \mathbb{B}(\mathbb{X})$  tal que  $\|\tilde{\mathcal{J}}\| < 1$  e

$$\mathcal{J} = \mathcal{W}\tilde{\mathcal{J}}\mathcal{W}^{-1}.$$

Considere a seguinte transformação, no algoritmo descrito pela equação (29)

$$\begin{aligned} \tilde{Y}(t) &= \mathcal{W}Y(t), \quad \tilde{\mathcal{D}}_i(t, k, \cdot) = \tilde{\mathcal{B}}_i(t, k) + \tilde{\mathcal{C}}_i(t, k, \cdot), \\ \tilde{\mathcal{B}}_i(t, k) &= \mathcal{W}\mathcal{B}_i(t, k), \quad \tilde{\mathcal{C}}_i(t, k, \cdot) = \mathcal{W}\mathcal{C}_i(t, k, \cdot)\mathcal{W}^{-1}. \end{aligned}$$

Então da equação (29) segue que

$$\tilde{Y}(t+1) = \tilde{Y}(t) + \gamma(t) \sum_{k=0}^{\infty} \lambda^k \tilde{\mathcal{D}}(t, k, \tilde{Y}(t)) \quad (31)$$

a qual pode ser reescrita como:

$$\tilde{Y}(t+1) = (1 - \gamma(t)) \tilde{Y}(t) + \gamma(t) \left( \tilde{Z} + \tilde{J}(\tilde{Y}(t)) + \tilde{W}(t) \right) \quad (32)$$

onde

$$\tilde{Z} = \mathcal{W}Z, \quad \tilde{W}(t) = \mathcal{W}W(t).$$

É fácil verificar-se que  $\mathbb{E}(\tilde{W}_i(t) \mid \tilde{\Sigma}_t) = 0$  e  $\mathbb{E}(\|\tilde{W}_i(t)\|^2 \mid \tilde{\Sigma}_t) \leq \tilde{a}\|\tilde{Y}(t)\|^2 + \tilde{b}$  para algum  $\tilde{a} > 0$ ,  $\tilde{b} > 0$  onde  $\tilde{\Sigma}_t$  é definido de forma apropriada com  $\tilde{Y}(s)$  no lugar de  $Y(s)$ . Assim, se  $\|\mathcal{J}\| \geq 1$ , podemos aplicar a transformação linear acima e trabalhar com a representação do algoritmo dado pela equações (31), (32), em lugar de (29), (30).

Assim, de agora em diante, assumimos, sem perda de generalidade que,  $\|\mathcal{J}\| < 1$  na equação (30). O seguinte resultado segue os mesmos argumentos apresentados em (Bertsekas e Tsitsiklis, 1996), páginas 162-167.

**Proposição 3** Se  $\lambda^2 r_\sigma(\mathcal{H}) < 1$  então a seqüência  $\{Y(t); t = 1, 2, \dots\}$  é limitada com probabilidade 1.

**Prova:** Esta prova segue os mesmos passos que a prova da Proposição 4.7 em (Bertsekas e Tsitsiklis, 1996), página 162-166. Como  $\|\mathcal{J}\| < 1$  podemos encontrar um  $G$  tal que  $G \geq 1$  e  $G > \frac{\|\mathcal{Z}\|}{1 - \|\mathcal{J}\|}$ , e definimos  $\eta$  e  $\epsilon$  como  $\|\mathcal{J}\| < \eta := \frac{\|\mathcal{Z}\|}{G} + \|\mathcal{J}\| \in (0, 1)$ ,  $\epsilon := \frac{1}{\eta} - 1 > 0$ . Como em (Bertsekas e Tsitsiklis, 1996), página 163, a seqüência  $\Sigma_t$ -adaptada não-decrescente  $\{G(t); t = 1, 2, \dots\}$  é definida como segue:  $G(0) = \max\{\|Y(0)\|, G\}$  e

$$G(t+1) = \begin{cases} G(t), & \text{se } \|Y(t+1)\| \leq (1 + \epsilon)G(t) \\ G(0)(1 + \epsilon)^\tau, & \text{caso contrário} \end{cases}$$



onde

$$\tau = \min\{s \geq 0; \|Y(t+1)\| \leq (1+\epsilon)^s G(0)\}.$$

Conseqüentemente para todo  $t = 1, 2, \dots$ ,  $\|Y(t)\| \leq (1+\epsilon)^t G(0)$  e  $\|Y(t)\| \leq G(t)$  se  $G(t-1) < G(t)$ . Definimos  $\widetilde{W}(t) = \frac{1}{G(t)}W(t)$ , de forma que

$$\mathbb{E}\left(\widetilde{W}_i(t) \mid \Sigma_t\right) = \frac{1}{G(t)}\mathbb{E}(W_i(t) \mid \Sigma_t) = 0 \quad (33)$$

e para  $c = a(1+\epsilon)^2 + b$

$$\begin{aligned} \mathbb{E}\left(\left\|\widetilde{W}_i(t)\right\|^2 \mid \Sigma_t\right) &\leq \frac{1}{G(t)^2}(a\|Y(t)\|^2 + b) \\ &\leq \frac{1}{G(t)^2}(a(1+\epsilon)^2 G(t)^2 + b) \leq c \end{aligned} \quad (34)$$

Definindo para  $i = 1, \dots, N$ ,  $t \geq t_0 \geq 1$ ,  $R_i(t_0, t_0) = 0$  e

$$R_i(t+1, t_0) = (1-\gamma(t))R_i(t, t_0) + \gamma(t)\widetilde{W}_i(t)$$

temos do Lema 6, (33) e (34), que  $R_i(t, t_0)$  tende para zero, quando  $t$  tende para infinito com probabilidade 1. Considere o conjunto  $\Gamma \subset \Sigma$ , tal que as equações (19), (20) são válidas e  $R_i(t, t_0)$  tende a zero, quando  $t$  tende a infinito para cada  $t_0 = 1, 2, \dots, i = 1, \dots, N$ . É fácil checar que  $\Gamma$  tem probabilidade 1 (é a intersecção de conjuntos contáveis com probabilidade 1). Suponha por contradição que  $Y(t)(\omega)$ ,  $\omega \in \Gamma$ , é ilimitado, de forma que  $G(t)(\omega) \rightarrow \infty$  a medida que  $t \rightarrow \infty$  e  $\|Y(t)(\omega)\| \leq G(t)(\omega)$  para infinitamente freqüentes  $t$ . Seja  $u(\omega) \geq 1$  tal que  $\gamma(s)(\omega) < 1$  para  $s \geq u(\omega)$ . Considere  $t_0(\omega) \geq u(\omega)$  tal que  $\|Y(t_0)(\omega)\| \leq G(t_0)(\omega)$  e  $\|R_i(t, 0)(\omega)\| \leq \frac{\epsilon}{2\sqrt{N}}$  para  $i = 1, \dots, N$ ,  $t \geq t_0(\omega) \geq u(\omega)$ . Visto que

$$R_i(t, 0)(\omega) = \prod_{s=t_0}^{t-1} (1-\gamma(s))R_i(t_0, 0)(\omega) + R_i(t, t_0)(\omega)$$

temos para  $t \geq t_0(\omega) \geq u(\omega)$  e  $R(t, t_0)(\omega) = (R_1(t, t_0)(\omega), \dots, R_N(t, t_0)(\omega)) \in \mathbb{H}^n$  que

$$\begin{aligned} \|R(t, t_0)(\omega)\| &= \left(\sum_{i=1}^N \|R_i(t, t_0)(\omega)\|^2\right)^{\frac{1}{2}} \\ &\leq \left(\sum_{i=1}^N (\|R_i(t, 0)(\omega)\| + \|R_i(t_0, 0)(\omega)\|)^2\right)^{\frac{1}{2}} \\ &\leq \left(\sum_{i=1}^N \frac{\epsilon^2}{N}\right)^{\frac{1}{2}} = \epsilon \end{aligned} \quad (35)$$

Além disso temos que

$$\begin{aligned} \|Z + \mathcal{J}(Y(t))\| &\leq \|Z\| + \|\mathcal{J}\| \|Y(t)\| \\ &\leq \|Z\| + \|\mathcal{J}\|(1+\epsilon)G(t) = (\eta - \|\mathcal{J}\|)G + \|\mathcal{J}\|(1+\epsilon)G(t) \\ &\leq (\eta + \|\mathcal{J}\|\epsilon)G(t) < \eta(1+\epsilon)G(t) = G(t) \end{aligned} \quad (36)$$

Falta mostrar por indução que para  $t \geq t_0(\omega)$

$$\|Y(t)(\omega) - R(t, t_0)(\omega)G(t_0)(\omega)\| \leq G(t_0)(\omega) \quad (37)$$

e

$$G(t)(\omega) = G(t_0)(\omega). \quad (38)$$

Para  $t = t_0(\omega)$  o resultado é trivial pois  $R(t_0, t_0)(\omega) = 0$  e por hipótese,  $\|Y(t_0)(\omega)\| \leq G(t_0)(\omega)$ . Suponha que as equações (37) e (38) são válidas para  $t$ . Então (por simplicidade suprimimos  $\omega$ ) a partir da equação (38), temos que

$$\begin{aligned} Y(t+1) - R(t+1, t_0)G(t_0) &= \\ &= 1 - \gamma(t)(Y(t) - R(t, t_0)G(t_0)) \\ &\quad + \gamma(t)(Z + \mathcal{J}(Y(t)) + W(t)) - \gamma(t)\widetilde{W}(t)G(t) \\ &= (1-\gamma(t))(Y(t) - R(t, t_0)G(t_0)) \\ &\quad + \gamma(t)(Z + \mathcal{J}(Y(t))) + \gamma(t)W(t) - \gamma(t)W(t) \\ &= (1-\gamma(t))(Y(t) - R(t, t_0)G(t_0)) + \gamma(t)(Z + \mathcal{J}(Y(t))) \end{aligned}$$

e das equações (36), (37), (38) que

$$\begin{aligned} \|Y(t+1) - R(t+1, t_0)G(t_0)\| &\leq (1-\gamma(t))\|Y(t) - R(t, t_0)G(t_0)\| \\ &\quad + \gamma(t)\|Z + \mathcal{J}(Y(t))\| \\ &\leq (1-\gamma(t))G(t_0) + \gamma(t)G(t_0) = G(t_0) \end{aligned}$$

provando (37) para  $t+1$ . Da equação (35) temos que

$$\begin{aligned} \|Y(t+1)\| &\leq \|Y(t+1) - R(t+1, t_0)G(t_0)\| \\ &\quad + \|R(t+1, t_0)G(t_0)\| \\ &\leq G(t_0) + \epsilon G(t_0) = G(t_0)(1+\epsilon) = G(t)(1+\epsilon) \end{aligned}$$

e pela definição  $G(t+1) = G(t) = G(t_0)$  provando (38) para  $t+1$ . Entretanto isto torna-se uma contradição, com o fato de que  $G(t)$  tende para infinito, provando o resultado. ■

Obtemos agora o resultado principal desta seção, provando a convergência de  $Y(t)$  para  $S$ .

**Teorema 7** Se  $\lambda^2 r_\sigma(\mathcal{H}) < 1$  então a seqüência  $\{Y(t); t = 1, 2, \dots\}$  converge para  $S$  com probabilidade 1.

**Prova:** Esta prova segue os mesmos passos que a prova da Proposição 4.5 no livro (Bertsekas e Tsitsiklis, 1996), páginas 166-167. Primeiro note que, a partir da equação (28),

$$S = Z + \mathcal{J}(S) = (1-\gamma(t))S + \gamma(t)(Z + \mathcal{J}(S))$$

e então definindo  $X(t) = Y(t) - S$ , temos

$$X(t+1) = (1-\gamma(t))X(t) + \gamma(t)(\mathcal{J}(X(t)) + W(t)).$$

Defina para  $i = 1, \dots, N$ ,  $t \geq t_0 \geq 1$ ,  $R_i(t_0, t_0) = 0$  e

$$R_i(t+1, t_0) = (1-\gamma(t))R_i(t, t_0) + \gamma(t)W_i(t).$$

A partir do Lema 6, Proposição 2 e 3 temos que  $R_i(t, t_0)$  tende a zero, à medida que  $t$  tende a infinito com probabilidade 1. Seja  $\Lambda \subset \Sigma$  um conjunto tal que a seqüência  $\{Y(t); t = 1, 2, \dots\}$  é limitada e  $R_i(t, t_0)$  tende a zero à medida que  $t$  tende a infinito para todo  $i = 1, \dots, N, t_0 \geq 1$ . Visto que este conjunto é a intersecção de conjuntos contáveis com probabilidade 1, temos que  $\mathbb{P}(\Lambda) = 1$ . Escrevamos  $R(t, t_0) = (R_1(t, t_0), \dots, R_N(t, t_0)) \in \mathbb{H}^n$ . Para cada  $\omega \in \Lambda$  existe um  $d(\omega)$  tal que para todo  $t \geq 1$ , temos que

$$\|X(t)(\omega)\| \leq \|Y(t)(\omega)\| + \|S\| \leq d(\omega).$$

Usamos  $\nu > 0$  tal que  $\|\mathcal{J}\| + \nu < 1$  e  $d_{k+1} = (\|\mathcal{J}\| + \nu)d_k$ ,  $d_0 = d$ . Pegamos  $t_0(\omega) = u(\omega)$  onde  $u(\omega)$  é tal que  $\gamma(s)(\omega) < 1$  para  $s \geq u(\omega)$ ) e provamos que existe sempre um  $t_{k+1}(\omega) \geq t_k(\omega)$  tal que

$$\|X(t)(\omega)\| \leq d_k(\omega), \quad t \geq t_k(\omega). \quad (39)$$

Para simplificar a notação, omitimos o  $\omega$ . Para  $t = 0$  o resultado é direto. Suponha que a equação (39) é válida para  $k$ . Considere a seguinte seqüência

$$y(t+1) = (1 - \gamma(t))y(t) + \gamma(t)\|\mathcal{J}\|d_k, \quad t \geq t_k, \quad y(t_k) = d_k. \quad (40)$$

Mostramos por indução que

$$\|X(t) - R(t, t_k)\| \leq y(t), \quad t \geq t_k. \quad (41)$$

Para  $t = t_k$  temos que  $y(t_k) = d_k$ ,  $R(t_k, t_k) = 0$  e o resultado segue da equação (39). Suponha que a equação (41) é válida para  $t$ . Então

$$X(t+1) - R(t+1, t_k) = (1 - \gamma(t))(X(t) - R(t, t_k)) + \gamma(t)\mathcal{J}(X(t))$$

e visto que  $\|\mathcal{J}(X(t))\| \leq \|\mathcal{J}\|\|X(t)\|$ , temos das equações (39), (40) e (41) que

$$\|X(t+1) - R(t+1, t_k)\| \leq (1 - \gamma(t))y(t) + \gamma(t)\|\mathcal{J}\|d_k = y(t+1)$$

mostrando (41) para  $t+1$ . Desta forma a equação (41) é válida para todo  $t \geq t_k$ . Como  $y(t)$  converge para  $\|\mathcal{J}\|d_k$  e  $R(t, t_k)$  tende a 0 à medida que  $t$  tende a infinito, podemos encontrar um  $t_{k+1} \geq t_k$  tal que  $y(t) \leq (\|\mathcal{J}\| + \frac{\nu}{2})d_k$  e  $\|R(t, t_k)\| \leq \frac{\nu}{2}d_k$  para todo  $t \geq t_{k+1}$ . Deste modo da equação (41) temos que para todo  $t \geq t_{k+1}$

$$\begin{aligned} \|X(t)\| &\leq \|X(t) - R(t, t_k)\| + \|R(t, t_k)\| \\ &\leq y(t) + \frac{\nu}{2}d_k \leq (\|\mathcal{J}\| + \nu)d_k = d_{k+1} \end{aligned}$$

provando a equação (39) para  $k+1$ . Como  $d_k$  tende a zero à medida que  $k$  tende a infinito, o resultado segue. ■

**Observação 3** Na prática devemos testar  $\lambda$  para verificar se a convergência de  $Y(t)$  para  $S$  ocorre, visto que  $\mathcal{P}$  é desconhecido, e deste modo não é possível checar de antemão se  $\lambda^2 r_\sigma(\mathcal{H}) < 1$ . Deve-se destacar também que o algoritmo necessita de um  $F^0$  que estabiliza  $(A, B)$  na média quadrática para iniciar as iterações.

## 4 EXEMPLO NUMÉRICO

Para ilustrar o uso do resultados desenvolvidos nas seções anteriores, escolhemos um sistema econômico simples baseado no modelo multiplicador-acelerador de Samuelson's (Blair e Sworder, 1975), o qual aparece na forma de equações de estado:

$$x(k+1) = A_{\theta(k)}x(k) + B_{\theta(k)}u(k).$$

Consideramos os seguintes três modos de operação do sistema linear a saltos Markovianos a tempo discreto:

$i$	Modos de operação					
	1		2		3	
$A_i$	0	1	0	1	0	1
	-2.5	3.2	-4.3	4.5	5.3	-5.2
$B_i$	0		0		0	
	1		1		1	
$C'_i C_i$	3.6	-3.8	10	-3	5	-4.5
	-3.8	4.87	-3	8	-4.5	4.5
$D'_i D_i$	2.6		1.165		1.111	

Tabela 1: Dados do modelos multiplicador-acelerador

A matriz de transição de probabilidades para este sistema é assumida como:

$$P = \begin{bmatrix} 0.67 & 0.17 & 0.16 \\ 0.30 & 0.47 & 0.23 \\ 0.26 & 0.10 & 0.64 \end{bmatrix}.$$

Definimos o algoritmo para  $\ell = 0, 1, \dots$ , da seguinte forma:

- i)  $F_i^\ell$  é calculado de acordo com a equação (26) (exceto para  $\ell = 0$ ).
- ii)  $S^\ell$  é o valor estacionário da equação (29).

Conforme mencionado anteriormente, fica evidenciado pelo somatório na equação (29) que para a implementação do algoritmo necessitaríamos de um grande número de realizações da cadeia de Markov  $\theta(t)$ , geradas computacionalmente ou através de observações passadas, e representadas na equação (29) por  $\theta_i(t, k), k = 0, 1, \dots$ . Deve-se notar também que, conforme mencionado na Observação 3, necessita-se de um  $F^0$  que estabilize  $(A, B)$  na média quadrática para iniciar as iterações.

$\lambda$	$\ell$	$\Delta_i$	
0.0	11	$\Delta_1 =$	0.1979 0.8081
		$\Delta_2 =$	0.0229 0.3016
		$\Delta_3 =$	0.0720 0.0613
0.3	11	$\Delta_1 =$	0.0027 0.2055
		$\Delta_2 =$	0.0475 0.3232
		$\Delta_3 =$	0.0707 0.0980
0.5	11	$\Delta_1 =$	0.2063 0.7898
		$\Delta_2 =$	0.0046 0.2450
		$\Delta_3 =$	0.0432 0.2307
0.7	11	$\Delta_1 =$	0.4042 0.6838
		$\Delta_2 =$	0.0835 0.4226
		$\Delta_3 =$	0.0653 0.1057
1.0	11	$\Delta_1 =$	0.1346 0.7351
		$\Delta_2 =$	0.0606 0.2358
		$\Delta_3 =$	0.0180 0.0403

Tabela 2: Resultados do algoritmo  $TD(\lambda)$  aplicado ao modelo multiplicador-acelerador

Foram realizados vários experimentos com diferentes valores  $\lambda$  (vide Tabela 2). Na coluna 3 da Tabela 2 mostra-se o erro, que foi calculado de acordo com a seguinte equação:

$$\Delta_i = \left\| \frac{F_i^{real} - F_i}{F_i^{real}} \right\| * 100.$$

onde  $F_i^{real}$  é o ganho do controlador ótimo associado ao modo  $i$ .

Além dos dados observados na tabela anterior, mostraremos também, nas figuras seguintes, a evolução do valor de  $F_i$  em relação ao número de iterações  $\ell$ , para um dos valores de  $\lambda$  (para os outros valores o comportamento é bastante semelhante).

## 5 CONCLUSÕES

Neste trabalho traçou-se um paralelo com o método de simulação Monte Carlo- $TD(\lambda)$  para PMD (veja por exemplo (Bertsekas e Tsitsiklis, 1996),(Sutton e Barto, 1998)), para obter o controle ótimo associado ao conjunto de equações algébricas de Riccati acopladas entre si (EARA), para controle ótimo de sistemas lineares com saltos Markovianos a tempo discreto. Assumimos que a matriz de transição de probabilidade  $\mathcal{P}$  é desconhecida, mas é possível simular as trajetórias da cadeia de Markov  $\theta(k)$ . Relacionamos as iterações (8),(9) com aquelas associadas à técnica de iteração de valores em PMD, e o método chamado de quase-linearização apresentado no Lema 5 com a técnica de iteração de estratégias para PMD, envolvendo uma parte de avaliação de estratégias, e a parte de melhoramento de estratégias (veja Observação 1).

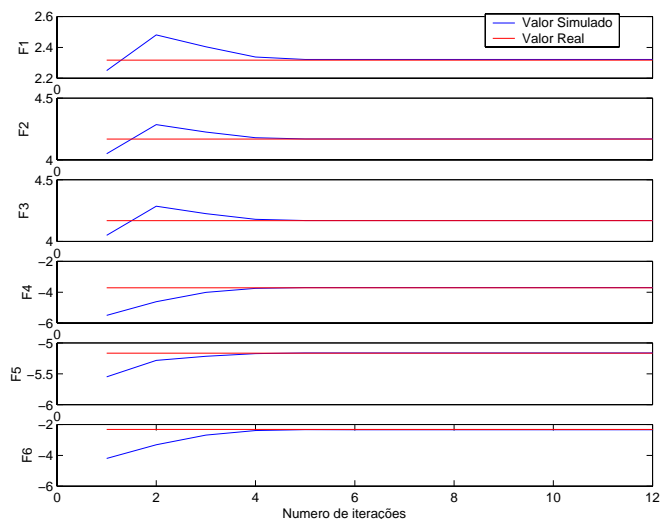


Figura 1: Gráfico dos valores de  $F_i$  para  $\lambda = 0.5$  - Exemplo do multiplicador-acelerador

Na técnica de iteração de estratégias, a avaliação de estratégias é feita via simulações de Monte Carlo, usando o método de diferenças temporais. Aplicando estas idéias ao método de quase-linearização apresentado no Lema 5, obtivemos um método iterativo que traça um paralelo com os métodos de simulação  $TD(\lambda)$  em PMD. Para isto foi mostrado no Teorema 7 que se o  $\lambda$  escolhido é suficientemente pequeno, a convergência do algoritmo  $TD(\lambda)$  na avaliação de custo acontece com probabilidade 1. São apresentados exemplos numéricos para ilustrar os resultados.

## REFERÊNCIAS

- Abou-Kandil, H., Freiling, G. e Jank, G. (1995). On the solution of discrete-time markovian jump linear quadratic control problems, *Automatica* **31**(5): 765–768.
- Ait-Rami, M. e Ghaoui, L. E. (1996). LMI optimization for nonstandard riccati equations arising in stochastic control, *IEEE Trans. Automatic Control* **41**(11): 1666–1671.
- Bertsekas, D. P. e Tsitsiklis, J. (1996). *Neuro-Dynamic Programming*, Athena Scientific.
- Blair, W. e Sworder, D. (1975). Feedback control of a class of linear discrete system with jump parameters and quadratic cost criteria, *Int. J. Control* **21**: 833–841.
- Brewer, W. (1978). Kronecker product and matrix calculus in system theory, *IEEE Trans. Circuits and Systems* **25**: 772–781.

- Costa, O. e Aya, J. (1999). Temporal difference methods for the maximal solution of discrete-time coupled algebraic riccati equations, *Proc. American Control Conference* pp. 1791–1795.
- Costa, O. e Fragoso, M. (1993). Stability results for discrete-time linear systems with markovian jumping parameters, *J. Math. Analysis and Applic* **179**: 154–178.
- Costa, O. e Fragoso, M. (1995). Discrete-time LQ-optimal control problems for infinite markov jump parameter systems, *IEEE Trans. Automat. Control* **40**: 2076–2088.
- Costa, O. e Marques, R. (1999). Maximal and stabilizing hermitian solutions for discrete-time coupled algebraic riccati equations, *Mathematics of Control, Signals and Systems* **12**(2): 167–195.
- Costa, O., Val, J. D. e Geromel, J. (1997). A convex programming approach to  $\mathcal{H}_2$ -control of discrete-time markovian jump linear systems, *Int. J. Control* **66**: 557–579.
- Gajic, Z. e Borno, I. (1995). Lyapunov iterations for optimal control of jump linear systems at steady state, *IEEE Trans. Automat. Control* **40**(11): 481–498.
- Ji, Y. e Chizeck, H. (1988). Controllability, observability and discrete-time markovian jump linear quadratic control, *Int. J. Control* **48**: 481–498.
- Ji, Y., Chizeck, H., Feng, X. e Loparo, K. (1991). Stability and control of discrete-time jump linear systems, *Control Th. and Adv. Tech* **7**: 247–270.
- Kubrusly, C. (1997). *An Introduction to Models and Decompositions in Operator Theory*, Springer Verlag, New York.
- Mariton, M. (1988). Almost sure and moments stability of jump linear systems, *Systems and Control Letters* **11**: 393–397.
- Mariton, M. (1990). *Jump Linear Systems in Automatic Control*, Marcel Dekker.
- Sutton, R. S. e Barto, A. G. (1998). *Reinforcement Learning - An Introduction*, MIT Press.
- Val, J. D., Geromel, J. e Costa, O. (1998). Uncoupled riccati iterations for the linear quadratic control problem of discrete-time markov jump linear systems, *IEEE Trans. Automat. Control* **43**(12): 1727–1733.
- Weidman, J. (1980). *Linear Operators in Hilbert Spaces*, Springer Verlag, New York.